Improving the Accessibility of Online Data Visualizations for Screen-Reader Users and Visualization Creators

Ather Sharif

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:
Jacob O. Wobbrock, Chair
Jeffrey Heer
Richard E. Ladner

Program Authorized to Offer Degree:
Computer Science and Engineering

University of Washington

Abstract

Improving the Accessibility of Online Data Visualizations for
Screen-Reader Users and Visualization Creators

Ather Sharif

Chair of the Supervisory Committee:
Jacob O. Wobbrock
The Information School

This dissertation introduces a novel interaction technique for screen-reader users to interact with and extract information from online data visualizations using voice-activated commands. To this end, this dissertation pioneers VoxLens, a multi-modal open-source JavaScript plug-in that—with a single line of code—improves the (1) information extraction experiences of screen-reader users with online data visualizations and (2) understanding and knowledge of visualization creators to make data visualizations accessible.

I present versions of VoxLens and independent artifacts, including VoxEx—a system that enables screen-reader users to customize the information they consume from online data visualizations. These artifacts collectively enable these users to extract data from simple and complex online data visualizations, both holistically and granularly, in the manner they prefer. The artifacts also provide these users with information on data uncertainty. VoxLens increased their accuracy of information extraction by 164% and reduced their interaction times by 50% over conventional methods to consume information from online data visualizations. Additionally, I present five interventions that minimize creators' challenges with accessibility. This work provides empirical and artifact contributions to the domains of accessibility and visualization.

The thesis of this dissertation is as follows: *A multi-modal, customizable, and interactive JavaScript plug-in called "VoxLens" improves the experiences of screen-reader users in extracting information from simple and complex online data visualizations while also enhancing the knowledge of visualization creators to make online data visualizations accessible.*

Table of Contents

List of Figures

List of Tables

Acknowledgments

While this dissertation uses singular first-person pronouns throughout, without an ounce of doubt, it would not have been possible without the relentless contributions of several people. Considering it a moral and ethical obligation to give credit where it is due, I am grateful to everyone who contributed to my personal and professional journey here at the University of Washington (UW) over the past six years.

First and foremost, I thank my advisor, Jacob O. Wobbrock. When I started my doctorate at UW, I didn't know much about research. Jake was patient with me. He taught me everything I needed to know to become the researcher I am today. He supported me during times of dispiritedness, not just when I needed to express my fuming vexation toward R2 but also when I lost my motivation to keep moving forward. I struggled to find the right advisor until the end of my third year, and I can confidently say that having Jake as my advisor was one of the best things that happened to me at UW. To Jake: *I aspire to be as eloquent and elegant as you are in stating complex thoughts. I appreciate you and everything you've done for me in my personal and professional growth more than you would ever know. I will always look up to you.*

I am grateful to my dissertation committee for their guidance and support, and for writing a prodigious amount of recommendation letters for me. I thank Richard E. Ladner for his unparalleled mentoring and benevolent advice. To Richard: *I know I'm neither the first nor the last to say this; you will always be my role model.* I thank Jeffrey Heer for his invaluable insights on this work and for kindly offering me career advice when I needed it the most. To Jeff: *Every time I reached out for your assistance, whether you were in Germany on sabbatical or walking back home, you have always made time for me; it means a lot.* I thank Joanne Woaik for her commitment to creating awareness of disabilities and for inviting me as a guest lecturer in her "Introduction to Disability Studies" course. To Joanne: *Conversations with you have strengthened my connection to my disabilities. I will always advise my students to take a course on disability studies.*

I am extremely fortunate to have worked with brilliant and thoughtful collaborators and researchers at UW and beyond who positively impacted me and my research work. These include (in no particular order): Anat Caspi, Jennifer Mankoff, Gary Hsieh, Jon E. Froehlich, Shaun Kane, Meredith Ringel Morris, Dhruv Jain, Victoria Pao, Qisheng Li, Negeen Aghassibake, Ruolin Wang, Emilia F. Gan, Miranda Wei, Jaden H. Wang, Alida T. Muongchan, Michael Saugstad, Manaswi Saha, Ploypilin Pruekcharoen, Ruoxi Shang, Spencer Williams, Andrew M. Zhang, Anna Shih, Arvind Krishnamurthy, Aedan L. McCall, Ruican Zhong, James Fogarty, Dorian Arnold, Arjun Srinivasan, Alexis Hiniker, Srihari N. Krishnaswamy, Tal August, Neha Aitharaju, Brianna Blaser, Jessie Z. Xu, Bongshin Lee, Joo Gyeong Kim, Kate Glazko, Kent R. Zeng, Qianqian Yu, Emma McDonnell, Lucy Jiang, Lisa Orii, Avery Mack, Chu Li, members of the ACE Lab, and several others who I owe apology notes to for forgetting their names. To all of you: *I will forever cherish sharing research ideas and doing thematic analysis over greasy pizza, boba, and One Direction songs.*

I am thankful for the support (financial and collaborative) that I received from organizations, including (in no particular order): AccessComputing, Center for Research and Education on Accessible Technology and Experiences at UW, Mani Charitable Foundation, Open Scholarship Commons, Paul G. Allen School of Computer Science and Engineering, Google, CMD-IT, LEAP Fellows, National Foundation of the Blind, UW Disability Resource for Students, UW Libraries, UW Graduate School, DO-IT, UW eScience Institute, Lime Connect, Golden Key Honour Society, Disability Empowerment Center, and the UAW 4121.

I remember Marco Casas Beaux, the owner of BOCA Argentine Restaurant, Bakery, and Pizzeria, who peacefully passed away earlier this year. BOCA Bakery was the location most used for the work presented in this dissertation. With Marco's passing, the bakery unfortunately no longer exists. To Marco: *Thank you for providing me with the best coffee and empanadas Seattle had to offer. Your warm greetings always made my day. Hasta que nos volvamos a encontrar, amigo.* I also thank Kitanda, Cafe Vivace, and Basecamp for providing me with the fuel (coffee) to write this dissertation.

This dissertation would not have been possible without the support I received from the CSE staff. I thank Elise deGoede Dorough, Joe Eckert, Les Sessoms, Tracy Erbeck, Olivia Banner, Elle Brown, Jan Cuny, Chloe Dolese Mandeville, Jason Howe, Hector Rodriguez, David Silcox-Rither, Aaron Timss, and Amber Cochran for always being there for me and going above and beyond to ensure that my disability needs were met and that my voice was heard. To you all, and especially Elise and Joe: *I don't think words* *could describe the impact you make on the lives of graduate students every single day, the impact that you have made on my life. You are, without a doubt, a national treasure.*

I am grateful to my family—my four ammas, my brother, my sister, my three uncles, and their respective spouses and children—for their immense support and encouragement throughout this journey. To all of you: *Thank you for always being patient with me, for all the love and care that I can feel despite being thousands of miles away.* I particularly thank Moien Khan, Rashid Baig, Yasmeen Memon, and Rana Kashif, without whom, I would have never gotten back to life after my spinal cord injury. To you all: *If I could be to someone what you all have been to me, I would consider my life complete.* And to my amma who now resides in heaven: *I'm sorry I couldn't say goodbye. I know you're looking after me from up above. I hope I can make you proud, one day.*

I am beyond thankful to my other family, without whom Seattle would have been as gloomy as it is advertised to be. I thank (in no particular order) Trung-Anh Nguyen, Nayon Park, Zeqiu "Ellen" Wu, Xuhai "Orson" Xu, Katherine Juarez, Yisu "Remy" Wang, Yadi Wang, Jialin Li, Venkatesh Potluri, Taylor Hunter, Phoebe Do, Yuren "Rock" Pang, Maheen Junaidy, Monica Cowdery, Nate Woods, Savanna Yee, Marissa Radensky, and all my friends in CSE, for keeping it real with me, for helping me become a better person, for keeping me sane, and for patiently waiting as I indulged myself in feeding my Instagram stories before they could feed themselves. Indeed, more important than the journey itself, is who we share it with, and I shared it with the best people I could have ever hoped for. To you all: *Although you all weren't the family that brought me into this world, you sure were the ones to keep me in it. This world is a beautiful place because you all are in it.* I thank my partner, Madelynne Zornes, for being my rock and the shoulder I have needed to lean on. To Madelynne: *I betrayed my Star Wars family and watched Star Trek for you. If that's not an indication that I hope to have a future with you, then I don't know what is. May the Force be with us.*

And finally, I am grateful to my cats, Zoey, Harry, and Gemma, for their *purr*fect feline support that kept me going every time my morale was at rock bottom. To Harry and Gemma: *I love and appreciate you more than any living thing on this planet. Please stop destroying the furniture*. To Zoey, a *fluffball* who relocated to cat heaven on June 23rd, 2022, and to whom this dissertation is dedicated to: *In the three cities we lived together, the seven years we spent together, I never knew what I would do without you. I still don't. Because you're still here. Thank you for being the light of my life.*

To **Zoey**

I miss you.

Chapter 1

Introduction

> *Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and* ***to seek, receive and impart information and ideas through any media and regardless of frontiers****.*
>
> —Article 19, Universal Declaration of Human Rights (Nations [2021])

Access to information is a fundamental right and necessity of all human beings, regardless of any factor, including their race, ethnicity, gender, sex, and ability levels. For disabled people[1], proper access to information, especially from digital content, has been an ongoing struggle, causing an even worsening digital divide. Such is particularly true for screen-reader users (e.g., blind individuals), who can not see the visual content and rely on their screen readers to access digital information. Milagros Costabel, a blind freelance journalist, elegantly describes her experience accessing digital information, "*Being blind in a digital age is as beautiful as it is difficult. It's beautiful, because all the content I can access today allows me to do things that would have been unthinkable not so long ago. It's difficult because I often hit informational dead ends due to inaccessible content, thus widening an accessibility gap that shouldn't be there*" (Costabel [2021]).

Among the digital content present online, data visualizations are a widely used medium to communicate essential insights about data. These visualizations assist users in making informed life decisions involving health (e.g., COVID-19 graphs), finances (e.g., stock market trends), and current events (e.g., polling

---

[1] Following the guidelines from my prior work (Sharif et al. [2022a]) and recommendations from the National Federation of the Blind (National Federation of the Blind [1940]), I hereafter use identity-first language to refer to disabled people and blind and low-vision (BLV) individuals. I mention specific diagnoses wherever necessary.

data). Due to the inherently visual nature of these visualizations, screen-reader users rely on visualization creators to incorporate adequate accessibility measures for them to consume information contained within these visualizations. Without these measures, screen readers announce online data visualizations as "object," "graphic," or "image," providing no meaningful information. Popular accessibility measures include providing alternative textual description ("alt-text") of the visualization, a non-speech audio representation of data (e.g., sonification), a non-textual representation of data (e.g., 3-D printing), and data tables.

When considering these measures, we may find that they comply with the Web Content Accessibility Guidelines (WCAG) (W3C [2021]) and are thus presumably sufficient for screen-reader users to extract information from online data visualizations. However, if we consider the limitations of these measures, then we find that these solutions do not provide screen-reader users equitable access to information. For example, an unfortunate reality is that visualization creators use the visualization title as alt-text, which does not include any information about the *data*. Even when the alt-text contains the information summary, it does not provide users access to individual data points. Similarly, sonification enables users to obtain holistic information about the data and may not be helpful in granular data consumption. Data tables are a possible remedy for this issue but can cause a significant cognitive overload, especially for large data sets, due to their linear traversal by screen readers. Similarly, 3-D printing is a plausible technique but may not be practical for spontaneous everyday web browsing.

Given the increasingly widespread usage of online data visualizations, enhancing the accessibility of these visualizations is a combined function of improving the experiences of screen-reader users and supporting visualization creators in making data visualizations accessible. Therefore, in this dissertation, I present a novel interaction technique that (1) improves information consumption for screen-reader users from simple and complex online data visualizations and (2) supports visualization creators by enhancing their knowledge of and minimizing their challenges in making data visualization accessible. This dissertation shall demonstrate the following thesis:

> *A multi-modal, customizable, and interactive JavaScript plug-in called "VoxLens" improves the experiences of screen-reader users in extracting information from simple and complex online data visualizations while also enhancing the knowledge of visualization creators to make online data visualizations accessible.*

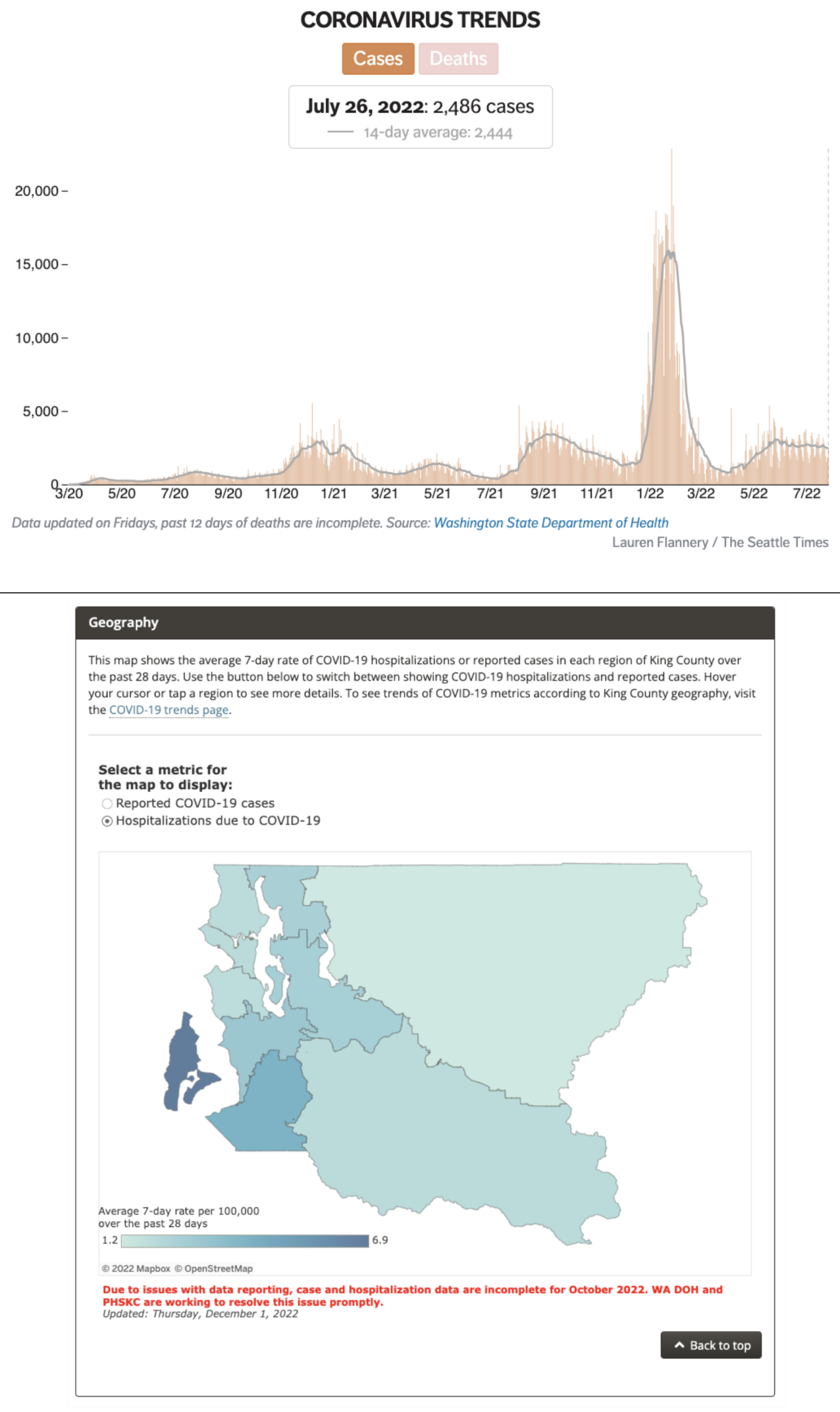
1.1 Scope

This dissertation focuses on improving the experiences of *screen-reader users* and *visualization creators* with *online* data visualizations. Screen-reader users are users who utilize a screen reader (e.g., JAWS (Scientific [1995]), NVDA (Access [2006]), or VoiceOver (Access [2006])) to read the contents of their computer screen; they may use a screen reader for permanent or temporary purposes. They might have complete or partial blindness, low vision, learning disabilities (such as Alexia), motion sensitivity, or vestibular hypersensitivity. As a large volume of related work on data visualization accessibility focuses on blind and low-vision (BLV) users, it is essential to note that BLV users constitute a subset of screen-reader users, speaking to the broader applicability of this work.

Visualization creators are individuals who create data visualizations; this term encompasses similar terms used in prior works, including “developers,” “practitioners,” “programmers,” and “designers.” Additionally, this dissertation concentrates on *online* data visualizations created using visualization libraries, such as D3 (Bostock et al. [2011]) and Google Charts (Developers [2014]). While this work recognizes future avenues for making data visualizations accessible more broadly, it is limited to visualizations on a web page; it does not include visualizations in other digital formats (e.g., PDFs) or external software, such as Microsoft Excel, Keynote, and Tableau.

1.2 Motivation

This dissertation presents an approach to improve data visualization accessibility for screen-reader users and visualization creators. While individual chapters of this dissertation include respective formative studies to shed further light on the motivation behind this work, as a whole, this dissertation is motivated by three primary factors. The first is the ubiquitous adoption of online data visualizations to convey essential information. The second is the inaccessibility of these visualizations. The third is the insufficiency of existing solutions to make these data visualizations accessible. These factors emphasize the disenfranchisement screen-reader users experience with consuming information from online data visualizations compared to non-screen-reader users. Therefore, the primary research goal is to reduce this disenfranchisement by improving the accessibility of online data visualizations.



Data updated on Fridays, past 12 days of deaths are incomplete. Source: Washington State Department of Health

Lauren Flannery / The Seattle Times

Geography

This map shows the average 7-day rate of COVID-19 hospitalizations or reported cases in each region of King County over the past 28 days. Use the button below to switch between showing COVID-19 hospitalizations and reported cases. Hover your cursor or tap a region to see more details. To see trends of COVID-19 metrics according to King County geography, visit the COVID-19 trends page.

Select a metric for the map to display:

○ Reported COVID-19 cases
● Hospitalizations due to COVID-19

Due to issues with data reporting, case and hospitalization data are incomplete for October 2022. WA DOH and PHSKC are working to resolve this issue promptly.
Updated: Thursday, December 1, 2022

Figure 1.1: Examples of data visualizations showing COVID-19 data. Image on top shows a visualization from The Seattle Times displaying daily COVID-19 cases in Seattle. Image on the bottom shows a choropleth map from King County, WA displaying hospitalizations due to COVID-19.

1.2.1 Wide Adoption of Online Data Visualizations

Online data visualizations have been increasingly assisting expert and non-expert users to communicate essential insights and explore, interact with, and extract meaningful information from simple and complex data (Lee et al. [2015]; Holland [2017]). They aid people in gathering information effectively and efficiently, taking advantage of the ability of the human mind to recognize and interpret visual patterns from large volumes of data (Marriott et al. [2021]). In addition, visualizations help users in making critical and informed life decisions for themselves and their families concerning health (*e.g.*, COVID-19 graphs), finances (*e.g.*, stock market trends), and the current events (*e.g.*, polling data), among other life domains (Huang et al. [2014]). For example, several news media, including The New York Times and The Seattle Times, and government websites used data visualizations on their websites to keep readers informed about COVID-19 cases and hospitalizations (see Figure 1.1). Additionally, recent work in politics (Zhao and Ye [2022]; Kubovics and Bielik [2021]), health (Zheng et al. [2021]; Grosjean et al. [2022]; Patrick and Junaini [2021]), finance (Tuarob et al. [2021]; Er and Sun [2021]), and business analytics (Biagi et al. [2022]; Zhang et al. [2021]) indicate the importance and wide adoption of data visualizations on the web.

Furthermore, data visualization research maintains a reputable standing as a research field, with several avenues dedicated to publishing visualization research, such as the IEEE Visualization Conference (VIS). It also showcases a significant presence at the ACM Conference on Human Factors in Computing Systems (CHI), with over 59% of publications appearing in the search results for "data visualization" in its 2023 proceedings. Similarly, the community for data visualization practitioners is ever-growing, evident through the membership of about 20,000 visualization creators in just one of the online communities called the "Data Visualization Society[2]." These statistics provide evidence of the widespread popularity of online data visualizations and the likelihood of this domain's growth in the future. As Few and Edge note, "*I expect that data visualization will continue for the next few years to pursue and mature those trends that have already begun. Dashboards, visual analytics, and even simple graphs will continue to develop and conform to best practices. I also have seen evidence that newer efforts are emerging that will soon develop into full-blown trends.*" (Few and Edge [2007]).

---

[2] https://www.datavisualizationsociety.org/

1.2.2 Online Data Visualizations are Inaccessible

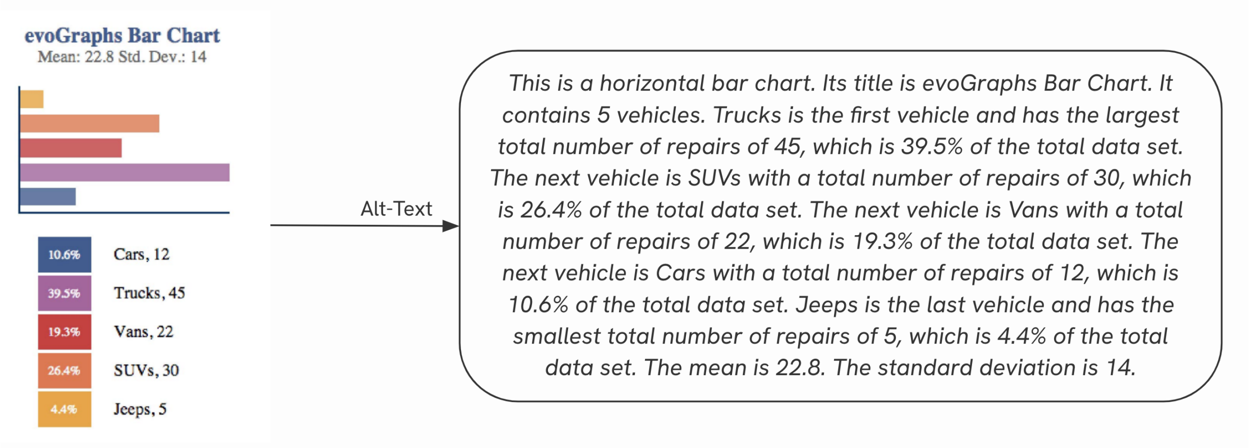
Online data visualizations offer paramount benefits to non-screen-reader users, enabling them to extract information quickly and effectively using visual means. Strecker states, "*Visual science has demonstrated that data visualizations are particularly effective in communicating or explaining data to an identified audience, if the visualizations are calibrated correctly to draw on the brain's ability to detect certain properties. If visualizations are properly designed they not only increase the speed at which data is comprehended, but can also increase the retention of data. Visual perception utilizes the eyes, a channel which has one of the largest 'bandwidths' to the brain of all our senses.*" (Strecker [2012]).

Unfortunately, this inherently visual nature of visualizations makes them inaccessible to screen-reader users, who may only be able to see part of the visualization or may not be able to see it at all. As Elmqvist notes, "*If you are blind, visualization is nearly impossible to use. And even if you can access the raw data, which is far from a given, the massive scale of many real-world datasets means that effective overview is beyond your reach without expertise in statistics and data science.*" (Elmqvist [2023]). Elmqvist's statement aligns with the findings presented in this dissertation, particularly those mentioned in Chapter 3.

Given the ubiquity of utilizing online data visualizations to convey information to users, inaccessible visualizations can cause detrimental disruptions in the lives of screen-reader users (Sharif et al. [2021]; Choi et al. [2019]; Zong et al. [2022]; Marriott et al. [2021]; Kim et al. [2021b]). For example, inaccessible COVID-19 graphs resulted in health concerns for the blind and low-vision community (Fan et al. [2023]; Praharaj et al. [2023]). Additionally, lack of access to the underlying information in visualizations exacerbates the disenfranchisement faced by screen-reader users (over 2.2 billion[3] users worldwide) who rely on visualization creators to make online visualizations accessible (Marriott et al. [2021]; Lee et al. [2021]; Sharif et al. [2021]; Davis [2002]). For example, inaccessible financial graphs can restrict them in making critical financial decisions. Findings from my research (discussed in Chapter 3) shows that due to inaccessible visualizations, screen-reader users extract information 61% less accurately and spend 211% more interaction time than non-screen-reader users (Sharif et al. [2021]). Furthermore, during the user studies, about 33% of the online data visualizations that the participants interacted with were completely indiscoverable by screen readers. This dissertation is motivated by the need to reduce this disenfranchisement.

---

[3] https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment



This is a horizontal bar chart. Its title is evoGraphs Bar Chart. It contains 5 vehicles. Trucks is the first vehicle and has the largest total number of repairs of 45, which is 39.5% of the total data set. The next vehicle is SUVs with a total number of repairs of 30, which is 26.4% of the total data set. The next vehicle is Vans with a total number of repairs of 22, which is 19.3% of the total data set. The next vehicle is Cars with a total number of repairs of 12, which is 10.6% of the total data set. Jeeps is the last vehicle and has the smallest total number of repairs of 5, which is 4.4% of the total data set. The mean is 22.8. The standard deviation is 14.

Figure 1.2: Autogenerated alt-text by `evoGraphs`, a jQuery plug-in to create web accessible graphs (Sharif and Forouraghi [2018]).

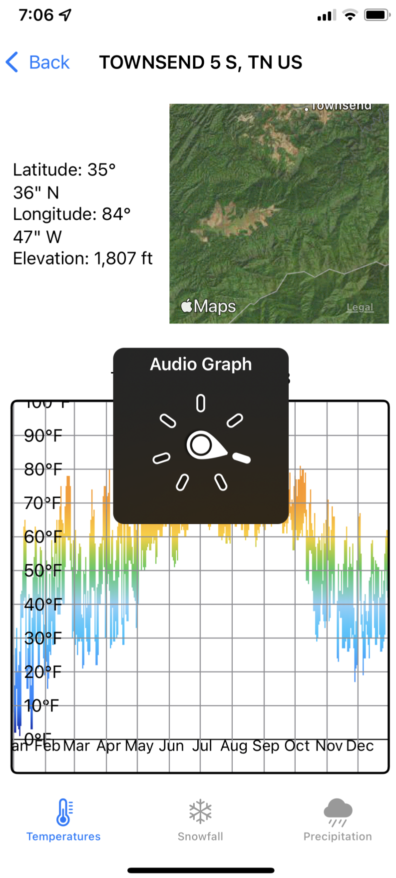
1.2.3 Existing Solutions are Insufficient

The inaccessibility of data visualizations is a widely recognized problem. Several researchers and practitioners have proposed solutions to this problem, yet screen-reader users experience disenfranchisement. As one of the participants in my user studies said, "*The progress is too slow in making things accessible. The data visualization space is progressing so rapidly that for anything that's being done, the gap is just being widened and it just keeps widening.*" These solutions mainly focus on four modalities: (1) Alternative textual description, (2) Sonification, (3) Tabular representation, and (4) 3-D printing. While solutions that utilize these modalities to make online data visualizations accessible are discussed in detail in Chapter 2, an overview of the modalities is presented below.

1.2.3.1 Alternative Textual Description

Alternative textual description ("alt-text") is the most commonly used technique to make digital graphics accessible, particularly online data visualizations. Traditionally, developers are responsible for providing the alt-text by describing the underlying information in the visualization in 125–150 words (University [2022]). However, alt-text is often missing, insufficient, or excessive, exacerbating the disenfranchisement screen-reader users experience consuming information from online data visualizations. Therefore, researchers and practitioners have explored autogenerating alt-text to reduce the burden on developers. For example, my prior work autogenerated a summary using all data points during visualization creation (Sharif and Forouraghi [2018]) (see Figure 1.2). Furthermore, with the recent advent of tools powered by large language models, such as ChatGPT[4], users can now obtain a summary of the visualization.

While alt-text is a plausible technique that provides users with a holistic overview of the information, it does not enable users to extract or compare individual data points, an affordance offered to non-screen-reader users from the get-go. Additionally, one may rightly question if a single alt-text satisfies the needs of *all* the screen-reader users, over 2.2 billion around the globe, considering every user is unique and has distinct preferences. Therefore, alt-text is limited in its offering to make data visualizations as accessible as they are to non-screen-reader users.



TOWNSEND 5 S, TN US

Latitude: 35° 36" N
Longitude: 84° 47" W
Elevation: 1,807 ft

Figure 1.3: A sample temperature chart that displays weather data for Tennessee, U.S. is shown with a VoiceOver rotor portraying the option of audio graphs.

---

[4] https://chat.openai.com/

1.2.3.2 Sonification

Similar to alt-text, sonification is also a widely utilized modality to make data visualizations accessible to screen-reader users. Sonification (or "audio graphs"[5]) is the non-speech audio representation of the data. A traditional usage of sonification can be imagined as a discrete or continuous sound that alters its pitch or frequency based on the data (e.g., higher frequency to represent higher values). Another innovative example of sonification is data representation using bird or waterfall sounds (Hoque et al. [2023]). Sonification's wide adoption is further evident from its application in commercial solutions. For example, since its 2021 operating system releases (Apple [2021b]), Apple has used sonification as a standard measure across all its platforms (Apple watch, Keynote, etc.) to make data visualizations accessible (see Figure 1.3).

However, sonification may not enable users to extract individual data points from a visualization. Additionally, sonification could lead to dangerous misinterpretations without properly conveying supplementary information, such as axis labels. For example, in a graph portraying stock market trends, the sonified response would be the opposite for axis values sorted alphabetically in an ascending order versus descending order. Similarly, sonification may be challenging to comprehend when used to represent complex data, especially data that does not show clear variation between data points.

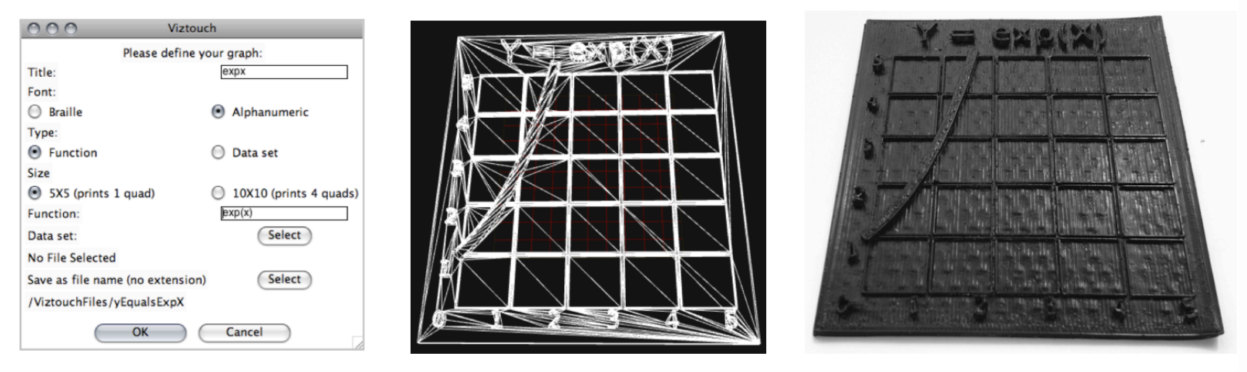
1.2.3.3 Tabular Representation

Contrary to alt-text and sonification, two common strategies discussed above that enable users to explore the data holistically, tabular representation provides screen-reader users with individual data points. Users can use the tables to perform their self-exploration and analysis of the data in a drilled-down manner. Visualization creators can present the data tables on web pages as standalone elements. They can also attach the tables with the visualization element and visually hide them using the "aria-hidden" attribute to support screen-reader-user-only access. For example, Google Charts (Developers [2014]) automatically generates and appends a visually hidden table from the underlying data for screen-reader users.

While data tables provide granular access to data, they can increase the cognitive load on users due to the linear traversing characteristic of screen readers. Such is especially true for data with large cardinalities. For example, for a data set comprising a hundred data points, a screen-reader user must go through

---

[5] I use "sonification" synonymously with "audio graphs," limiting its domain of usage to data visualizations.



Viztouch

Please define your graph:

Title: expx

Font:

○ Braille ● Alphanumeric

Type:

● Function ○ Data set

Size

● 5X5 (prints 1 quad) ○ 10X10 (prints 4 quads)

Function: exp(x)

Data set: Select

No File Selected

Save as file name (no extension) Select

/ViztouchFiles/yEqualsExpX

Figure 1.4: Pipeline to generate a 3-D printed visualization of $Y = exp(X)$ using `VizTouch` (Brown and Hurst [2012]). Image on the left shows an interface to provide data. Image in the middle shows the wireframe view of the 3-D model. Image on the right shows the 3-D printed output.

each data point and mentally compute data to extract information as simple as the minimum or maximum. Consequently, screen-reader users may experience considerably higher interaction times with online data visualizations than non-screen-reader users.

1.2.3.4 3-D Printing

3-D printing enables the production of *physical* visualizations that facilitate blind and low-vision users to feel and explore static visualizations in their graphical state (see Figure 1.4). In opposition to alt-text, sonification, and data tables, 3-D printed visualizations are not a practical solution for spontaneous online web browsing based on the assumption that users will likely not have access to a 3-D printer at all times. Even if they have access, obtaining the underlying data and feeding it into 3-D printing software to generate the visualization may not be feasible for everyday web browsing. Additionally, they require extra hardware (3-D printer), which can be costly for some users. As this dissertation focuses on *online* data visualizations, a domain out of scope for 3-D printing, discussions on 3-D printed visualizations are limited to only their recognition as an alternate modality to make data visualizations accessible to screen-reader users.

1.3 Research Approach

This research follows the user-centered iterative design protocol in the development of a novel interaction technique for screen-reader users to interact with online data visualizations. Specifically, I conducted formative studies to motivate the creation of and inform the design decisions for artifacts for screen-reader users and visualization creators, including the VoxLens versions. These formative studies include pilot studies, Wizard-of-Oz studies, surveys, interviews, and workshops.

Artifacts are built using iterative prototyping and continuous user feedback at each development stage. These artifacts are then evaluated using appropriate qualitative, quantitative, and subjective methods, including contextual and in-depth interviews, task-based user experiments, usability studies, diary studies, and longitudinal studies. In particular, two types of comparisons are made. First, the performance of screen-reader users is evaluated against the baseline of non-screen-reader users' performance to assess the access gap at each stage of this work. Second, the performance of VoxLens is evaluated using the baseline of traditional visualizations without VoxLens. When possible, follow-up interviews are conducted for further assessment and identification of areas of improvement.

Overall, this research uses ability-based design (Wobbrock et al. [2018, 2011b]) as guidance to improve screen-reader users' experiences with online data visualizations and enhance the knowledge of visualization creators to make online data visualizations accessible.

1.4 Dissertation Structure

This chapter has presented the potential value of a multi-modal, customizable, and interactive plug-in to improve the accessibility of online data visualizations. The subsequent chapters are structured as follows:

Related Work Chapter 2	Describes prior work on the accessibility of online data visualizations, interaction experiences of screen-reader users with online data visualizations, implementation experiences of visualization creators with data visualization accessibility, and topics relevant to specific chapters in this dissertation.

Formative Research Chapter 3	Discusses formative work to understand the challenges of screen-reader users with online data visualizations compared to non-screen-reader users, the information they commonly seek, and the techniques and strategies that could improve their interaction experiences with online data visualizations.
VoxLens Design Chapter 4	Introduces and illustrates the design and implementation of VoxLens, an open-source multi-modal JavaScript plug-in that improves the accessibility of online data visualizations.
Supporting Screen-Reader Users Chapter 5–9	Describes the formative work, design, and evaluation of VoxLens iterations and VoxEx—a system that enables screen-reader users to customize their information consumption, developed to improve the experiences of screen-reader users with online data visualizations. Chapter 5 discusses work to support holistic information extraction from online data visualizations by screen-reader users. Chapter 6 introduces work to improve data sonification as an accessibility measure to make online data visualizations accessible for screen-reader users. Chapter 7 presents work to support drilled-down information extraction from simple and complex online data visualizations by screen-reader users. Chapter 8 illustrates work to convey information on data uncertainty from online data visualizations to screen-reader users. Chapter 9 presents work to enable the customization of information consumption from online data visualization by screen-reader users.

Supporting Visualization Creators Chapter 10	Presents the formative work, design, and evaluation of the VoxLens iteration and five interventions implemented to enhance the experiences and knowledge of visualization creators in making online data visualizations accessible to screen-reader users.
Conclusion Chapter 11	Provides reflections and insights, major empirical findings, contributions of this dissertation, future avenues for research, and final remarks.

Chapter 2

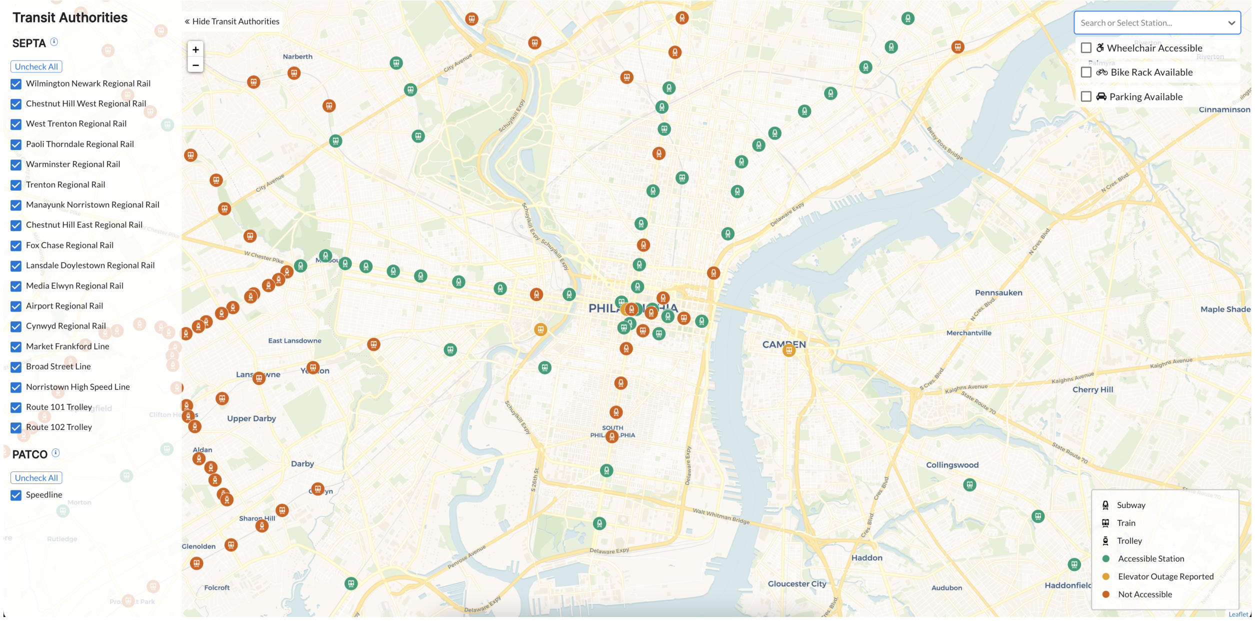
Related Work

In this chapter, I present previous research on the accessibility of online data visualizations that motivated and guided my work. I highlight research works that have emphasized the need for accessible visualizations, provided recommendations and techniques to make online visualizations accessible, and contributed solutions to improve the accessibility of online visualizations. I also discuss the common modalities used to make online online data visualizations accessible to screen-reader users. Furthermore, I review works that have employed customization for screen-reader users to provide agency to these users in extracting information online. Finally, I discuss prior works that shed light on the challenges faced by visualization creators in making online data visualizations accessible to screen-reader users.

2.1 Interaction Experiences of Screen-Reader Users with Technology

Several research efforts have explored the interaction of screen-reader users with technology via user studies (Kane et al. [2008]; Billah et al. [2017]; Grussenmeyer et al. [2017]; Adams et al. [2013]; Abdolrahmani et al. [2018]), showing that screen-reader users encounter several challenges. For example, Kane *et al.* (Kane et al. [2008]) conducted interviews with eight screen-reader users, identifying usability issues with mobile devices, especially touchscreen smartphones (which were new at the time). Billah *et al.* (Billah et al. [2017]) conducted a study with 21 screen-reader users, reporting on the usage of screen readers in remote access scenarios. They utilized various screen-readers on different types of computers to measure screen-reader users' experiences with using computers at home, in the workplace, and at school.

Similarly, Schaadhardt *et al.* (Schaadhardt et al. [2021]) studied screen-readers users' experiences with 2-D digital artboards, such as those appearing in Microsoft PowerPoint and Adobe Illustrator. Their findings detail the challenges of using screen readers in 2-D environments compared to 1-D text streams. Their results are similar to the ones in this work, such as high cognitive loads and a need for better feedback. In my work, I conducted contextual interviews and longitudinal studies with screen-reader users to understand the holistic and drilled-down information they seek in simple and complex online data visualizations (Chapters 3, 5, and 7). Additionally, I reported on the pain points in their interactions and the techniques they prefer to improve the accessibility of online data visualizations.



Transit Authorities

SEPTA

Uncheck All

- ☑ Wilmington Newark Regional Rail
- ☑ Chestnut Hill West Regional Rail
- ☑ West Trenton Regional Rail
- ☑ Paoli Thorndale Regional Rail
- ☑ Warminster Regional Rail
- ☑ Trenton Regional Rail
- ☑ Manayunk Norristown Regional Rail
- ☑ Chestnut Hill East Regional Rail
- ☑ Fox Chase Regional Rail
- ☑ Lansdale Doylestown Regional Rail
- ☑ Media Elwyn Regional Rail
- ☑ Airport Regional Rail
- ☑ Cynwyd Regional Rail
- ☑ Market Frankford Line
- ☑ Broad Street Line
- ☑ Norristown High Speed Line
- ☑ Route 101 Trolley
- ☑ Route 102 Trolley

PATCO

Uncheck All

- ☑ Speedline

Search or Select Station...

- ☐ Wheelchair Accessible
- ☐ Bike Rack Available
- ☐ Parking Available

Figure 2.1: The *map page* of UNLOCKEDMAPS' user interface (UI) showing transit stations in Philadelphia. *Filtering options for station-related attributes* are displayed at the top-right corner, *filtering options for transit authorities* are at the left corner, and *map legend* is at the bottom-right corner of the page. Accessible stations are shown using icons with a green background, inaccessible stations with a red background, and stations experiencing an elevator outage with an orange background.

2.2 Accessibility of Visualizations Versus Visualizations of Accessibility

As a preliminary clarification, in this work, I consider "accessibility of online data visualizations" distinct from "visualizations of accessibility." The former follows the concepts of "web accessibility" (Brophy and

Craven [2007]; Lazar et al. [2004]; Paciello [2000]) and seeks to understand and improve the accessibility of online data visualizations for disabled people (e.g., screen-reader users) (Sharif et al. [2021]; Lundgard and Satyanarayan [2021]; Zong et al. [2022]; Elavsky et al. [2022]). In contrast, the latter displays information about accessibility, such as *urban infrastructure* (e.g., transit, sidewalks, roads) for disabled people (e.g., people with mobility disabilities) (Brock et al. [2018]; Froehlich et al. [2019]; Zeng et al. [2011]; Sharif et al. [2023a]; Bolten and Caspi [2019]).

One example of visualizations of accessibility is my work on UnlockedMaps (Sharif et al. [2023a]), an open-data map that visualizes the real-time elevator status (accessible, not accessible, experiencing an outage) of urban rail stations in six North American cities and displays nearby accessible restaurants and restrooms, assisting users in making informed decisions regarding their commute (see Figure 2.1). In this dissertation, I focus on my work on the "accessibility of online data visualizations" for screen-reader users.

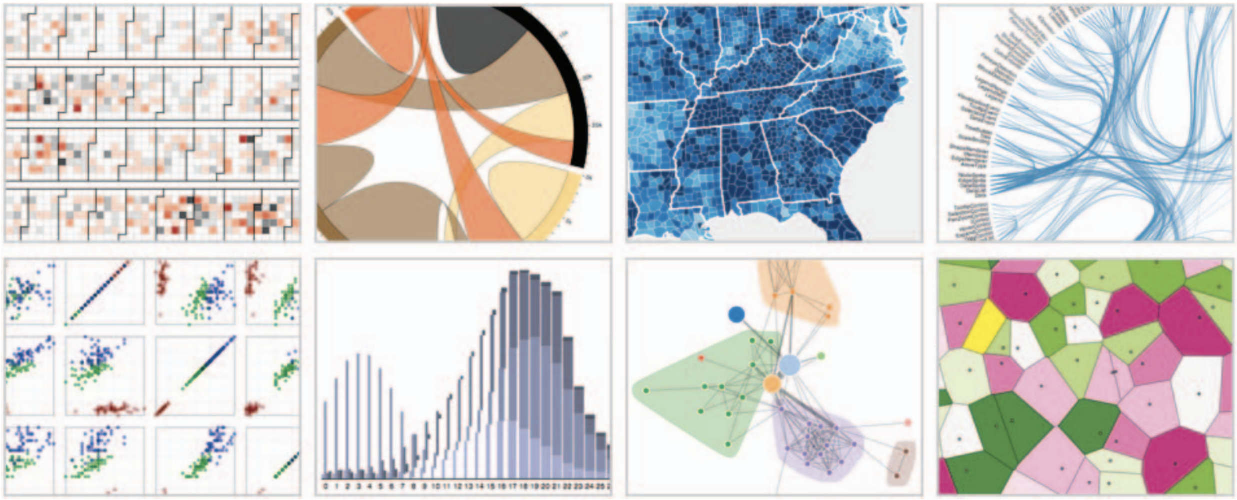


Figure 2.2: Bostock *et al.*'s interactive visualizations built with D3, running inside Google Chrome. From left to right: calendar view, chord diagram, choropleth map, hierarchical edge bundling, scatterplot matrix, grouped & stacked bars, force-directed graph clusters, Voronoi tessellation.

2.3 Existing JavaScript Data Visualization Libraries

Creators commonly use JavaScript visualization libraries to create online data visualizations. In my work, I utilized three visualization libraries to conduct user studies: (1) D3 (Bostock et al. [2011]), (2) Google

Charts (Developers [2014]), and (3) ChartJS (ChartJS [2015]). I selected these libraries based on the differences in their underlying implementations and accessibility measures, capturing unique interaction experiences of screen-reader users. Bostock *et al.* (Bostock et al. [2011]) developed D3—a powerful visualization library that uses web standards to generate graphs (see Figure 2.2). D3 uses Scalable Vector Graphics (SVG) (Dengler et al. [2011]) to create such visualizations, relying on the developers to provide adequate alternative text for screen-reader users to comprehend the information contained in the visualizations.

Google Charts (Developers [2014]) is a visualization tool widely used to create graphs that provides a visually hidden tabular representation of data for screen-reader users. While this approach allows screen-reader users to access the raw data, extracting information from tables can induce excessive user workloads. The workload is further exacerbated as data cardinality increases, forcing screen-reader users to memorize each data point to extract even rudimentary information such as minimum or maximum values. ChartJS (ChartJS [2015]) uses HTML Canvas to render the visualization as an image and relies on the developers to add alternative text ("alt-text") and Accessible Rich Internet Applications ("ARIA") attributes (W3C [2006]). In contrast to these approaches, VOXLENS introduces an alternate way for screen-reader users to obtain their desired information without relying on visualization creators and computing complex information through memorization of data.

2.4 Accessibility of Online Data Visualizations

Here, I review prior research that has highlighted the need for accessible data visualizations, provided recommendations to make online data visualizations accessible, and created solutions to improve the interaction experiences of screen-reader users with online data visualizations. Additionally, I discuss prior work on sonification, existing JavaScript libraries used to generate online visualizations, and identifying trends in the accessibility of online data visualizations.

2.4.1 Need for Accessible Data Visualizations

Several researchers have emphasized the importance of making data visualizations accessible by highlighting the inequities caused by inaccessible visualizations (Keilers et al. [2023]; Marriott et al. [2021]; Sharif et al. [2021]; Lundgard et al. [2019]; Lee et al. [2020]; Elmqvist [2023]; Konecki et al. [2018]). Keilers *et* *al.* (Keilers et al. [2023]) surveyed 45 blind and low-vision adults to explore data visualization accessibility on computers, phones, tablets, paper, and TVs, finding that insufficient accessibility practices significantly impact these users to access the underlying data in visualizations. Lundgard *et al.* (Lundgard et al. [2019]) identified the danger of perpetuating a vision-first approach that further marginalizes non-visual users when these users are not equal participants throughout the visualization design process. My work (Chapter 3) provided empirical evidence of this inequity by conducting mixed-methods studies with 36 screen-reader- and 36 non-screen-reader users. My results showed that due to the inaccessibility of online data visualizations, screen-reader users extract information 61% less accurately and spend 211% more time interacting than non-screen-reader users.

Marriott *et al.* (Marriott et al. [2021]) put forward a call-to-action for inclusive data visualizations, declaring the lack of access to visualizations and their underlying data a significant equity issue. Therefore, recognizing the critical need for accessible visualizations, researchers have provided recommendations to visualization creators (Sharif et al. [2021]; Lundgard et al. [2019]; Strantz [2021]; Carroll et al. [2013]; Oliveira [2013]; Elavsky et al. [2022]), including auto-generating alternative text, multi-modality, participatory design, and appropriate usage of Accessible Rich Internet Application (ARIA) attributes.

2.4.2 Trends and Gaps in Accessibility of Online Data Visualizations

Prior works have made survey contributions to reveal trends and identify gaps to understand the growth of accessibility as a field (Mack et al. [2021b]; Sharif et al. [2022b]). Similarly, researchers have provided a survey of accessible visualizations (Kim et al. [2021b]; Torres and Barwaldt [2019]; Lawrence and Lobben [2011]; Wabiński et al. [2021]; Hennig et al. [2017]). For example, Kim *et al.* (Kim et al. [2021b]) collected and analyzed papers published for the last 20 years on visualization accessibility, mapping a design space for accessible visualizations. They presented a preliminary model and identified future directions.

2.4.3 Accessible Modalities for Online Data Visualizations

Visualization creators use various modalities to make online data visualizations accessible to screen-reader users, such as alt-text, sonification, and data tables. I introduced these modalities in the previous chapter. In this section, I discuss prior work that utilized these modalities in turn below.

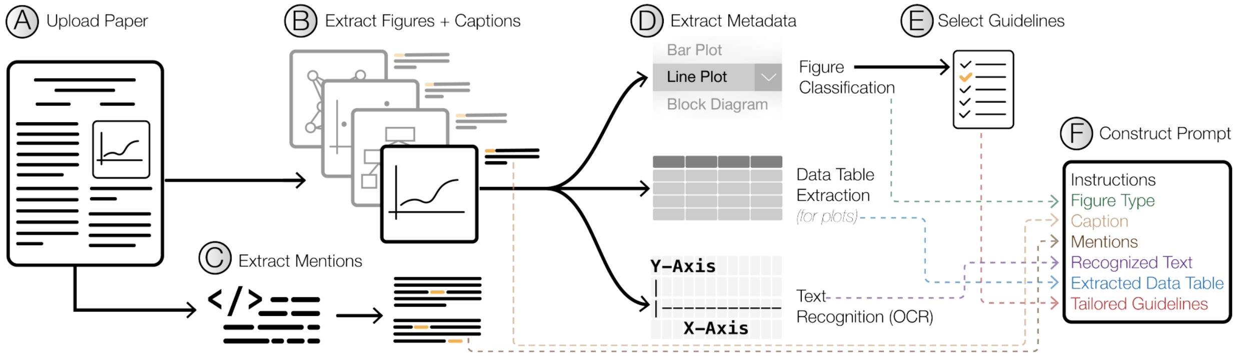


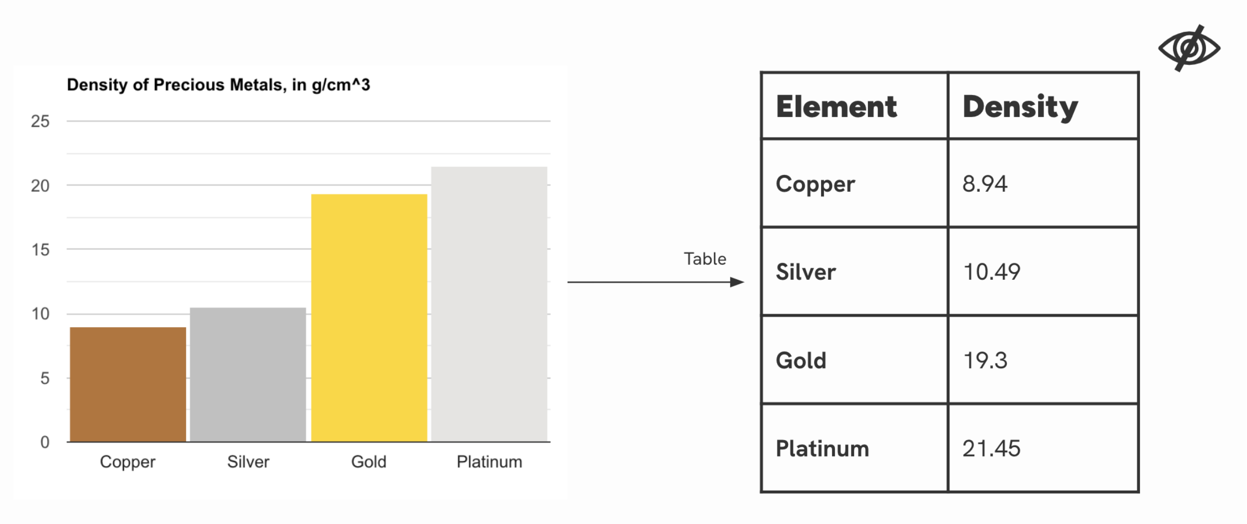
Figure 2.3: Pipeline for extracting information from figures, and using this information in a prompt to generate draft alt text and suggestions for enhancement. The author first (A) uploads a paper, from which (B) figures and their captions, and (C) mentions of each figure in the paper are extracted. Then, (D) the figure is classified, a data table is extracted if it is a plot, and the figure text is recognized. Finally, (E) based on the figure type, a set of guidelines are selected. (F) all of this information is put together with instructions into a prompt for the LLM to use in generating drafts and suggestions.

2.4.3.1 Alt-Text

Alternative text (“alt-text”) is one of the most widely used techniques to provide screen-reader users with a description of visual content, such as images (Mott et al. [2023]; McEwan and Weerts [2007]; Mack et al. [2021a]). Researchers have made contributions by assisting developers in auto-generating alt-text for images and visualizations (Sharif and Forouraghi [2018]; Mirri et al. [2017]; Kim et al. [2021a]; Lundgard and Satyanarayan [2021]; Singh et al. [2024]; Wu et al. [2017]). For example, Singh *et al.* (Singh et al. [2024]) developed `FigurA11y`, an interactive system that supplies authors of scientific papers with draft alt-text and offers suggestions based on figure and paper metadata (see Figure 2.3). Their findings from a user study with 14 authors indicate that their system improved authors’ effectiveness in producing descriptive alt-text in scientific papers. Similarly, Wu *et al.* (Wu et al. [2017]) created `Automatic Alt-Text`, an applied computer vision system, to identify faces, objects, and themes from photos to generate alt-text for screen-reader users on Facebook. They evaluated their system in a two-week field study with 9000 users, reporting an improvement in users’ engagement with and perceived usefulness of Facebook. However, unlike non-screen-reader users, screen-reader users can only consume information chosen for them. In this work, I address this disenfranchisement by providing these users the agency to customize the information they consume from online data visualizations (Chapter 9).

2.4.3.2 Sonification

Sonification, sometimes referred to as audio graphs, communicates information to screen-reader users using non-speech audio representations (for the Blind [2023]; Walker and Nees [2011]; Barrass and Kramer [1999]). Prior research has explored sonification to improve the experiences of screen-reader users with data visualizations (Siu et al. [2022]; Fan et al. [2022]; Holloway et al. [2022]; Austin and Sorge [2023]; Roy and Boppana [2022]) and developed open-source solutions (Highcharts [2009]; Sharif et al. [2022e]; Langston [2022]; for Digital Music Queen Mary University of London [2023]; Software [2023]). Hoque *et al.* (Hoque et al. [2023]) used natural sounds, such as waterfalls, to sonify data visualizations through their tool, `Susurrus`. They found that natural sounds help interpret multi-series data. I extended Wang *et al.*'s (Wang et al. [2022a]) work by investigating various oscillators and synthesizers and their effects on the pleasantness of sonified responses, resulting in an increased confidence of users in interpreting simple and complex data. I developed `Sonifier` (Sharif et al. [2022e]), an open-source JavaScript library that creates data sonification using these oscillators and synthesizers. Additionally, my work enables screen-reader users to specify their preferences for sonification, including the sound type and speed.



Element	Density
Copper	8.94
Silver	10.49
Gold	19.3
Platinum	21.45

Figure 2.4: A bar chart showing density of previous metals implemented using Google Charts (Developers [2014]). A tabular representation of the data is also shown with a "hidden" icon to represent the visually-hidden table that Google Charts appends to the visualization accessible to screen-reader users.

2.4.3.3 Data Tables

In contrast to alt-text and sonification, which only provide holistic information, data tables offer screen-reader users granular access to underlying data. Several prior works have discussed the utility of data tables as an accessibility technique (Kim et al. [2023]; Lai [2013]; Wang et al. [2022b]). For example, Ferres *et al.* (Ferres et al. [2013]) reported that screen-reader users relied on raw data to extract information from data visualizations in their user studies. These findings were particularly applicable to Google Charts (Developers [2014]), which automatically append an invisible data table to the document tree of the HTML page to provide screen-reader users with the underlying data (see Figure 2.4). However, data tables require linear access to data points and can cause cognitive overload for users, especially when the data cardinality is high (Sharif et al. [2021]; Laney et al. [2013]). To this end, in this work, I enable screen-reader users to specify their information consumption preferences. Visualization creators can use these preferences to present data tables and visualizations to screen-reader users in the manner these users prefer. For example, visualization creators can organize data tables based on the sorting preferences of screen-reader users.

2.4.4 Recommendations on Making Online Data Visualizations Accessible

To reduce the disenfranchisement screen-reader users face in extracting information from online data visualizations, researchers have recommended techniques to improve the accessibility of such visualizations (Lundgard et al. [2019]; Strantz [2021]; Carroll et al. [2013]; Oliveira [2013]). Lundgard *et al.* (Lundgard et al. [2019]) presented a set of sociotechnical considerations for accessible visualization designs, identifying participatory design and the usage of Accessible Rich Internet Application (ARIA) attributes as crucial elements in creating online data visualizations. Strantz (Strantz [2021]) provided best practices to create visually accessible data visualizations for developers. Their recommendations included using whitespace, creating contrast, maintaining size/scale, labeling the graph clearly, specifying textual descriptions and ARIA attributes, providing the overview and data context, and testing the visualizations through user studies and automated tools. Building on these prior works, my work (Chapter 3) recommended auto-generating alternative text (“alt-text”) to represent dynamic data and using multi-modality (e.g., tables, summaries, sonification) to enable screen-reader users to explore visualizations based on their preferences.

2.4.5 Tools to Make Data Visualizations Accessible

Following the recommendations put forward by prior work, several researchers have created tools to make data visualizations accessible to screen-reader users. These tools include auto-generating alternative text (Sharif and Forouraghi [2018]; Mirri et al. [2017]; Kim et al. [2021a]; Lundgard and Satyanarayan [2021]), sonification (Sharif et al. [2022f]; McGookin and Brewster [2006]; Zhao et al. [2008]; Ahmetovic et al. [2019b]; Siu et al. [2022]; Fan et al. [2022]; Holloway et al. [2022]; Austin and Sorge [2023]; Roy and Boppana [2022]), summarization (Kim and McCoy [2018]), tables (Developers [2014]), haptic graphs (Yu et al. [2000]; Van Scoy et al. [2005]), 3-D printing (Brown and Hurst [2012]; Shi et al. [2016]; Hurst and Kane [2013]), and multi-modality (Sharif et al. [2022f]; Thompson et al. [2023]; Blanco et al. [2022]). Most recently, Thompson *et al.* (Thompson et al. [2023]) introduced `Chart Reader`, an open-source prototype accessibility engine that renders accessible data visualizations. They created their tool following an iterative co-design study with 10 Microsoft employees and reported the evolution of the design of `Chart Reader` during this five-month study. Blanco *et al.* (Blanco et al. [2022]) built an open-source library called `Olli` that converts visualizations into a keyboard-navigable structure accessible to screen-reader users, which enables visualization creators to easily create accessible visualizations across various toolkits, including `Vega-Lite` (Satyanarayan et al. [2016]). In this work, I developed VOXLENS, an open-source JavaScript plug-in that enables screen-reader users to interact with online data visualizations using a multi-modal approach, assisting them in obtaining data through sonification, summary, and verbal querying (question and answer). I report a 164% and 50% improvement in screen-reader users' information extraction and interaction times, respectively, compared to conventional methods, through multiple user studies with over 100 users over time.

2.5 Sonification Solutions to Make Online Data Visualizations Accessible

In this section, I review prior work that has utilized sonification in practice and research. Additionally, I highlight research on the assessment of sonification as a technique to make data visualizations accessible for screen-reader users.

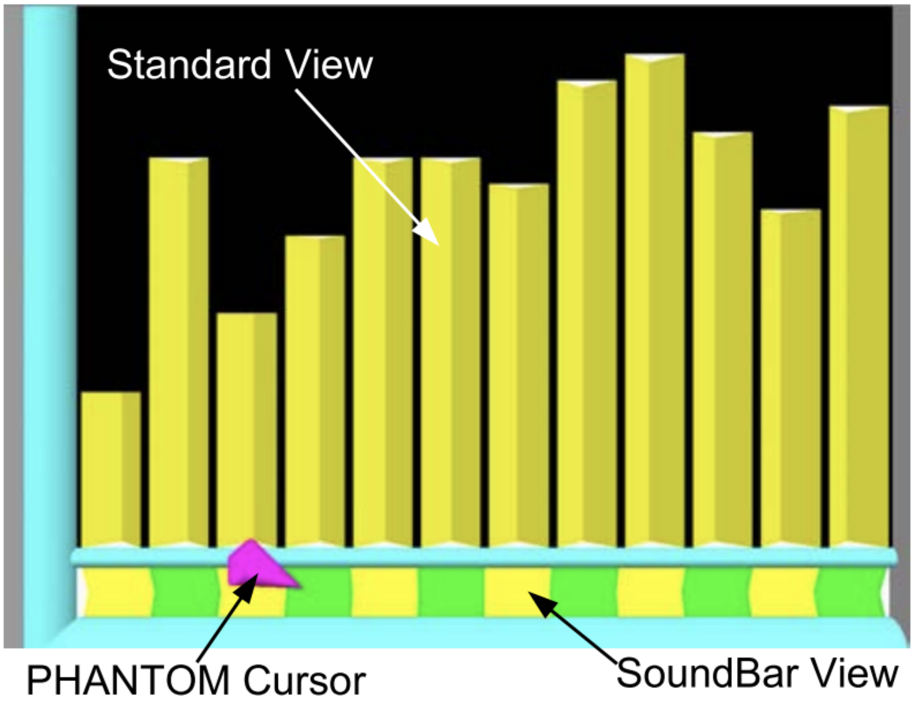


Figure 2.5: McGookin *et al.*'s screenshot of the SoundBar System. Bars are represented as recessed grooves, the SoundBar is located below the bars. When a segment of the SoundBar is touched with the PHANTOM (represented by the cone shaped object), a note proportional to the height of the bar immediately above is played.

2.5.1 Sonification in Practice

Developers and researchers have created several sonification solutions for screen-reader users; some of these solutions are open-source (Highcharts [2009]; Sharif et al. [2022e]; Langston [2022]; for Digital Music Queen Mary University of London [2023]; Media [2020]; LabSound [2021]), whereas some are proprietary (Software [2023]; Apple [2021b]; Wall et al. [2012]; Studios [2011]). However, only a few are suitable for *online* data visualizations. For example, Statistical Analysis Software (SAS) Graphing Calculator (Software [2023]) is a browser extension that enables screen-reader users to interact with online data visualizations using sonification but is limited to graphs created using the SAS software. Apple Audio Graphs (Apple [2021a]) provides an API for Apple application developers to construct an audible representation of the data in charts, giving screen-reader users access to valuable insights into data. Highcharts (Highcharts [2009]) is a proprietary JavaScript library that aids visualization creators in developing online data visualizations and offers free built-in sonification. Similarly, Sonifier (Sharif et al. [2022e]) and Chart2Music (Langston [2022]) are open-source JavaScript libraries that make online data visualizations accessible to screen-reader users by generating sonified responses using various oscillator waveforms and synthesizers. However, these libraries do not aid in determining when sonification may not be suitable for use.

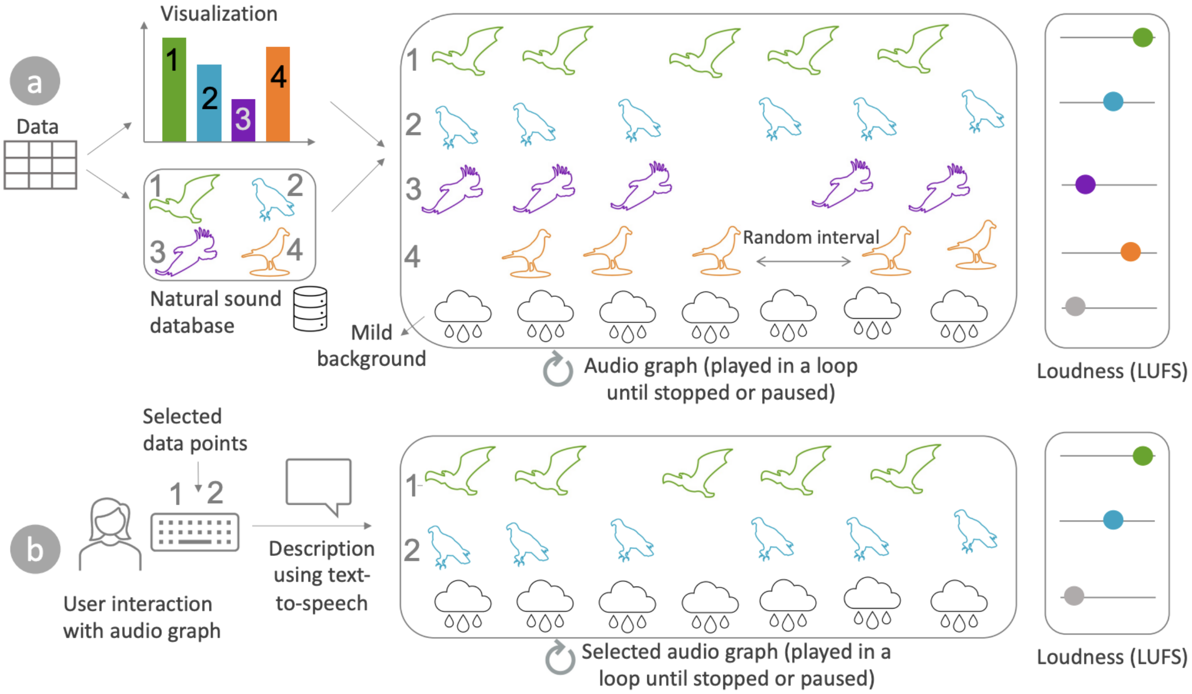
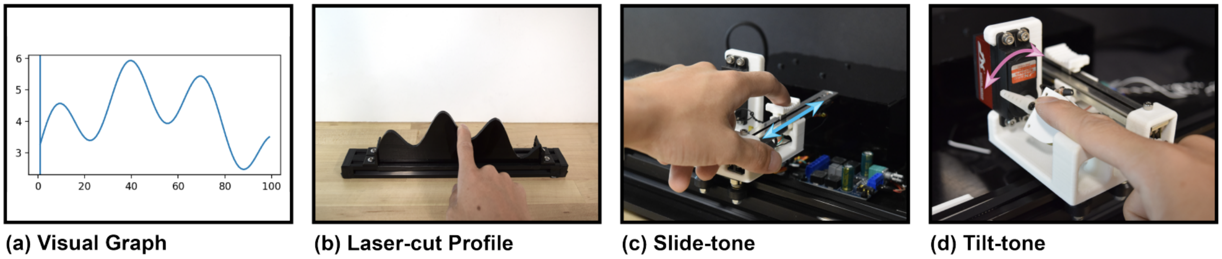


Figure 2.6: Hoque *et al.*'s sonification of a bar chart with Susurrus. (a) Their technique maps each bar to a natural sound drawn from an ambient theme (e.g., forest and birds). Using Loudness Levels Relative to Full Scale (LUFS), they convey the data values by setting the loudness of the sounds in decibels proportionately (i.e., height) to the bars. In this instance, they have mapped four bars in a bar chart to four bird sounds (e.g., robin, woodpecker, raven, and dove). They play the sounds together (i.e., in parallel) in a loop with random intervals and use a calm forest ambiance as background, thus making sonification of the bar chart similar to listening to bird sounds in the forest. (b) A user can interact with the audio graph using specific keys. For example, the user can select the first two bars using 1 and 2 number keys and listen to the corresponding sounds. With the selection, the user can also listen to the description of the selected data values using Text-to-Speech. The audio for this example is provided in the supplement.

2.5.2 Sonification in Research

Several research projects have explored sonification to improve the experiences of screen-reader users with data visualizations (Sharif et al. [2022f]; McGookin and Brewster [2006]; Zhao et al. [2008]; Ahmetovic et al. [2019b]; Siu et al. [2022]; Fan et al. [2022]; Holloway et al. [2022]; Austin and Sorge [2023]; Roy and Boppana [2022]). Most recently, Hoque *et al.* (Hoque et al. [2023]) developed Susurrus, which sonifies visualizations using natural sounds, such as birds singing in a forest (see Figure 2.6). Their findings show that natural sounds can benefit screen-reader users in interpreting data from visualizations, especially charts representing multiple categories.



(a) Visual Graph **(b) Laser-cut Profile** **(c) Slide-tone** **(d) Tilt-tone**

Figure 2.7: Fan *et al.*'s work informed by formative workshops that explored how haptic cues for fingerpad position and inclination support shape perception of data graphs (a-b). They introduce two refreshable, 1-DOF audio-haptic interfaces for data exploration. Slidetone (c) relies on fnger position with sonifcation, and Tilt-tone (d) relies on fingerpad contact inclination with sonification to provide shape feedback to users.

Ahmetovic *et al.* (Ahmetovic et al. [2019b]) developed AUDIOFUNCTIONS.WEB, which enables blind people to explore mathematical function graphs using sonification. They evaluated their system with 13 BLV participants, finding a high usability rating for their system's interaction modalities. Similarly, McGookin *et al.* (McGookin and Brewster [2006]) developed SoundBar, a system that allows blind users to gain a quick overview of bar graphs using musical tones (see Figure 2.5). Fan *et al.* (Fan et al. [2022]) built two audio-haptic interfaces that provide shape feedback to blind and low-vision users using sonification, reporting an increased appreciation for shape information from their users (see Figure 2.7.

At least one of the following is true for all of these systems: (1) they are either proprietary or limited to their respective commercial products (e.g., (Highcharts [2009])); (2) they are either standalone hardware or software applications (e.g., (Ahmetovic et al. [2019b])); (3) they require installation of either extra hardware or software (e.g., (McGookin and Brewster [2006])); or (4) they are incompatible with existing

JavaScript libraries (e.g., (Apple [2021a])). Thus, I developed SONIFIER (Sharif et al. [2022f])—an open-source JavaScript library that generates a sonified response from two-dimensional data. Additionally, I built on prior work by Wang *et al.* (Wang et al. [2022a]) and sought to examine and improve the usability and user-friendliness of sonified responses generated from online data visualizations created using JavaScript libraries (Sharif et al. [2022c]).

2.5.3 Assessment of Sonification

Prior research has assessed the usability of sonified responses (Presti et al. [2021]; Ahmetovic et al. [2019a]; Gerino et al. [2015]; Sharif et al. [2022c]; Wang et al. [2022a]). Most recently, Wang *et al.* (Wang et al. [2022a]) examined the impact of various auditory channels (e.g., pitch and volume) on users' perception of data visualizations. I extended their work by investigating the effects of various oscillator waveforms and synthesizers on the pleasantness and confidence of users in interpreting simple and complex sonified responses (discussed in Chapter 6).

In addition to assessing the usability of sonification, researchers have also used sonification as a baseline to examine the utility of multi-modal solutions. For example, Siu *et al.* (Siu et al. [2022]) investigated the usefulness of audio data narratives compared to a standard sonification representation. Their results suggest that audio data narratives help users gain a more complete gist of the data. Similarly, Chundury *et al.* (Chundury et al. [2022]) reported that their blind and low-vision participants preferred a combination of sound and touch to interpret data visualizations compared to using only auditory feedback.

In contrast to prior work, this work contributes to the existing accessibility and visualization literature by investigating the experiences of screen-reader users with sonification as a standalone technique to interpret data from visualizations. My research (see Chapter 6) provides empirical findings from a large-scale need-finding survey and an interview study with 106 and 12 screen-reader users, respectively. Furthermore, this work utilizes these findings to build a decision tree as a recommendation aid to assist visualization creators in using sonification to make data visualizations accessible to screen-reader users.

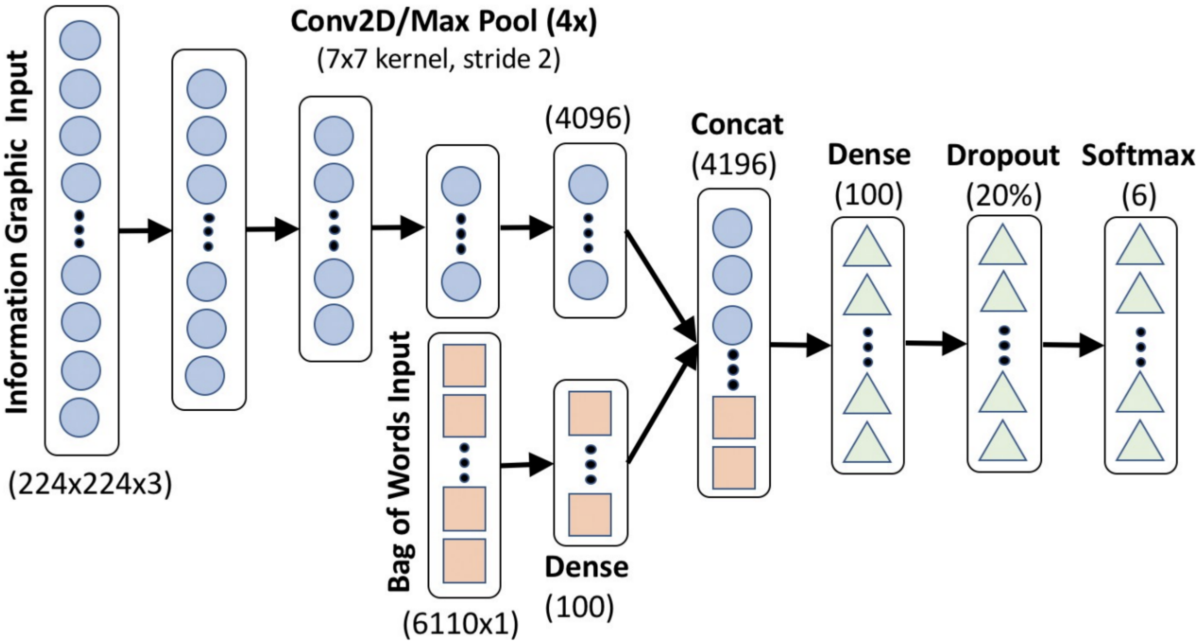


Figure 2.8: Kim *et al.*'s illustration of their multimodal model used for intended message classification. The model takes in a 224x224x3 pixel information graphic and a bag of words (BOW) input from the OCR extraction. The bag of words vector size is equal to the dictionary size, e.g. 6110. The visual input is fed through a convolutional neural network (represented by blue circle nodes). Notice that the visual network consists of four layers shrinking through the pooling operation. The BOW is bottlenecked through a dense, fully connected layer (represented by orange square nodes). Both modality streams are concatenated into a 4196 dimension vector. The joint embedding (represented by green triangle nodes) is bottlenecked again through a dense, fully connected layer with 20% dropout added for regularization. The final softmax layer has six outputs, corresponding to the six classes.

2.6 Multi-Modal Solutions for Improving Data Visualization Accessibility

Several researchers have developed solutions employing strategies identified in these recommendations to enhance the experiences of screen-reader users in extracting information from online data visualizations. These solutions include auto-generation of alt-text (Sharif and Forouraghi [2018]; Mirri et al. [2017]; Kim et al. [2021a]), sonification (Flowers et al. [1997]; McGookin and Brewster [2006]; Zhao et al. [2008]; Brown et al. [2003]; Highcharts [2009]; Apple [2021a]; Ahmetovic et al. [2019b]), data summarization (Kim and McCoy [2018]), keyboard navigation (Zong et al. [2022]), tabular representations of the data (Developers [2014]), and multi-modality (Kim and McCoy [2018]; Yu and Brewster [2002]).

Kim *et al.* (Kim and McCoy [2018]) generated summarization text displaying the high-level information from image-based line graphs using a multi-modal deep learning framework (see Figure 2.8). They used 1000 line graphs to train their model and conducted experiments with 100 blind and low-vision (BLV) participants to assess their model's performance. Their results indicated that their multi-modal model performed better than any modality alone, including the average human annotator. Yu *et al.* (Yu and Brewster [2002]) compared a multi-modal data visualization system and traditional tactile diagrams, measuring the accuracy of information extracted and interaction times. They found that the multi-modal approach improved the accuracy of information extraction from graphs. Similarly, Brewster (Brewster [2002]) employed the same dependent variables to compare the performance of speech and pitch sound graphs for screen-reader users, finding that non-speech sound and haptics can significantly improve interaction with visualizations.

However, these solutions focus on simple graphs (e.g., single-series two-dimensional graphs) and the extraction of holistic information (e.g., data summary or trend information). Therefore, in contrast, recognizing the utility and benefits of multi-modality, I developed VoxLens (Sharif et al. [2022f, 2023b]), an open-source JavaScript plug-in that enables screen-reader users to interact with and extract information online data visualizations using a multi-modal approach. VoxLens supports three modes: (1) *Question-and-Answer* mode, (2) *Summary* mode, and (3) *Sonification* mode. These modes enable screen-reader users to extract granular information from simple and complex online data visualizations. Additionally, prior studies have only explored the performance of screen-reader users. My work is the first empirical work to employ the same dependent variables (accuracy of extracted information and interaction times) to assess the *performance difference* between screen-reader and non-screen-reader users.

2.6.1 User Agency for Screen-Reader Users

Prior work in human-computer interaction has defined user agency as "a user's inherent capacity to form goals and intentionally take action to achieve those goals" (Kim et al. [2011]; Toivonen and Lelli [2024]), and reported that it is "central in managing technostress" (Dillon [2002]; Toivonen and Lelli [2024]). Researchers have widely discussed the importance of user agency for disabled people (Lakhani et al. [2018]; Clinkenbeard [2020]; Dai et al. [2023]), including screen-reader users (Manikoth [2016]; Barter [2023]; Jones et al. [2023]; Herskovitz et al. [2023]). For example, Barter (Barter [2023]) discussed the limitations of blind and low-vision users' agency in accessing information equitably compared to their sighted peers.

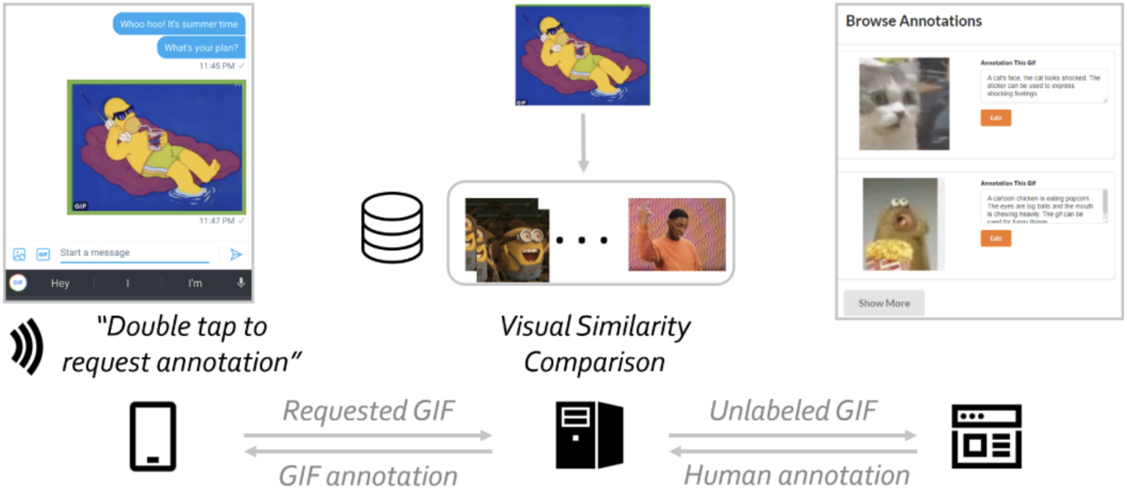


Figure 2.9: Zhang *et al.*'s components of Ga11y. The user requests the GIF annotation on the mobile client via the screen reader, and the requested GIF is searched for in the human annotation database on the server. If there is a visually similar GIF with a human annotation, that annotation will be returned; otherwise, a machine-generated annotation is returned, and the unlabeled GIF is then displayed in the Web-based human annotation interface. Once the GIF is annotated by volunteers on the website, the annotation is updated in the server's database for future retrieval.

They provided design recommendations for developers and creators of information and communication technologies to support equal agency for these users.

Researchers have also explored user agency through customization in their solutions (Thompson et al. [2023]; Kulkarni et al. [2016]; Jayant et al. [2011]; Stephanidis et al. [1998]; Wobbrock et al. [2018]; Ahmetovic et al. [2019c]). Several researchers have employed customization to build solutions for screen-reader users, including blind and low-vision (BLV) individuals (Zhang et al. [2022]; Ferretti et al. [2016]; Zhang et al. [2017]; Lee and Ashok [2020]; Kacorri [2017]). Zhang *et al.* (Zhang et al. [2022]) developed GA11Y that creates annotations for animated GIF images using machine intelligence and crowdsourcing to improve the accessibility of these images for BLV users. They assessed their annotation interfaces and conducted a multi-stage evaluation with 12 BLV participants from the United States and China, receiving positive feedback (see Figure 2.9). However, these customization options are limited to the functionality offered by their respective solutions. For example, Thompson *et al.* (Thompson et al. [2023]) developed a web-based accessibility engine called `ChartReader` and implemented customizations in its "filter" feature to minimize the auditory and processing loads on the user (see Figure 2.10).

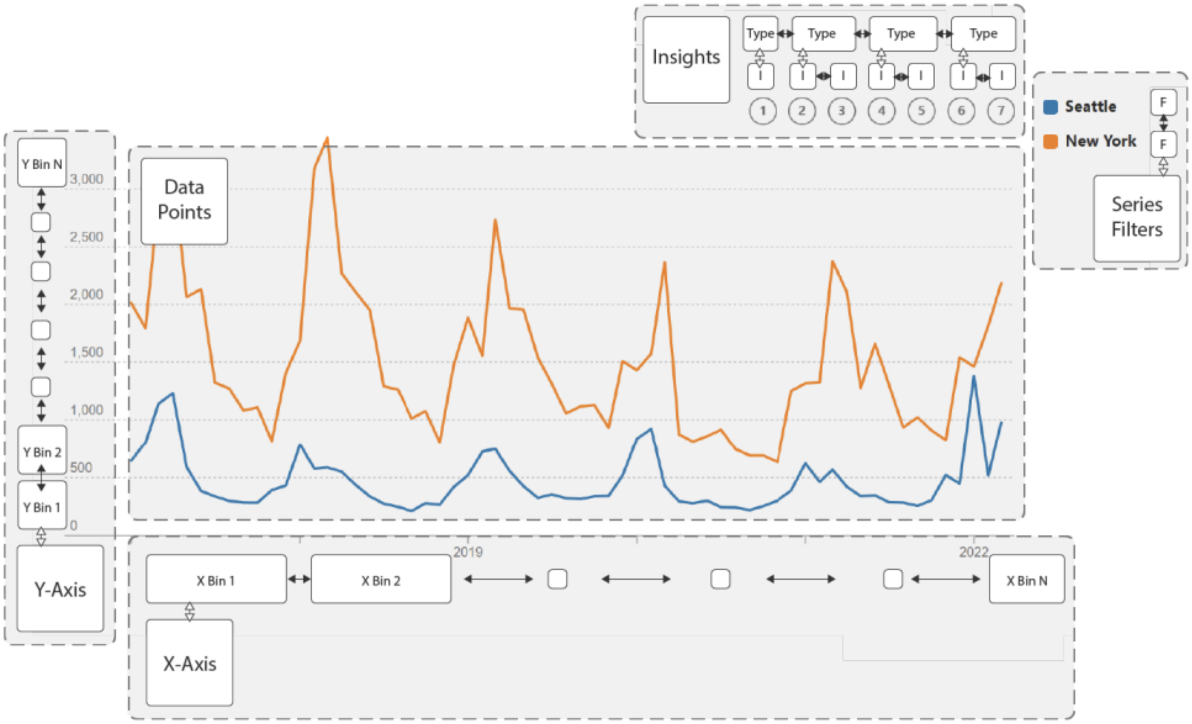
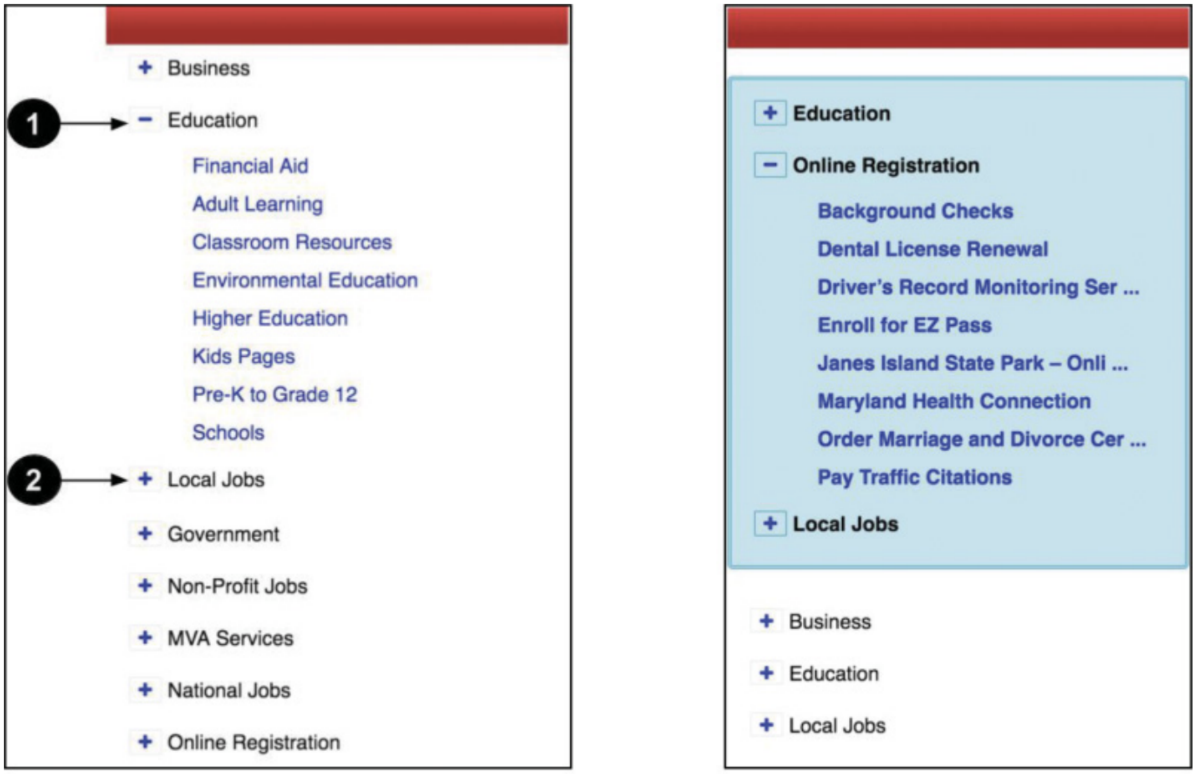


Figure 2.10: Thompson *et al.*'s five regions in charts rendered using Chart Reader to support accessible visualization experience. Data Insights, with sub-regions for each insight type and insight, further subdivided for each individual insight. X-Axis, with sub-regions for each bin along the axis. Y-Axis, with sub-regions for each bin along the axis. Data Points region. Filters, with sub-regions for each individual series.

Ferretti *et al.* (Ferretti et al. [2016]) recognized the benefits of personalized web interfaces for disabled people and created a system that used web intelligence and reinforcement learning to adjust web elements automatically based on user preferences. They evaluated their system through real and virtual longitudinal user studies, finding their solution feasible for disabled people. Zhang *et al.* (Zhang et al. [2017]) built a Personalized Assistive Web (PAW) interface that aims to improve skimming in mobile web browsing for BLV users through hierarchical outline view and personalization adaptations (see Figure 2.11). They evaluated their interface with 21 blind and 34 sighted participants, finding strong evidence for the positive impacts of customization on performance. In contrast, my work is the first to provide a centralized interface for screen-reader users to specify their preferences that visualization creators can utilize through the browser's session storage to create a customized experience for these users. Adopting this system places a minimal



(a) hierarchical outline view (b) A personalized outline view

Figure 2.11: Zhang *et al.*'s interfaces of PAW. (1) Select "–" to collapse an expanded node. (2) Select "+" to expand a node.

development burden on visualization creators and does not demand additional installation.

2.7 Experiences of Visualization Creators With Accessible Visualizations

Several researchers have recognized the obstacles to making data visualizations accessible (Singh et al. [2023]; Strantz [2021]; Plaisant [2004]; Kim et al. [2021b]; Wu et al. [2021]). However, to my knowledge, only a few have studied challenges that visualization creators face in making data visualizations accessible to screen-reader users (Trewin et al. [2010]; Joyner et al. [2022]). Most relevant to this work is the exploration by Joyner *et al.* (Joyner et al. [2022]), in which they surveyed 144 developers and conducted follow-up interviews with 10 selected respondents to understand the rationale and context behind the design choices of visualization designers. Their findings provide insight into visualization designers' challenges, knowledge, and prioritization of making their data visualizations accessible.

In contrast, my work is the first to identify educational and technological interventions to *reduce* the challenges visualization creators experience with making online data visualizations accessible while additionally shedding light on their challenges with data visualization accessibility. Furthermore, I complement Joyner *et al.* (Joyner et al. [2022])'s work by quantitatively analyzing the survey respondents' ratings of their accessibility challenges, knowledge, and prioritization. I also provide the results from the task-based user study that I conducted to assess and validate my findings.

2.8 Role-Playing Methodology

Role-playing is a technique used to assign participants particular roles to gather insights into their actions, reactions, and interactions with hypothetical scenarios or systems (Lewis-Beck et al. [2003]; Buchenau and Suri [2000]). Prior works have discussed the benefits of using role-playing methodologies to conduct user studies (Aranda et al. [2015]; Miller [1972]; Alexander and Scriven [1977]; Greenberg [1967]; Cooper [1976]; Svanaes and Seland [2004]). Researchers have broadly adopted this methodology in human-computer interaction (Boess [2008]; Lui [2012]; Colella et al. [1998]). However, the use of role-playing in accessibility research has been minimal. In this work, I utilized role-playing to understand the interactions of screen-reader users with online data visualizations (Chapter 7). Specifically, I asked participants to adopt the roles of "explorer," "teacher," and "news reporter," each of which imbued participants with different perspectives and elicited different responses from them.

Chapter 3

Understanding Screen-Reader Users' Experiences With Online Data Visualizations*

This chapter describes two empirical studies I conducted to understand and assess the challenges screen-reader users experience with online data visualizations. My objective from these two studies was to shed light on the experiences of screen-reader users with online data visualizations compared to non-screen-reader users, the information they commonly seek, and the techniques and strategies that could improve their interaction experiences with online data visualizations. It is important to note that my evaluation is of the online data visualizations (the technology) and not the cognitive abilities of people who use screen readers to interpret data. I discuss additional research on the experiences of screen-reader users with online data visualization in subsequent chapters of this dissertation.

3.1 Motivation

Online data visualizations are widely used to communicate essential insights, assisting users to explore, interact with, and extract meaningful information from complex data (Lee et al. [2015]). These insights

---

*Parts of this chapter are adapted from Sharif et al. [2021].

help people make critical and informed decisions for themselves and their families concerning health (*e.g.*, COVID-19), finances (*e.g.*, stock trends), and the current events (*e.g.*, polling data), among other vital life domains (Huang et al. [2014]).

The wide adoption of online data visualizations for information uptake, learning, and decision-making means that those unable to access the data presented in these visualizations may be disadvantaged. In particular, screen-reader users[1] must extract the information contained within data visualizations in alternative ways. However, no prior work has provided empirical studies of whether and to what degree screen-reader users are disenfranchised due to lack of access to information in online data visualizations.

Prior work has explored automatically generating alternative text from data visualizations summarizing commonly used statistics (Sharif and Forouraghi [2018]; Mirri et al. [2017]) as well as providing screen-reader users with alternative mediums for interacting with digital visualizations (such as sonification (Flowers et al. [1997]; McGookin and Brewster [2006]; Zhao et al. [2008]), haptic graphs (Yu et al. [2000]; Van Scoy et al. [2005]), and 3-D printing (Brown and Hurst [2012]; Shi et al. [2016]; Hurst and Kane [2013])). However, such alternative mediums require auxiliary resources and are not practical for daily web browsing by screen-reader users. To my knowledge, no work has evaluated the needs and performance of screen-reader users when interacting with online visualizations.

In this chapter, I present two empirical studies I conducted to understand and assess screen-reader users' challenges with online data visualizations compared to non-screen-reader users. The results show that online data visualizations were often undiscoverable to screen readers. In cases when screen readers recognized these visualizations, the users seemingly endured an excessive workload burden to extract information from them. Additionally, my findings show that the inaccessibility of online visualizations causes screen-reader users to extract information 61% less accurately and to spend 211% more time interacting with online data visualizations compared to non-screen-reader users. My results also identified ways to reduce this gap.

---

[1] There are over 7.6 million blind and low-vision people in the United States (of the Blind [2019]), who can benefit from using screen readers.

3.2 Study Design

I conducted two empirical studies to understand and assess current challenges screen-reader users face with online data visualizations compared to non-screen-reader users: (1) semi-structured contextual interviews (Holtzblatt and Jones [1995]) with nine screen-reader users and (2) a task-based user study with 36 screen-reader users and 36 non-screen-reader users. I present the study design for each study below.

3.2.1 Contextual Interviews

I conducted contextual interviews with nine screen-reader users to understand their interaction experiences with online data visualizations. I observed the participants interacting with real-world online data visualizations embedded in websites and subsequently interviewed them about their experiences. My research questions were:

1. What challenges do screen-reader users face when interacting with online data visualizations?
2. What information do they commonly seek in online data visualizations?
3. What techniques could improve their interaction experience with online data visualizations?

3.2.1.1 Participants

The study participants (Table 3.1) were nine screen-reader users. I recruited them using word-of-mouth, snowball sampling, social media advertisements (Facebook and Twitter), and email distribution lists for people with disabilities. Four participants identified as women and five as men. Their average age was 50.2 years (SD=18.4). Two participants were blind since birth, and seven lost vision gradually. The highest level of education attained or in pursuit was a doctoral degree for seven participants; for the remaining two participants, it was a bachelor's degree and a high school diploma, respectively. The daily computer usage was more than 5 hours for five participants, 3-5 hours for two participants, and 1-2 hours for the remaining two participants. The average visualization interaction frequency was over two visualizations per day. All participants were present in the United States.

I ceased recruitment of participants once I reached saturation of insights, as recommended by prior work (Wyche and Grinter [2009]). Participants received a $15 Amazon gift card for an hour of their time.

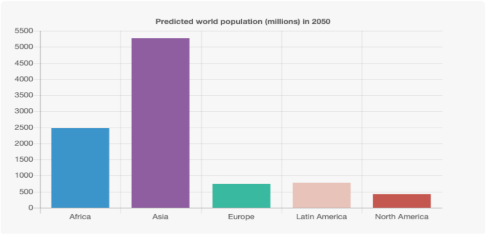
Table 3.1: Screen-reader participants for contextual interviews, their gender identification, age, screen-reader, vision-loss level, and diagnosis. Under the "G" (*Gender*) column, M = *Man*, W = *Woman*, NB = *Non-binary*, and "-" means preferred not to disclose.

	G	Age	Screen-Reader	Vision Loss Level	Diagnosis
P1	M	26	NVDA	Blind since birth	Optic Nerve Hypoplaxia
P2	M	55	JAWS	Lost vision gradually	Retinitis Pigmentosa
P3	F	30	NVDA	Blind since birth, Partial vision	Rhetonopy Prematurity
P4	F	67	Fusion	Lost vision gradually	Juvenile Macular Degeneration
P5	F	72	JAWS	Lost vision gradually	Retinitis Pigmentosa
P6	M	51	JAWS	Lost vision gradually	Demacular Degeneration
P7	F	75	JAWS	Lost vision gradually	Rhegonitis Stignitosa
P8	M	35	JAWS	Lost vision gradually	Retinitis Pigmentosa
P9	M	41	JAWS	Lost vision gradually	Angle Glaucoma

3.2.1.2 Materials

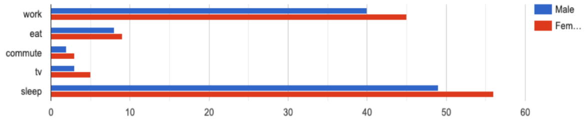
I curated a set of 50 web pages that contained a visualization created using one of three commonly used visualization libraries (Google Charts, ChartJS, or D3) and one of three common chart types (Bar, Scatter, or Line) (Saket et al. [2018]). I compiled this set of visualizations by searching for terms including "d3 visualizations," "google charts visualizations," and "chartjs visualizations" on Google. I incorporated different chart types to present participants with diverse real-world visualizations. To avoid inaccessible web pages affecting participants' experiences with the visualizations contained therein, I screened these pages to ensure they were accessible to screen readers. As my objective was to discover the accessibility of the visualizations, I did not screen for pages with accessible visualizations. Out of the 50 web pages, I randomly selected 27, nine for each of the three chart types. I used stratified random sampling to assign each participant three unique web pages, each with a different chart type.

(a) ChartJS



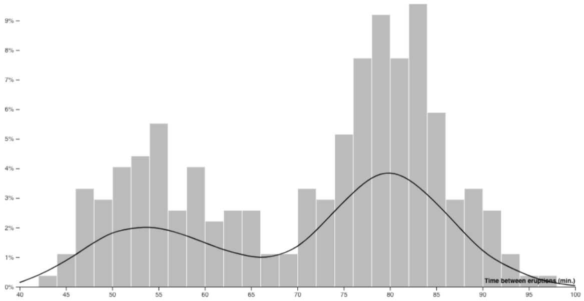
Screen-reader output:
[did not detect the element]

(b) Google Charts



Screen-reader output:
A chart, group. Male. Fem. 0. 10. 20. 30. 40. 50. 60. Sleep. TV. Commute. Eat. Work.

(c) D3



Screen-reader output:
Frame. 40. 45. 50. 55. 60. 65. 70. 75. 80. 85. 90. 95. 100. 0%. 1%. 2%. 3%. 4%. 5%. 6%. 7%. 8%. 9%

Figure 3.1: Examples of three visualizations that participants interacted with during the contextual interviews, one for each visualization library. Transcribed text from VoiceOver (Inc. [2005]): **(a)** was implemented using ChartJS (ChartJS [2015]) was not detected by the screen reader; **(b)** was implemented using Google Charts (Developers [2014]) and was identified as a "chart"; and **(c)** was implemented using D3 (Bostock et al. [2011]) and was identified as a "frame." Axis labels were also read for **(b)** and **(c)**.

3.2.1.3 Procedure

I conducted the study online using Zoom video conferencing in August 2020. Interview sessions lasted 45-60 minutes. First, I asked participants about their self-identified gender, age, screen reader, vision-loss level, and diagnosis (Table 3.1). I also asked about their education level, daily computer usage, and interaction frequency with online data visualizations.

To prepare for the observational part of the contextual interviews, I asked the participants to share their screens and make the audio output from their screen readers audible. I recorded all interviews using Zoom and subsequently used its built-in recording feature for transcription.

I asked the participants to interact with the web pages containing a visualization as they would in their daily lives, following a "think-aloud" protocol. Figure 3.1 shows a subset of three visualizations, one for each visualization library, that participants interacted with during the study session. I randomized the order of the chart type across participants. Participants spent 10-15 minutes interacting with each web page. I took detailed notes during their interactions. I interviewed them using a semi-structured structure with open prompts at the end of their interactions. Specifically, I asked participants about the information they sought, the difficulties they faced during their interactions, and the improvements that could optimize their performance in extracting information from visualizations. Additionally, I inquired whether and how their experiences with online data visualizations from this study differed from their daily life interactions.

3.2.1.4 Data Analysis

I used thematic analysis, following a semantic approach, in which themes are identified within the explicit or surface meanings of the data (Patton [1990]), guided by an essentialist paradigm (Widdicombe and Wooffitt [1995]; Potter and Wetherell [1987]). The essentialist paradigm focuses on reporting the experiences, meanings, and reality of the participants (Braun and Clarke [2006]). I developed an initial set of codes based on two interviews before coding all nine interview transcripts, adding new codes as necessary. As a result, I compiled a set of 29 codes. I coded each interview transcript independently with one co-author and resolved disagreements through mutual discussions until I reached a consensus. I calculated inter-rater reliability (IRR), expressed as percentage agreement among raters before resolving disagreements, dividing the number of codes agreed upon by the total number of identified codes across nine transcripts (Hartmann [1977]). IRR was therefore calculated as $88 \div 106 \times 100 = 83\%$, demonstrating an acceptable level of agreement (Hartmann [1977]; Graham et al. [2012]).

I combined the 29 codes into 10 axial codes using affinity diagramming. Axial codes are the product of combining the initial codes into broader, over-arching categories. I followed the thematic analysis approach by Braun and Clarke (Braun and Clarke [2006]) for analysis. After searching and reviewing themes (Phases 3 and 4 in (Braun and Clarke [2006])), my final analysis revealed nine themes across the three research questions. Additionally, I did not include the chart types in the final analysis, as I used chart types only to diversify the visualization dataset.

3.2.2 Task-Based Usability Study

To evaluate the accessibility of online data visualizations and their effects on participants' ability to extract information accurately and effectively, I conducted a mixed factorial experiment with screen-reader- and non-screen-reader users. The experiment was conducted online without supervision.

Table 3.2: Screen-reader participants for task-based user study, their gender identification, age, screen-reader, vision-loss level, and diagnosis. Under the "G" (*Gender*) column, M = *Man*, W = *Woman*, NB = *Non-binary*, and "-" means preferred not to disclose.

	G	Age	Screen-Reader	Vision-Loss Level	Diagnosis
S3	F	67	Fusion	Partial vision, Lost vision gradually	Juvenile Macular Degeneration
S4	M	55	JAWS	Lost vision gradually	Retinitis Pigmentosa
S5	F	30	NVDA	Lost vision gradually, Partial vision	Retinopathy of Prematurity
S6	F	63	JAWS	Lost vision gradually	Retinitis Pigmentosa
S12	F	35	JAWS	Blind since birth	Leber's Congenital Amaurosis
S13	M	41	JAWS	Lost vision gradually	Juvenile Onset Open Angle
S15	M	40	JAWS	Partial vision, Lost vision gradually	Retinitis Pigmentosa
S16	M	47	JAWS	Lost vision gradually	Leber's Congenital Amaurosis
S17	M	35	JAWS	Blind since birth	Leber's Congenital Amaurosis
S18	F	51	JAWS	Blind since birth	-
S19	M	51	JAWS	Blind since birth	-
S20	M	31	NVDA	Blind since birth, Lost vision gradually	Peter's Anomaly
S21	M	48	NVDA	Lost vision gradually	Retinitis Pigmentosa
S22	GF	24	VoiceOver	Partial vision	Partial Sight Impairment
S23	F	27	NVDA	Blind since birth, Lost vision gradually	Retinal Detachment
S25	F	64	JAWS	Partial vision	-
S26	F	39	Fusion	Lost vision gradually	-

S28	F	53	JAWS	Lost vision gradually	Optic Neuropathy
S29	F	22	NVDA	Lost vision gradually	Retinitis Pigmentosa
S30	M	60	JAWS	Partial vision	Optic Neuropathy
S31	M	46	JAWS	Lost vision gradually	Retinitis Pigmentosa
S32	M	29	NVDA	Blind since birth	Leber's Congenital Amaurosis
S33	F	46	JAWS	Lost vision gradually	Optic Neuritis/Atrophy, Diabetic Retinopathy
S34	M	57	JAWS	Lost vision gradually	Glaucoma
S35	M	18	NVDA	Blind since birth	Retinopathy of Prematurity
S36	F	63	JAWS	Lost vision gradually	Cataracts
S40	F	28	NVDA	Blind since birth, Lost vision gradually	Optic Nerve Hypoplasia and Glaucoma
S41	M	27	NVDA	Blind since birth	Optic Nerve Hypoplasia
S42	F	68	JAWS	Blind since birth	Retinopathy of Prematurity
S44	F	34	JAWS	Blind since birth	Renal Retinal Dysplaysia
S46	F	72	JAWS	Lost vision gradually	Retinitis Pigmentosa
S47	M	39	JAWS	Lost vision gradually	Retinitis Pigmentosa, Maculae Degeneration
S48	M	47	JAWS	Lost vision gradually, Partial vision	Leber's Congenital Amaurosis
S49	M	41	JAWS	Blind since birth	Micropthalmia
S50	M	43	Other	Blind since birth	Retinopathy of Prematurity
S51	F	46	JAWS	Lost vision gradually	Optical Nerve Damage

3.2.2.1 Participants

I recruited 72 participants using word-of-mouth, snowball sampling, and through social media channels (Facebook and Twitter); 36 were screen-reader users (see Table 3.2). I also advertised on email distribution lists for people with disabilities. I calculated the sample size at 0.8 power to detect a large effect size at the standard .05 alpha significance threshold, assuming normal distribution. Among screen-reader users, 17 identified as women, 18 as men, and one as genderfluid. Their average age was 44.1 years (SD=14.1). In

Table 3.3: Cardinality summary from Borkin et al. [2013]'s dataset comprising 2,068 two-dimensional single-panel visualizations used to determine the range for the different complexity levels.

	N	Minimum	Maximum	Median	Mode	Mean	25th Percentile	75th Percentile
Line	50	5	168	29	22	41	17	56
Bar	50	6	42	20	20	21	15	27
Scatter	50	7	170	39	25	52	25	72

the group of non-screen-reader users, 21 identified as women and 15 as men, with their average age being 43.4 years (SD=12.8). The age was not significantly different between the two groups (t(70)=0.23, p=.823).

Screen-reader users were compensated with a $15 Amazon gift card for 45 minutes of their time. Non-screen-reader users received a $10 Amazon gift card for 20 minutes. I calculated the compensation amount based on the average study completion time. No participant was allowed to partake more than once.

3.2.2.2 Apparatus

I implemented an online task-based study using the React (Inc. [2013]) framework, ensuring maximum and proper accessibility measures by testing it myself and with screen-reader users. I deployed the study online on my server and shared the link to the website with the participants.

3.2.2.2.1 Visualization Dataset Following prior work (Pini et al. [2019]; Smith et al. [2018]), I collected 27 different datasets from Kaggle[2]—one of the largest online resources for open datasets. In choosing the visualization datasets, my goal was to ensure that the datasets were both topically diverse and realistic. The topics were mutually agreed upon by all the authors and selected such that (a) a dataset for a given topic only existed at most once in the pool of datasets and (b) datasets represented different fields of interest, filtered using "tags" on Kaggle. For line charts, nine of the 27 datasets represented temporal data, which showed a clear trend to avoid misinterpretation.

To account for visualizations having a varying range of data points, I subdivided the dataset into different levels of complexity. I randomly sampled 50 visualizations for each chart type, totaling 50×3=150 visualizations, from Borkin *et al.*'s(Borkin et al. [2013]) dataset comprising 2,068 two-dimensional single-

---

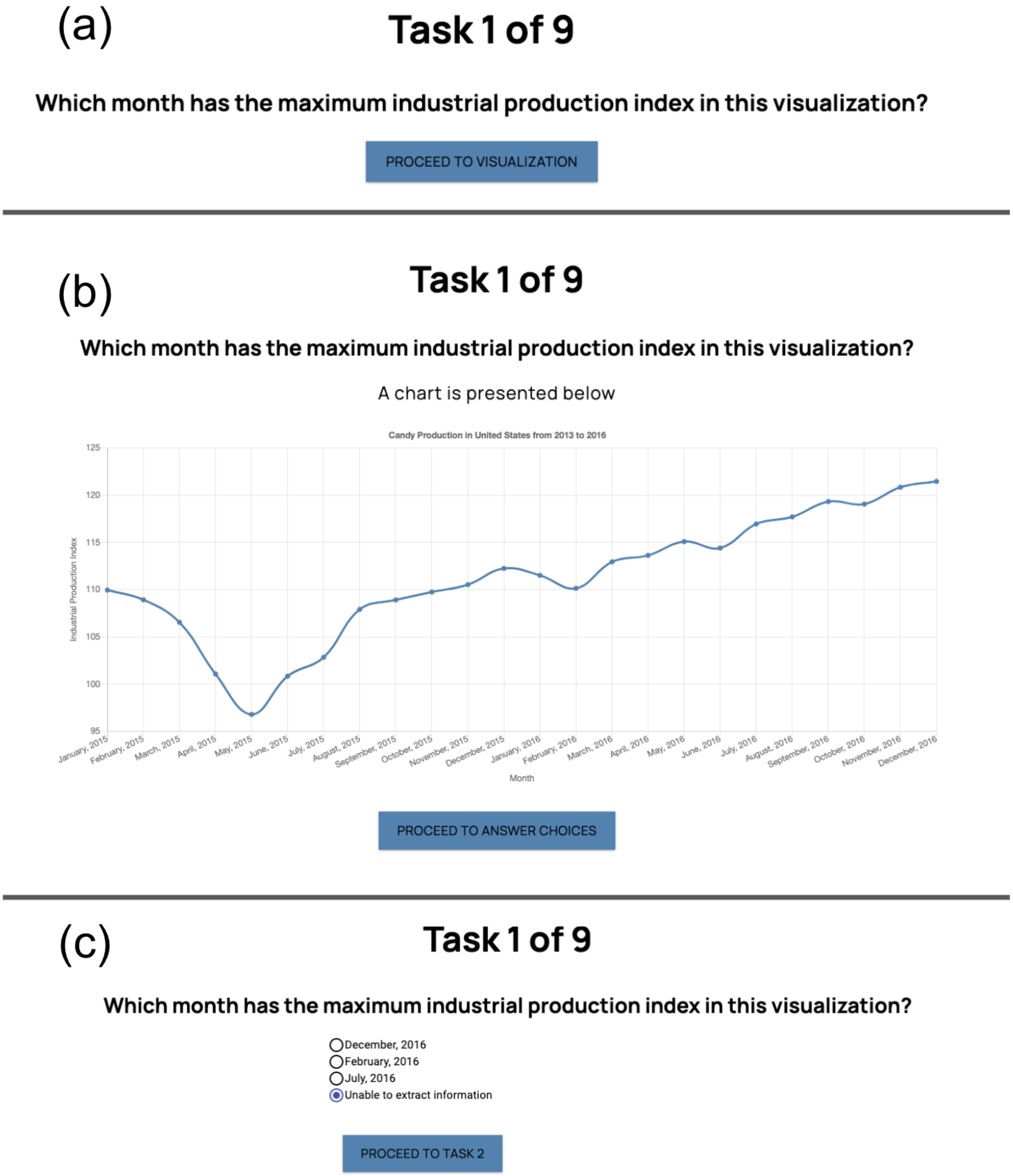
[2] https://www.kaggle.com/datasets

panel visualizations. I used this sample to find the minimum, maximum, 25th percentile, and 75th percentile of their cardinalities (Table 3.3), which I used to determine the range for the complexity levels. Specifically, the datasets had a random cardinality between the minimum and 25th percentile, the 25th and 75th percentile, and the 75th percentile and the maximum, for low, medium, and high complexity levels, respectively.

I used the datasets acquired from Kaggle to implement the visualizations for the online experiment, following the WCAG 2.0 Guidelines (Caldwell et al. [2008]) in conjunction with the official accessibility recommendations from the visualization libraries. All visualizations were interactive by default and generated using all three visualization libraries. I present these visualizations in Appendices A, B, and C.

3.2.2.2.2 Question Categories The study utilized four question categories to measure the accuracy of extracted information from the data visualizations. The categories were derived based on Brehmer and Munzner's task topology (Brehmer and Munzner [2013]). Specifically, I considered one Search action (*lookup and locate*) and two Query actions (*identify and compare*), similar to prior work (Brehmer et al. [2018]). Each category was assigned a difficulty level, determined by a discussion and mutual agreement between at least two authors based on their knowledge and familiarity with the subject matter. The categories, in ascending order of difficulty, were:

1. *Order Statistics*: I asked the participants about either the maximum data point or the minimum data point, chosen randomly (for example, "What is the minimum data point in the visualization?" or "What is the maximum data point in the visualization?")
2. *Symmetry Comparison*: I asked the participants to identify the relationship between two data points (for example, "How is [random data point 1]'s value in comparison to that of [random data point 2]?")
3. *Chart Type-Specific Questions*:
    - *Value Retrieval*: I asked the participants to extract the information from a given individual data point (for example, "What is the corresponding value for [random data point]?"). I only asked this question for bar charts.
    - *Trend Summary*: I asked the participants about the overall data trend (for example, "What was the overall trend of the visualization?"). I curated the dataset to ensure no ambiguity in the answer. I only asked this question for line charts.



Task 1 of 9

Which month has the maximum industrial production index in this visualization?

Task 1 of 9

Which month has the maximum industrial production index in this visualization?

A chart is presented below

Task 1 of 9

Which month has the maximum industrial production index in this visualization?

○ December, 2016
○ February, 2016
○ July, 2016
● Unable to extract information

Figure 3.2: Participants in Study 2 were shown three pages in a single task. (**a**): Page 1 presented the question to explore. (**b**): Page 2 displayed the same question and a visualization. (**c**): Page 3 showed the question again with a set of four multiple choice responses.

- *Correlation*: I inquired about the correlation between the dependent and independent variables in the visualization (for example, “What was the correlation between [dependent variable] and [independent variable]?”). I curated the dataset, ensuring no ambiguity in the answer while keeping the data points scattered along the axes. I only asked this question for scatter plots.

All questions were multiple-choice with four choices: the correct answer, two incorrect answers, and the option for “Unable to extract information.” I randomized the order of the choices per trial.

3.2.2.3 Procedure

The study was conducted online without supervision. I showed the participants the study purpose, eligibility criteria, and the statement of IRB approval on the first page. On the next page, the participants were asked to fill out a pre-study questionnaire to record their demographic information, screen reader, vision-loss level, and diagnosis (see Table 3.2). I also asked about their education level, daily computer usage, and interaction frequency with visualizations. Then, I showed the participants instructions for completing the study tasks.

Each participant interacted with three visualizations created using one of three commonly utilized visualization libraries (Google Charts, ChartJS, D3). Figure 3.2(b) shows an example visualization. For each visualization, I asked the participants to answer three questions. Each of the three questions represented a different difficulty level (Low, Medium, High), assigned by mutual agreement from at least two authors based on the ease of extracting the answers from the visualizations. I counterbalanced the complexity (Low, Medium, High) and chart type (Bar, Line, Scatter) for the visualizations for each order of visualization libraries across participants.

For each of the three *Visualization Library* $\times$ *Complexity* conditions, participants were shown three pages: Page 1 contained the question to explore; page 2 displayed the question and visualization; and page 3 presented the question with a set of four multiple-choice responses from which participants chose the answer (see Figure 3.2). I randomized the order of the questions per visualization; each question appeared at the top of the page. I asked the participants to interact with the visualization as they would in their daily lives. For screen-reader users, a study session took 30-45 minutes from start to finish, whereas for non-screen-reader users, the total time for the study session ranged between 10-20 minutes.

3.2.2.4 Design & Analysis

The experiment was a mixed factorial design with the following factors and levels:

- *Screen-Reader User*, between-Ss.: {yes, no}
- *Visualization Library*, within-Ss.: {ChartJS, D3, Google Charts}
- *Data Complexity*, within-Ss.: {Low, Medium, High}
- *Question Difficulty*, within-Ss.: {Low, Medium, High}

My dependent variables were *Accuracy of Extracted Information* (AEI) and *Interaction Time* (IT). For tractability, I treated AEI as binary, classifying AEI as "inaccurate" if users incorrectly answered the question or were unable to extract information. It was "accurate" otherwise. As for IT, for screen-reader users, IT was calculated as the total time of focus on the root visualization element. I made this decision to represent a screen-reader user's interaction experience accurately. For non-screen-reader users, IT was calculated simply as the total focus time on the webpage containing the visualization.

To analyze AEI, I used mixed logistic regression (Gilmour et al. [1985]) with the above factors, their interactions, a covariate of *Age*, and a random effect for *Subject* to account for repeated measures. The statistical model was *AEI* = *SRU* $\times$ *VL* + *SRU* $\times$ *CMP* + *SRU* $\times$ *DF* + *Age* + *Subject*. To analyze IT, I used a linear mixed model analysis of variance (Frederick [1999]; Littell et al. [1998]), with the same model as for AEI.

Participants were tested over three *Visualization Library* $\times$ *Complexity* conditions, resulting in a total of $3{\times}3 = 9$ trials per participant. With 72 participants, a total of $72{\times}9 = 648$ trials were produced and analyzed in this study.

3.3 Results

This section presents the results from my two empirical studies, shedding light on the interaction experiences of screen-reader users with online data visualizations. Furthermore, I provide empirical results highlighting the disenfranchisement these users face compared to non-screen-reader users. I present the results of each study separately below.

3.3.1 Contextual Interviews

I identified nine main themes across my three main research questions.

3.3.1.1 RQ1: Challenges and Pain Points

My analysis revealed three themes relevant to challenges that screen-reader users experience with online data visualizations: (1) Visualizations are often completely indiscoverable by screen readers, (2) Information read out by screen readers often lacks context, making the information difficult to comprehend, and (3) Data contained in the visualizations is not available to the screen-reader users, restricting screen-reader users to explore the data.

3.3.1.1.1 The First Problem: Invisible Visualizations The first theme that emerged in my study was that in 33% of the web pages presented to the participants, their screen reader did not detect the visualization at all. Consequently, participants interacted with the web page but were unaware it contained a visualization. For example, P9 said:

> `It didn't seem like there's anything.` (P9)

Participants expressed similar experiences when interacting with visualizations in their daily lives. P3, who has only had partial vision since birth, said:

> `So the actual visualizations and the graphs, I can't access at all. Unless they have an image description with them, which they usually don't.` (P3)

This finding shows that screen-reader users commonly interact with websites without knowing that a visualization is present.

3.3.1.1.2 The Second Problem: Incomprehensible Visualizations My second theme showed that even when the screen reader detected the visualization, the information read to the participants was insufficient for fully comprehending the visualization data. Screen readers often only identified the visualization as "blank," "graphic," "frame," or "object". Noticeably frustrated, P6 commented:

> `It says "graphic, graphic, graphic..." It means nothing to a blind person. Have to be more descriptive, saying "graphic" means nothing.` (P6)

Similarly, P2 was confused when their screen reader identified the visualization as an "object"—a term that usually identifies the HTML object tag:

> `I guess a pain point would be when [screen reader] said "object." I wouldn't normally expect to hear that in any kind of visualization, so that was a little confusing for me.` (P2)

Screen-reader users were also often unable to infer what the data was referring to. P3, who has been blind since birth, reported being confused about whether the screen reader was reading out axis labels or other data:

> `When it was saying like "12k," "1k," I had no idea what that was referring to.` (P3)

When screen readers read the data from a data table (as provided in Google Charts), screen-reader users must go through each data point, which can be tedious and cognitively challenging, especially with larger datasets. P9, who interacts with several visualizations every week, found complex visualizations in the study very time-consuming:

> `It seemed like I went through a lot of [data] points.` (P9)

3.3.1.1.3 Where's the Data: Lack of Access to Data Points My third theme showed that screen-reader users were frustrated due to the lack of direct access to the underlying data points. For example, P3 and P8 considered access to the data points a key piece to explore visualizations. They shared their frustrations:

> `The information that was actually graphed, so again, like the points--I wasn't able to access that. Right now I have zero access to the data point, which is the whole point of having a graph.` (P3)

> `I mean, data's not accessible, so that's why I got stuck.` (P8)

Overall, the answer to the first research question is that visualizations are often entirely invisible to screen readers, taking away the opportunity for screen-reader users to explore the data and extract meaningful information. In cases when the visualization element is detectable, it is recognized meaninglessly as "blank," "graphic," "frame," or "object." Additionally, screen-reader users trying to access visualizations produced by ChartJS or D3 do not have access to the underlying data. Google Charts provides access to the data, but the participants found going through each data point tedious and time-consuming.

3.3.1.2 RQ2: Commonly Sought Information

Two main themes emerged from my analysis that are relevant to the second research question on the information screen-reader users commonly seek in visualizations: (1) Screen-reader users first explore the data holistically, and (2) after an initial holistic exploration, they look for and compare individual data points.

3.3.1.2.1 The Flyover: Holistic Exploration and Trend Assessment My first theme for RQ2 showed that screen-reader users sought to obtain a holistic sense of the data in the online visualizations, getting the "feel" for the information in the visualization before deciding to explore the data further, similar to non-screen-reader users. For example, P6 described how screen-reader users approach online visualizations:

> `When we first interact with something new to us, first we want to try to read everything at a quick glance and then the second time, the third time, the fourth time we really want to understand.` (P6)

To get a holistic overview of the data, the participants specifically looked for the *overall trend*, *extrema* (data points representing minimum and maximum values), and *axis information* (labels and ranges for each axis) in the visualizations presented to them during the study session.

When looking for the overall trend, I found that the participants developed a mental image of the data. They did so by interacting with one data point at a time and navigating through the SVG elements or the items in the data table, depending on the visualization library. For example, P2 and P4, both of whom lost their vision gradually, described their thought process:

> `In my mind, I try to move along the graph and visualize the trend as it goes up and down.` (P2)

> `I can just listen to data and get an idea of what that represents over time.` (P4)

Additionally, P5, in their interaction, was interested in the maximum and minimum values (extrema):

> `I would gather the bulk of the info, such as the minimum and maximum.` (P5)

Similarly, P1 was able to get a holistic overview of the data from the axes labels and ranges:

> `X-axis is the time and Y-axis is basically giving the popularity score where 100 is the most popular.` (P1)

3.3.1.2.2 The Drill-Down: Investigating Specific Data Points My second theme showed that participants were interested in the *individual data points*. It is worth noting that the participants could only explore the data granularly after gaining a holistic view of the information in the data visualization. Additionally, participants explored how specific data points compared to other data points (greater, lesser, or the same). For example, P9 and P2 shared their experiences of comparing data points:

> `[I look for] the opening price and the closing price of one minute worth of data all the way up to days, or maybe for months, for data.` (P9)

> `What I would think is that the male is higher, and the female is a lower number right there.` (P2)

Overall, I found screen-readers first explore the data holistically, seeking the overall trend, extrema, and axis information. For visualizations that pique the interests of screen-reader users after an initial holistic exploration, they explore the individual data points both exclusively and in comparison to other data points.

3.3.1.3 RQ3: Techniques and Strategies for Accessible Visualizations

To understand future possibilities of information visualizations for screen-reader users, I explored strategies that would improve the accessibility of the visualizations. I identified four strategies for RQ3: (1) Tabular representation, (2) Textual representation, (3) Overall trends in non-visual formats, and (4) Multi-modality.

3.3.1.3.1 Tabular Representation My first theme that emerged for RQ3 was that the participants highlighted the importance of having a tabular representation of the data in place of a visualization. For example, P3 and P4 described their positive experiences from everyday life with tabular representations of data:

> `One of my friends developed a website for blind and visually impaired people with COVID-19 data. It's actually based in a table instead of a bar graph, so he created this tool that pulls data every day and puts it in a chart form.` (P3)

> `The daily newspapers that I listen to, they list all the municipalities in that county and it's like a three column table, and so it's the name of the municipality, the number of cases to date, and then the number of deaths to date, and some of them have what increase or decrease that represents since March 1st, so that's wonderful for me. I don't need to look at a chart that way I can just listen to data and get an idea of which municipalities have the greatest number of cases, which of them have had the most fatalities, and what that represents over time.` (P4)

3.3.1.3.2 Textual Representation My second theme showed that screen-reader users value the importance and benefits of using textual representations of data to aid in the visualization interaction experience, within and outside the study session. For example, P1 highlighted the benefits of including alt-text:

> `Most visualizations are, of course, inaccessible. If somebody cares to put alternative text, yes, that helps a bit.` (P1)

In another instance during the study, P1 found the alt-text to help obtain the holistic overview of data:

> `This help blob really helped distill the information down more clearly. The alternative text gave me enough information about what the visualization was about.` (P1)

P6 considered alt-text as the best solution:

> `Like I said, alt-text is the best solution to solve this [inaccessible visualizations] problem. So we need to have alternative text to describe the graphics.` (P6)

3.3.1.3.3 Overall Trend In Non-Visual Formats

My third theme showed that screen-reader users emphasized the importance of presenting the trend in an alternate format compared to the visual format. They identified that a non-visual format would reduce the burden of deducing the overall trend on screen-reader users and increase the holistic exploration of the data. As P1, who is blind since birth, said:

> `It would be helpful to know how the visualization would look in a visual aspect. For example, when people talked about COVID-19, they said it's an exponential growth and flatten the curve--but what does that mean? It means one thing numerically, and it means another thing visually. And for you to participate in a conversation, you should know what the visual aspect means for you to sensibly contribute.`

The participants also identified *sonification*, *summarization*, and *braille printouts* as three techniques to present online visualizations involving data trends to screen-reader users. For example, P3 and P8 shared their thoughts:

> *Sonification*: `Having an auditory graph is also super helpful; I just learned how to do that in the health app on my phone yesterday, and it was really helpful to get that audio feedback about what this graph looked like.` (P3)

> *Summarization*: Showing the highest point, lowest point, maybe mean or average, and a trend, is it generally going up, down, up and down, is it stagnant? Just a brief description that you would be able to see by just looking at it for one or two seconds. Oh, just a quick paragraph before the data just explaining it, you know, this is the dataset from whatever organization "X" that is showing, you know, "X" trend over whatever amount of time or something, you know what I mean? Just something like preferably in plain language as well. I'm a big fan of just kind of explaining, you know. (P8)

> *Braille Printouts*: My preferred method is still, because I am a visual learner, so I do appreciate when I have access to a hard copy braille of a graph. Just because I am more visual. So it depends on which information you're trying to convey on whether or not a table is more useful than an actual graph. (P3)

3.3.1.3.4 The More Options, The Better: Multi-Modality My last theme showed that screen-reader users consider a multi-modal solution beneficial in exploring the data contained in the visualizations. For example, P8 and P1 expressed the importance of multi-modality and having various options for accessing the information contained in a visualization:

> The most ideal way to access [information] is like having options, multi-modal approach. Where, like, I can view it as a table or a list or I can download it as an Excel spreadsheet, you know? (P8)

> Like overview kind of a thing. Like to look at the general trend. And I want to be able to control how I consume the data. And I want to know the visual attributes of the plot or the graph. (P1)

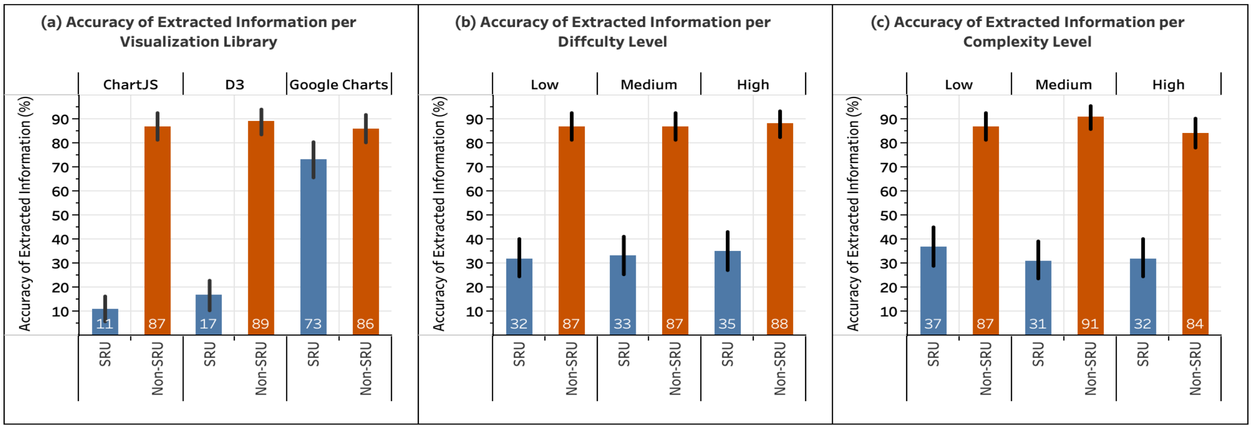


Figure 3.3: Accuracy of Extracted Information (AEI), as a percentage, for screen-reader and non-screen-reader users, per **(a)** Visualization Library, **(b)** Difficulty Level, and **(c)** Complexity Level. AEI was classified as "inaccurate" if the user incorrectly answered the question or was unable to extract the information, and as "accurate" otherwise. The percentage represents the "accurate" answers. Therefore, higher is better.

Overall, I found that screen-reader users prefer the following techniques and strategies to improve the accessibility of online data visualizations: (1) Tabular representation, (2) Textual representation, (3) Overall trends in non-visual formats, and (4) Multi-modality.

3.3.2 Task-Based User Study

In this section, I present the results of the task-based user study focusing on the *Accuracy of Extracted Information* (AEI) and *Interaction Time* (IT) for screen-reader and non-screen-reader users with online data visualizations. It is worth re-emphasizing that my work does not assess the cognitive or intellectual abilities of the participants, especially screen-reader users; instead, my work focuses on AEI and IT as a function of the *accessibility* of online data visualizations.

As a preliminary matter, I checked the regression models for multicollinearity by calculating the variance inflation factor (VIF) and found that multicollinearity between AEI and IT was not a concern (VIF=1.49).

3.3.2.1 Accuracy of Extracted Information (AEI)

The results show a significant main effect of *Screen-Reader User* (SRU) on AEI overall (χ^2(1, N=72) = 67.22, p<.001, Cramer's V=0.18), indicating that AEI differs significantly between the two *Screen-Reader User* groups. In fact, AEI is considerably lower for screen-reader users (34%) compared to non-screen-

Table 3.4: Numerical results for the $N = 648$ questions asked of screen reader users (SRUs) and non-SRUs for each level of *Visualization Library*, *Difficulty Level*, and *Complexity Level*. N is the total number of questions asked, AA and $AA(\%)$ are the number and percentage of "accurate answers," respectively.

	Both Groups			*SRUs*			*Non-SRUs*		
	N	AA	AA (%)	N	AA	AA (%)	N	AA	AA (%)
Overall	648	392	60%	324	109	34%	324	283	87%
Visualization Library									
- ChartJS	216	106	49%	108	12	11%	108	94	87%
- D3	216	114	53%	108	18	17%	108	96	89%
- Google Charts	216	172	80%	108	79	73%	108	93	86%
Difficulty Level									
- Low	216	129	60%	108	35	32%	108	94	87%
- Medium	216	130	60%	108	36	33%	108	94	87%
- High	216	133	62%	108	38	35%	108	95	88%
Complexity Level									
- Low	216	134	62%	108	40	37%	108	94	87%
- Medium	216	132	61%	108	34	31%	108	98	91%
- High	216	126	58%	108	35	32%	108	91	84%

reader users (87%), constituting a percentage difference of 61.48%.

There was also a significant main effect of *Visualization Library* (VL) on AEI overall (χ^2(2, N=72) = 40.45, p<.001, Cramer's V=0.14). This result indicates that AEI differs significantly between different visualization libraries. Figure 3.3 and Table 3.4 show the AEI percentages across different visualization libraries. *Google Charts* had the best performance (73%) for screen-reader users, followed by *D3* (17%) and *ChartJS* (11%). For non-screen-reader users, all three visualization libraries performed almost identically.

I also examined whether changes in AEI were proportionally similar or different between the visualization libraries for participants in each *Screen-Reader User* group. To do so, I examined the *SRU* $\times$ *VL* interaction and found it to have a significant effect on AEI overall (χ^2(2, N=72)=50.35, p<.001, Cramer's V=0.15). This result indicates that AEI not only significantly differs between visualization libraries over-

Table 3.5: Summary results from 72 screen-reader users (SRUs) and non-SRUs using mixed logistic regression (Gilmour et al. [1985]). The statistical model was *AEI = SRU × VL + SRU × CMP + SRU × DF + Age + Subject*. *Subject* was included as a random variable. AEI represents the accuracy of extracted information ("inaccurate" or "accurate"). Cramer's V is a measure of effect size (Ferguson [2016]).

	N	χ^2	p	Cramer's V
Screen Reader Usage (SRU)	72	67.22	$< .001$	0.18
Visualization Library (VL)	72	40.45	$< .001$	0.14
SRU × VL	72	50.35	$< .001$	0.15
Complexity (CMP)	72	1.94	.380	0.03
SRU × CMP	72	2.20	.332	0.03
Difficulty (DF)	72	0.35	.838	0.01
SRU × DF	72	0.06	.972	0.00
Age	72	2.70	.100	0.04

all but also between screen-reader and non-screen-reader users. Table 3.4 shows AEI percentages across different visualization libraries for each user group.

Additionally, I also investigated the effects of *Complexity, Difficulty, Age*, and interactions between *SRU* × *Complexity*, and *SRU* × *Difficulty*, but did not find a significant effect on AEI (see Table 3.5).

3.3.2.2 Interaction Time (IT)

Anderson-Darling (Anderson and Darling [1954]) goodness-of-fit tests of normality showed that the interaction times were conditionally non-normal. Further inspection revealed these values to be conditionally lognormal, and so a logarithmic transformation was applied before analysis, as is common practice for time measures (Hoyle [1973]; Limpert et al. [2001]; Berry [1987]). Examination of this log-transformed response indicated that it was indeed normal ($p \approx .234$). For ease of communication, I show the plots of this dependent variable using the original non-transformed values.

The factor *Screen-Reader User* (SRU) had a significant main effect on *Interaction Time* (IT) overall ($F(1,69)=115.33$, $p<.001$, $\eta_p^2=0.63$). Specifically, the average IT for screen-reader users was 84.6 seconds (SD=75.2). For non-screen-reader users, it was 27.2 seconds (SD=16.8). The average IT for participants

Table 3.6: Summary results from 72 screen-reader users (SRUs) and non-screen-reader users showing the numerical results for *Interaction Time* (IT), for each age range. N is the total number of participants for the given age range, *Mean* is the average IT in seconds, and SD represents the standard deviation.

	Both Groups			*SRUs*			*Non-SRUs*		
Age Range	N	Mean	SD	N	Mean	SD	N	Mean	SD
18-20	2	65.8	43.9	1	96.8	-	1	34.8	-
20-30	12	40.6	28.0	6	63.7	20.3	6	17.4	4.9
30-40	14	44.2	23.1	7	60.6	16.5	7	27.7	16.1
40-50	19	67.5	78.7	10	106.7	93.1	9	24.0	10.7
50-60	14	47.6	27.3	5	79.1	19.2	9	30.1	8.1
60-70	9	64.8	33.6	6	76.8	33.7	3	40.6	19.0
> 70	2	127.0	127.4	1	217.1	-	1	36.9	-

who used screen readers was 210.96% more than for participants who did not.

There was a significant main effect of *Visualization Library (VL)* (F(2,564)=25.10, p<.001, η_p^2=0.08) on IT. Furthermore, the interaction between *SRU* × *VL* was also significant (F(2,564)=31.58, p<.001, η_p^2=0.10). These results indicate that IT not only significantly differed between visualization libraries overall but also differentially between visualization libraries for participants in each group. Figure 3.4 and Table 3.7 show interaction times across different visualization libraries for each user group. For screen-reader users, *ChartJS* had the minimum interaction time, followed by *D3* and *Google Charts*. For non-screen-

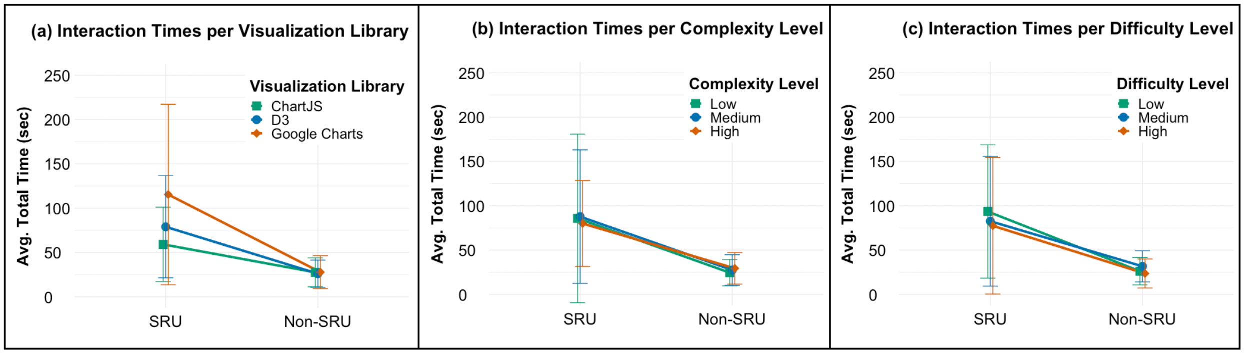


Figure 3.4: Interaction Times (IT), in seconds, for screen-reader and non-screen-reader users, per **(a)** Visualization Library (VL), **(b)** Complexity Level (CMP), and **(c)** Difficult Level (DF).

Table 3.7: Summary results from 72 screen-reader and non-screen-reader participants using a mixed-effects model analysis of variance (Frederick [1999]; Littell et al. [1998]). The statistical model was *IT* = *SRU* × *VL* + *SRU* × *CMP* + *SRU* × *DF* + *Age* + *Subject*, where *Subject* was modeled with a random intercept. Partial eta-squared (η_p^2) is a measure of effect size (Cohen [1973]).

	df_n	df_d	F	p	η_p^2
Screen Reader Usage (SRU)	1	69	115.33	$< .001$	0.63
Visualization Library (VL)	2	564	25.10	$< .001$	0.08
SRU:VL	2	564	31.58	$< .001$	0.10
Complexity (CMP)	2	564	6.00	.003	0.02
SRU:CMP	2	564	2.01	.136	0.01
Difficulty (DF)	2	564	24.07	$< .001$	0.08
SRU:DF	2	564	12.75	$< .001$	0.04
Age	1	69	7.03	.010	0.09

reader users, all three visualization libraries performed almost identically.

I found a significant main effect of *Difficulty (DF)* (F(2,564)=24.07, p<.001, η_p^2=0.08) on IT. Furthermore, the interaction between *SRU* × *DF* was also significant (F(2,564)=12.75, p<.001, η_p^2=0.04). These results indicate that IT not only significantly differed between question difficulty levels overall but also between difficulty levels for participants in each screen-reader user group. Figure 3.4 and Table 3.7 show interaction times across different difficulty levels for each user group.

Additionally, *Complexity (CMP)* had a significant main effect on IT overall (F(2,564)=6.00, $p \approx$.003, η_p^2=0.02), indicating that IT differed significantly between different complexity levels. Figure 3.4 and Table 3.7 show interaction times across different complexity levels. I also examined the interaction between *SRU* × *CMP*, but did not find a significant effect.

I investigated the effects of *Age* on IT. *Age* had a significant effect on IT (F(1,69)=7.03, p<.05, η_p^2 =0.09), indicating that IT differed significantly across the ages of the participants, with participants over the age of 50 years showing higher interaction times by about 12.18% than participants under the age of 50. Table 3.6 shows the average IT for each age range of screen-reader- and non-screen-reader users.

In addition to the above analyses, I examined the interactions between *VL* × *CMP*, *VL* × *DF*, and *CMP*

$\times$ *DF* to explore the relationship between the independent variables, but did not find any significant effects.

3.4 Design Recommendations

My results have several important implications for the design of future online data visualizations that equalize access for screen-reader users.

3.4.1 Online Data Visualizations Should be Discoverable and Comprehensible

One of the disadvantages screen-reader users face concerning online visualizations is that many visualizations are indetectable to screen readers. Therefore, the information within such visualizations is completely hidden from screen-reader users. I recommend developers and visualization creators adequately provide alternative text and use ARIA attributes to improve the discoverability of the visualization elements.

3.4.2 Visualization Libraries Should Offer Both Holistic and Drilled-Down Explorations

As my studies showed, current popular online visualization libraries either encourage using alternative text (ChartJS and D3) or automatically include a data table (Google Charts). However, even though alternative text can support gaining a holistic overview, it is usually not comprehensive enough to allow an understanding of the data in detail. In contrast, data tables allow detailed exploration but add a mental burden for screen-reader users to process all data. I recommend supporting screen-reader users to explore preferences, as found in this study.

3.4.3 Alternative Text Should be Auto-Generated Based on the Underlying Data

My findings showed that alternative text is often absent, and when it is present, it is mostly inadequate. Additionally, the quality of the alternative text primarily depends on the developer, which can produce inconsistencies in visualization interaction experiences for screen-reader users from one visualization to another. Therefore, I recommend dynamically generating alternative text, similar to how it has been proposed in prior work (Sharif and Forouraghi [2018]; Mirri et al. [2017]). Additionally, given that every user is unique and may prefer different information in the alternative text, I recommend generating customized alternative text based on the individual preferences of screen-reader users, collected using a centralized system.

3.4.4 Visualization Libraries Should Offer Multi-Modality for Exploring Data

My study found that while data tables are beneficial, many screen-reader users preferred additional approaches, such as data sonification, to explore data. Presenting screen-reader users with multiple modes for exploring visualizations would improve their overall experience with online data visualizations. Multi-modality is especially beneficial for complex visualizations, which could require more exploration time.

3.5 Summary

This chapter presented an empirical evaluation of the experiences of screen-reader users interacting with online data visualizations compared to those of non-screen-reader users. I used a mixed-methods approach, employing contextual interviews and a quantitative task-based user study with 45 screen-reader- and 36 non-screen-reader users. The findings from these studies show that screen readers often do not even "see" data visualizations, and when they do, these visualizations still inadequately support screen-reader users. Due to the inaccessibility of online data visualizations, extracting information using a screen reader is both inaccurate and time-consuming, creating significant disparities between screen-reader and non-screen-reader users. This work further emphasizes the need for solutions to make online data visualizations accessible. In the next chapter, I present VoxLens, an open-source multi-modal JavaScript plug-in that makes online data visualization accessible to screen-reader users.

Chapter 4

VoxLens: A JavaScript Plug-In to Make Online Data Visualizations Accessible*

This chapter presents the foundational design of a novel interaction technique called VOXLENS, which is an open-source multi-modal JavaScript plug-in that improves the accessibility of online data visualizations. VOXLENS is the first system to enable screen-reader users to interact with and extract information from online data visualizations using voice-activated commands. Furthermore, it is the first to improve visualization creators' understanding of screen-reader users' experiences, enhance their knowledge of visualization accessibility, and assist them in making data visualizations more accessible. I named my tool VOXLENS, combining "vox," meaning "voice" in Latin, and "lens," since it provides a way for screen-reader users to explore, examine, and extract information from online data visualizations.

I intend the design of VOXLENS present in this chapter to acclimatize readers to VOXLENS' core concept and functionalities. Chapters 5 and 7–10 provide the design and implementation of respective VOXLENS versions, along with the formative and summative studies to motivate and evaluate those versions.

4.1 Wizard-of-Oz Studies

As a preliminary measure to motivate the design of VOXLENS and gather feedback, I conducted a Wizard-of-Oz study (Dahlbäck et al. [1993]; Hajdinjak and Mihelic [2004]) with five screen-reader users (see Ta-

---

*Parts of this chapter are adapted from Sharif et al. [2022f].

Table 4.1: Screen-reader participants for the Wizard-of-Oz experiment, their gender identification, age, screen reader, vision-loss level, and diagnosis. Under the "G" (*Gender*) column, M = *Man*, F = *Woman*, and NB = *Non-binary*.

	G	**Age**	**Screen Reader**	**Vision-Loss Level**	**Diagnosis**
W1	M	25	VoiceOver	Partial vision	Extremely low vision
W2	M	28	NVDA	Blind since birth	Optic Nerve Hypoplaxia
W3	M	23	VoiceOver	Blind since birth	Septo-optic Dysplasia
W4	F	26	JAWS	Blind since birth	Leber Congenital Amaurosis
W5	M	31	JAWS	Blind since birth	Retinopathy of Prematurity

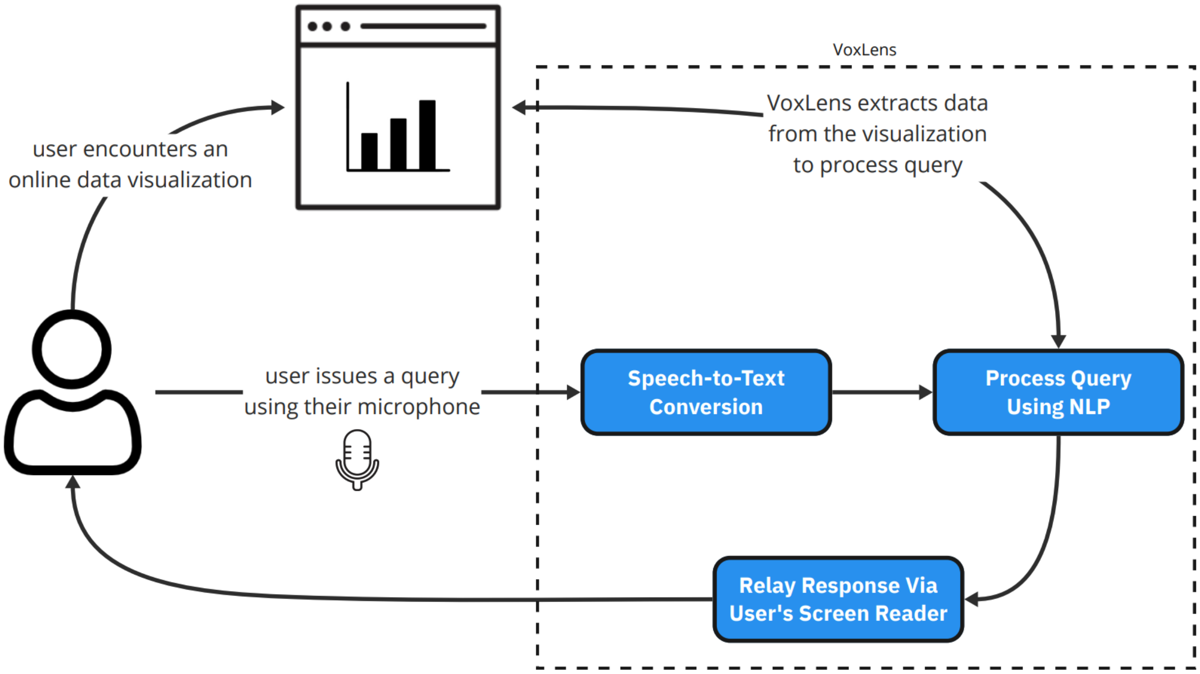


Figure 4.1: System diagram of VoxLens, showing each step from a screen-reader user issuing a query to getting a response.

Table 4.2: Keyboard shortcuts for VoxLens' interaction modes and preliminary commands. `Modifier Keys` for Windows and MacOS were `Control+Shift` and `Option`, respectively.

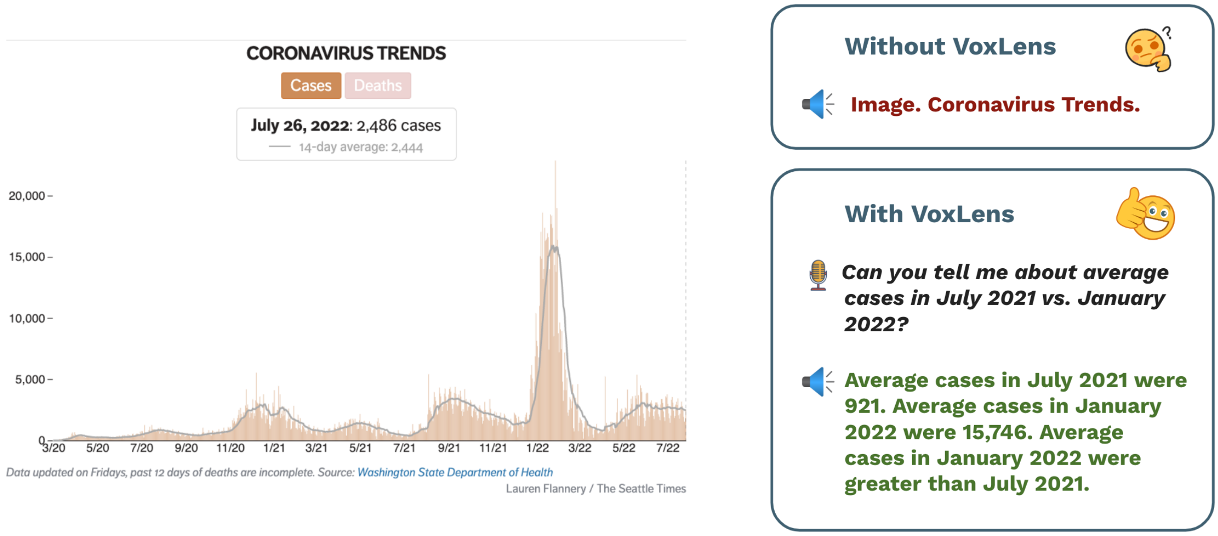
	Keyboard Shortcuts
Question-and-Answer Mode	`Modifier Keys + A` / `Modifier Keys + 1`
Summary Mode	`Modifier Keys + S` / `Modifier Keys + 2`
Sonification Mode	`Modifier Keys + M` / `Modifier Keys + 3`
Repeat Instructions	`Modifier Keys + I` / `Modifier Keys + 4`
Pause Output	`Modifier Keys + P` / `Modifier Keys + 5`

ble 4.1). I used the findings from these studies, together with the results from my prior work (Sharif et al. [2021]) presented in Chapter 3, to inform design decisions when iteratively building VoxLens. In the studies, I, the "wizard," simulated the auditory responses from a hypothetical screen reader.

Specifically, I asked the participants to "interact" with different modalities, including summary, sonification, and tables, to extract information from online data visualizations. I also asked participants to ask verbal questions from the visualization to obtain information. Subsequently, I briefly interviewed them in a semi-structured manner with open prompts at the end of their interactions with each modality and qualitatively analyzed their responses. Through these studies, I identified three interaction modes of VoxLens, which I present in the section below. Additionally, I show the system diagram of VoxLens in Figure 4.1.

4.2 Interaction Modes

Following the recommendations from my prior work (Sharif et al. [2021]), my goal was to enable screen-reader users to gain a holistic overview of the data and perform drilled-down explorations when needed. Therefore, I implemented three modes of interaction: (1) *Question-and-Answer* mode, where the user verbally interacts with the visualizations; (2) *Summary* mode, where VoxLens verbally offers a summary of the information contained in the visualization; and (3) *Sonification* mode, where VoxLens maps the data in the visualization to a musical scale, enabling listeners to interpret possible data trends. I iteratively built the features for these modes, seeking feedback from screen-reader users through Wizard-of-Oz studies. Users can activate these three modes of interaction by pressing their respective keyboard shortcuts (Table 4.2).



Without VoxLens

Image. Coronavirus Trends.

With VoxLens

Can you tell me about average cases in July 2021 vs. January 2022?

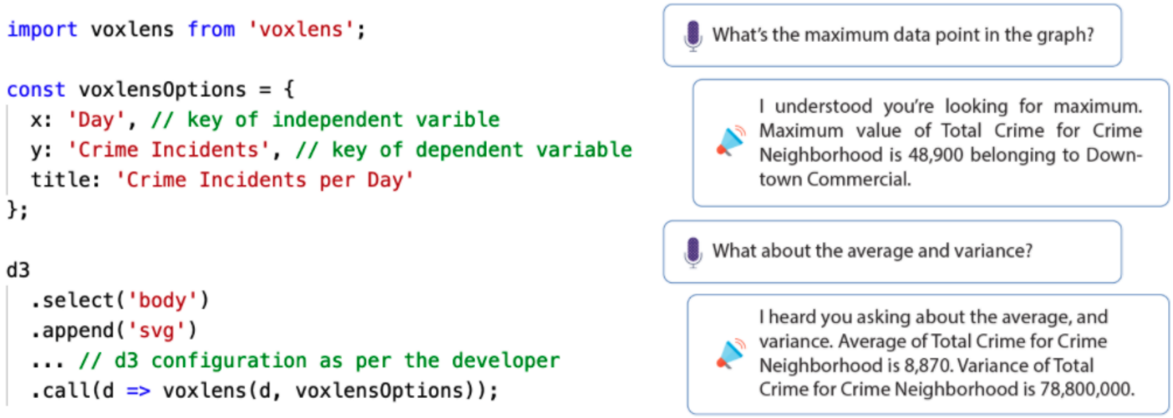
Average cases in July 2021 were 921. Average cases in January 2022 were 15,746. Average cases in January 2022 were greater than July 2021.

Figure 4.2: Interactions of a screen-reader user with a COVID-19 visualization from The Seattle Times with and without VoxLens.

4.2.1 Question-and-Answer Mode

In *Question-and-Answer* mode, screen-reader users can extract information from data visualizations by asking questions verbally through their microphones (see Figure 4.2). I used the Web Speech API (Natal et al. [2020]) and the P5 Speech library (DuBois [2017]) for speech input, removing the need for any additional software or hardware installation by the user. Through manual testing, I found that the P5 Speech library produced adequate results in recognizing speech from different accents, pronunciations, and background noise levels. After getting the text from the speech, I used an approximate string-matching algorithm from Hall and Dowling (Hall and Dowling [1980]) to recognize the commands. Additionally, I verified VoxLens' command recognition effectiveness through manual testing, using prior work's (Srinivasan et al. [2021]) data set on natural language utterances for visualizations.

The Wizard-of-Oz studies revealed that participants liked clear instructions and responses, integration with the user's screen reader, and the ability to query by specific terminologies. They specified having an interactive tutorial to become familiar with the tool, a help menu to determine which commands are supported, and the ability to include the user's query in the response as helpful features. Therefore, after recognizing the commands and processing their respective queries, VoxLens delivers a single combined



```
import voxlens from 'voxlens';

const voxlensOptions = {
  x: 'Day', // key of independent varible
  y: 'Crime Incidents', // key of dependent variable
  title: 'Crime Incidents per Day'
};

d3
  .select('body')
  .append('svg')
  ... // d3 configuration as per the developer
  .call(d => voxlens(d, voxlensOptions));
```

What's the maximum data point in the graph?

I understood you're looking for maximum. Maximum value of Total Crime for Crime Neighborhood is 48,900 belonging to Downtown Commercial.

What about the average and variance?

I heard you asking about the average, and variance. Average of Total Crime for Crime Neighborhood is 8,870. Variance of Total Crime for Crime Neighborhood is 78,800,000.

Figure 4.3: The code at left shows that integration of VoxLens requires only a single line of code. At right, I portray an example interaction with VoxLens using voice-activated commands.

Table 4.3: Voice-activated commands for VoxLens' *Question-and-Answer* mode.

Information Type	**Command**	**Aliases**
Extremum	Maximum	Highest
	Minimum	Lowest
Axis Labels and Ranges	Axis Labels	-
	Ranges	-
Statistics	Mean	Average
	Median	-
	Mode	-
	Variance	-
	Standard Deviation	-
	Sum	Total
Individual Data Point	[x-axis label] value	[x-axis label] data
Help	Commands	Instructions
	-	Directions, Help

response to the user via their screen readers. This approach enables screen-reader users to get a response to multiple commands as one single response. Additionally, I added each query as feedback in the response (Figure 4.3). For example, if the user said, “What is the maximum?”, the response was, “I heard you ask about the maximum. The maximum is...” For unidentified commands, the response was, “I heard you say [user input]. Command not recognized. Please try again.”

Screen-reader users can also get a list of supported commands by asking for the commands list. For example, the user can ask, “What are the supported commands?” to hear all of the commands that VoxLens supports. Table 4.3 presents the list of supported commands and their aliases.

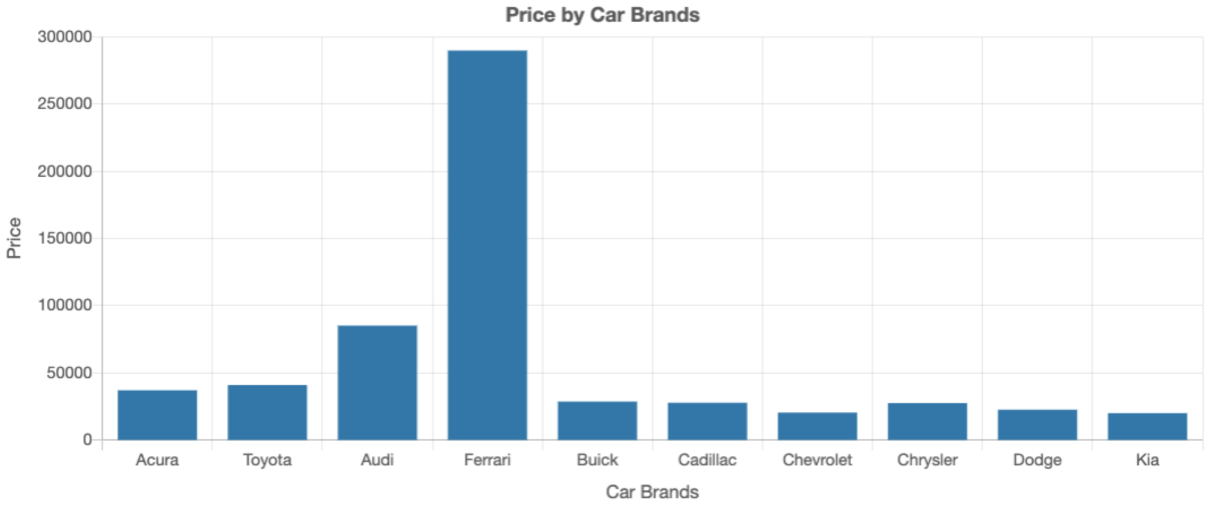


Figure 4.4: Sample visualization showing price by car brands.

4.2.2 Summary Mode

The Wizard-of-Oz studies, in line with the findings from my prior work (Sharif et al. [2021]), revealed that participants liked the preliminary exploration of the data. They suggested the information be personalized based on the user’s preferences; however, by default, it should only expose the minimum amount of information that a user would need to decide if they want to delve further into the data exploration. To delve further, they commonly seek the title, axis labels and ranges, maximum and minimum data points, and the average in online data visualizations. (The title and axis labels are required configuration options for VoxLens from creators. Axis ranges, maximum and minimum data points, and average are computed by VoxLens.)

At the same time, screen-reader users preferred concisely stated information. Therefore, the goal for VoxLens' *Summary* mode was to generate a summary only to provide the foundational holistic information about the visualization and not serve as a replacement for the visualization itself. I used the "language of graphics" (Bertin [1983]) through a pre-defined sentence template, identified as *Level 1* by Lundgard *et al.* (Lundgard and Satyanarayan [2021]), to decide the sentence structure. I altered the sentence template to include my findings from the Wizard-of-Oz studies. The final sentence template was:

```
Graph with title:  [title].  The X-axis is [x-axis title].  The
Y-axis is [y-axis title] and ranges from [range minimum] to
[range maximum].  The maximum data point is [maximum y-axis
value] belonging to [corresponding x-axis value], and the
minimum data point is [minimum y-axis value] belonging to
[corresponding x-axis value].  The average is [average].
```

For example, here is a generated summary of a data visualization depicting the prices of various car brands (Figure 4.4):

```
Graph with title:  Price by Car Brands.  The X-axis is car
brands.  The Y-axis is price and ranges from $0 to $300,000.
The maximum data point is $290,000 belonging to Ferrari, and
the minimum data point is $20,000 belonging to Kia.  The
average is $60,000.
```

As noted by prior work (Sharif et al. [2021]; Lundgard and Satyanarayan [2021]), the preference for information varies from one individual to another. Therefore, I explored customization to generate the summary, catering to the individual needs of screen-reader users; I present my work in Chapter 9.

4.2.3 Sonification Mode

For *Sonification* mode, the Wizard-of-Oz participants liked the ability to explore the data trend and identify the maximum and the minimum data points. Therefore, VoxLens' sonification mode presents screen-reader users with a sonified response (also known as an "audio graph" (Smus [2013])) by mapping the data in the visualization to a musical scale. A sonified response enables the listeners to interpret the data trend or pattern and gain a big-picture perspective of the data that is not necessarily achievable otherwise (Sawe et al. [2020]). To generate the sonified response, I utilized Tone.js (Mann [2020]), a JavaScript library that offers a variety of customizable options to produce musical notes. My goal was to enable the listeners to distinguish between data points directionally and to interpret the overall data trend.

Varying tonal frequency is more effective at representing trends than varying amplitude (Flowers [2005]; Hu et al. [2020]). Therefore, I mapped each data point to a frequency between 130 and 650 Hz, calculated based on its magnitude. For example, for the minimum data point, the frequency was 130 Hz; for the maximum data point, it was 650Hz; the intermediate data points were assigned values linearly in-between, similar to prior work (Ciuha et al. [2010]; Ohshiro et al. [2021]). Additionally, similar to design choices made by Ohshiro *et al.* (Ohshiro et al. [2021]), I used the sound of a sawtooth wave to indicate value changes along the x-axis. These approaches enabled us to distinctively differentiate between data values directionally, especially values that were only minimally different from each other. I chose this range based on the frequency range of the human voice (Titze and Martin [1998]; Baken and Orlikoff [2000]; McGookin and Brewster [2006]), and by trying several combinations ourselves, finding a setting that was comfortable for human ears. I open-sourced my sonification library on my GitHub repository[1].

I used the three common chart types (bar, scatter, and line) (Saket et al. [2018]), following prior work (Sharif et al. [2021]). All of these chart types use a traditional Cartesian coordinate system. Therefore, VOXLENS' sonified response is best applicable to graphs represented using a Cartesian plane.

4.3 Interaction Experience

A pain point for screen-reader users when interacting with online data visualizations is that most visualization elements are undiscoverable and incomprehensible to screen readers (Sharif et al. [2021]). In building VOXLENS, I ensured that the visualization elements were recognizable and describable by screen readers. Hence, as the very first step, when the screen reader encounters a visualization created with VOXLENS, the following is read to users:

```
    Bar graph with title:  [title].  To listen to instructions on
```

---

[1] https://github.com/athersharif/sonifier

```
how to interact with the graph, press Control + Shift + I or
Control + Shift + 4.  Key combinations must be pressed all
together and in order.
```

The modifier keys (`Control + Shift` on Windows, and `Option` on MacOS) and command keys were selected to not interfere with the dedicated key combinations of the screen reader, the Google Chrome browser, and the operating system. Each command was additionally assigned a numeric activation key, as per suggestions from the participants.

When a user presses the key combination to listen to the instructions, their screen reader reads:

```
To interact with the graph, press Control + Shift + A or
Control + Shift + 1 all together and in order.  You'll hear a
beep sound, after which you can ask a question such as, "What
is the average?"  or "What is the maximum value in the graph?"
To hear the textual summary of the graph, press Control + Shift
+ S or Control + Shift + 2.  To hear the sonified version of
the graph, press Control + Shift + M or Control + Shift + 3.
To repeat these instructions, press Control + Shift + I or
Control + Shift + 4.  Key combinations must be pressed all
together and in order.
```

At this stage, screen-reader users can activate the question-and-answer mode, listen to the textual summary, play the sonified version of the data contained in the visualization, or hear the instructions again. Activating the question-and-answer mode plays a beep sound, after which the user can ask a question in a free-form manner, without following any specific grammar or sentence structure. They can also ask for multiple pieces of information in no particular order. For example, in a visualization containing prices of cars by car brands, a screen-reader user may ask:

```
Tell me the mean, maximum, and standard deviation.
```

The response from VoxLens would be:

```
I heard you asking about the mean, maximum, and standard
deviation.  The mean is $60,000.  The maximum value of price
for car brands is $290,000 belonging to Ferrari.  The standard
deviation is 30,000.
```

Similarly, users can hear the textual summary or play the sonified version by pressing the appropriate key combination, as discussed above.

4.4 Developer Configuration Options

The accessibility of online data visualizations depends on visualization creators and their knowledge and practice of accessibility standards. When an alt-text or description is not provided, the visualization is useless to screen-reader users. In cases where alternative text *is* provided, the quality and quantity of the text is also a developer's choice, which may or may not be adequate for screen-reader users. For example, a common unfortunate practice is to use the title of the visualization as its alternative text, which helps screen-reader users understand the topic of the visualization but does not help in understanding the content contained *within* the visualization. Therefore, VoxLens is designed to reduce the burden and dependency on developers to make accessible visualizations, keeping the interaction consistent, and independent of the visualization library used. Additionally, VoxLens is engineered to require only a single line of code, minimizing any barriers to its adoption (Figure 4.3). Full code samples for implementing a bar chart with D3, Google Charts, and ChartJS is illustrated in Appendices D, E, and F, respectively.

VoxLens supports the following configuration options: "x" (name of the independent variable), "y" (name of the dependent variable), "title" (title of the visualization), "xLabel" (label for x-axis), and "yLabel" (label for y-axis). "x," "y," and "title" are required parameters, whereas the "xLabel" and "yLabel" are optional and default to the key names of "x" and "y," respectively. VoxLens allows visualization creators to set the values of these configuration options, as shown in Figure 4.3. I added more configuration options to support new functionalities in each VoxLens version; I discuss these additions in their respective chapters.

4.5 Channeling VoxLens' Output to Screen Readers

One challenge I faced was channeling the auditory response from VoxLens to the screen readers. As noted by the participants during Wizard-of-Oz studies, screen-reader users have unique preferences for their screen readers, including the voice and speed of the speech. Therefore, it was important for VoxLens to utilize these preferences, providing screen-reader users with a consistent, familiar, and comfortable experience. To relay the output from VoxLens to the screen reader, I created a temporary `div` element only visible to screen readers and positioned it off-screen, following WebAIM's recommendations (WebAIM [1999]).

Then, I added the appropriate Accessible Rich Internet Applications (ARIA) attributes (W3C [2006]) to the temporary element to ensure maximum accessibility. ARIA attributes are a set of attributes to make web content more accessible to people with disabilities. Notably, I added the "aria-live" attribute, allowing screen readers to immediately announce the query responses that VoxLens adds to the temporary element. For MacOS, I had to additionally include the "role" attribute, with its value set to "alert." This approach enabled VoxLens to promptly respond to screen-reader users' voice-activated commands using their screen readers. After the response from VoxLens is read by the screen reader, a callback function removes the temporary element from the HTML tree to avoid overloading the HTML Document Object Model (DOM).

4.6 Additional Implementation Details

VoxLens relies on the Web Speech API (Natal et al. [2020]). Therefore, it is only fully functional on browsers with established support for the API, such as Google Chrome. JavaScript was naturally the choice of programming language for VoxLens, as VoxLens is a plug-in for JavaScript visualization libraries. Additionally, I used EcmaScript (Harband et al. [1999]) to take advantage of modern JavaScript features such as destructured assignments, arrow functions, and the spread operator. I also built a testing tool to test VoxLens on data visualizations, using the React (Inc. [2013]) framework as the user-interface framework and Node.js (Foundation [2009]) as the back-end server—both of which also use JavaScript as their underlying programming language. Additionally, I used GraphQL (Foundation [2021]) as the API layer for querying and connecting with the Postgres (Group [1996]) database, which I used to store data and participants' interaction logs.

Creating a tool like VOXLENS requires significant engineering efforts. VOXLENS' GitHub repository[2] has seven releases and over 100,000 lines of developed code, excluding comments. The repository also includes six "forks"[3] from other contributors with additional releases and lines of codes not included in the numbers reported in the previous statement. To support testing VOXLENS on various operating systems and browsers with different screen readers, I collected 30 data sets of varying data points, created their visualizations using Google Charts, D3, and ChartJS, integrated VOXLENS with each of them, and deployed a "playground" testing website on my server. The testing website was instrumental in ensuring the correct operation of VOXLENS under various configurations, bypassing the challenges of setting up a development environment for testers.

4.7 Conflicts with Other Plug-Ins

To the best of my knowledge, two kinds of conflicts are possible with VOXLENS: *key combination* conflicts and *ARIA attribute* conflicts. As mentioned earlier, I selected default key combinations to avoid conflicts with the dedicated combinations of the screen reader, the Google Chrome browser, and the operating system. However, some users might have external plug-ins that use key combinations conflicting with those from VOXLENS. In such cases, VOXLENS provides visualization creators to alter the modifier keys via the configuration options. Work is underway to enable screen-reader users to modify the modifier keys and the key combinations, providing them with a more granular control over their information consumption experiences with online data visualizations.

VOXLENS modifies the "aria-label" attribute of the visualization container element to describe the interaction instructions for VOXLENS. Another plug-in may intend to modify the "aria-label" attribute as well, in which case the execution order of the plug-ins will determine which plug-in achieves the final override. The execution order of the plug-ins depends on several external factors (Picazo-Sanchez et al. [2020]), and is, unfortunately, a common limitation for any browser plug-in. However, VOXLENS does not affect the "aria-labelledby" attribute, allowing other systems to gracefully override the "aria-label" attribute set by VOXLENS, as this attribute takes precedence over the "aria-label" attribute in the accessibility tree. Fur-

---

[2] https://github.com/athersharif/voxlens

[3] A fork is a new repository that shares code and visibility settings with the original repository.

thermore, at present, I am experimenting with the "longdesc" HTML attribute to devise a more comfortable solution that improves the experiences of screen-reader users in extracting information while also reducing the possibility of a conflict with other plug-ins.

It is important to note that VOXLENS's sonification library is supplied independently from the main VOXLENS plug-in and does not follow the same limitations. My testing did not reveal any conflicts between the sonification library with other plug-ins.

Table 4.4: VOXLENS versions presented in this dissertation, organized by their version number, description, and the chapters in which they are introduced.

Version	Chapter(s)	Description
1.0	4 and 5	Supports holistic information extraction from simple and single-series visualizations (*e.g.,* bar chart, scatter plot, line graph)
2.0	7	Supports drilled-down information extraction from complex multi-series visualizations (*e.g.,* geospatial map, multi-series line graph)
3.0	8	Supports extraction of information about data uncertainty from online data visualizations
4.0	9	Supports integration with VOXEX, a system that enables screen-reader users to customize the information they consume from online data visualizations
5.0	10	Supports visualization creators through implementation and integration of four technological interventions

4.8 VOXLENS Versions

This chapter introduced the first version of VOXLENS (version 1.0), setting up the foundation for subsequent versions presented in succeeding chapters of this dissertation. Table 4.4 shows the five VOXLENS versions offered in this dissertation.

4.9 Supporting Additional Visualization Types

In this foundational VOXLENS version, I supported simple online data visualizations (*i.e.,* bar charts, scatter plots, line graphs) created using single-series data. In subsequent versions of VOXLENS, present in the succeeding chapters of this dissertation, I extended VOXLENS' support to multi-series line graphs and geospatial maps. Three factors should be considered to expand the utility of VOXLENS to support additional visualization types: (1) data series, (2) visual encodings, and (3) context.

4.9.1 Data Series ("Content")

VOXLENS currently supports both single- and multi-series data. Therefore, VOXLENS is functional for any data visualization regardless of the visualization type, provided that the data is two-dimensional. For supporting *n*-dimensional data (where $n > 2$), user studies are necessary to understand the interactions of screen-reader users to extract information from the respective data dimensionality. Chapter 7 presents an approach to conducting these studies.

4.9.2 Visual Encodings ("Container")

At present, VOXLENS supports two-dimensional bar charts, line graphs, scatter plots, and geospatial maps. Other data visualization types (*e.g.,* heatmaps, area charts) warrant similar investigations present in this dissertation to understand how visual encodings in these data visualizations can be communicated to screen-reader users effectively.

For example, if a developer wishes to extend VOXLENS' support to heatmaps, they can first investigate how screen-reader users interact with heatmaps, the information they seek, and identify encodings and specifications that they wish to relay to the users. Then, they can add code to the VOXLENS repository in the appropriate places to implement their findings. These additions will enable VOXLENS to activate the heatmap "module" to process queries when visualization creators specify "heatmap" as the value of the `chartType` configuration option. The implementations for different visualization types may differ depending on the findings from the initial investigation. Chapter 7 provides an example of extending VOXLENS to new visualization types, including geospatial maps. Additionally, the documentation in VOXLENS' repository could further assist developers in onboarding new visualization types.

4.9.3 Context

VoxLens is currently not a context-aware system. This limitation is further discussed in Chapter 11 of this dissertation. For data visualizations that are context-dependent (*e.g.,* weather data visualizations), an enhancement to VoxLens' query identification and information retrieval process may be needed. Additionally, these enhancements would need to follow a user-centered iterative design protocol to ensure acceptable levels of usability and usefulness for screen-reader users.

4.10 Summary

In this chapter, I presented the design and implementation details of a novel interaction technique, VoxLens, for screen-reader users to extract information from online data visualizations. VoxLens is a multi-modal open-source JavaScript plug-in (as opposed to a programming language or a standalone software) that improves the accessibility of online data visualizations for screen-reader users. It is limited to the visualization types discussed in this chapter; extending its support to additional visualizations is also discussed. The subsequent chapters in this dissertation introduce additional versions of VoxLens.

Chapter 5

Holistic Information Extraction*

Chapter 4 presented the core design and functionalities of VoxLens, a JavaScript plug-in I created to make online data visualizations accessible by following the findings and recommendations from my prior work presented in Chapter 3. This chapter describes my evaluation of VoxLens' performance based on the accuracy of extracted information and interaction times. I note that this evaluation is for the VoxLens version described in Chapter 4 (version 1.0). Subsequent chapters provide the design and evaluation of their respective VoxLens versions.

5.1 Motivation

Online data visualizations are a widely used source for experts and non-experts alike to explore and analyze simple as well as complex data. They assist people in extracting information effectively and efficiently, taking advantage of the ability of the human mind to recognize and interpret visual patterns (Marriott et al. [2021]). However, the visual nature of data visualizations inherently disenfranchises screen-reader users, who may not be able to see or recognize visual patterns (Marriott et al. [2021]; Lee et al. [2021]).

Due to the inaccessibility of data visualizations, screen-reader users cannot access the underlying information in these visualizations, and oftentimes, at all. Even when the data visualization includes basic accessibility measures (e.g., alternative text or a data table), screen-reader users still spend 211% more time interacting with online data visualizations and answer questions about the data in the visualizations 61%

---

*Parts of this chapter are adapted from Sharif et al. [2022f].

less accurately, compared to non-screen-reader users (Sharif et al. [2021]). Screen-reader users rely on the creators of visualizations to provide adequate alternative text, which is often incomplete. Additionally, they have to remember and process more information mentally than is often humanly feasible (Thiele et al. [2011]), such as when seeking the maximum or minimum value in a chart.

My prior work studied the experiences of screen-reader users with online data visualizations, highlighting the challenges they face, the information they seek, and the techniques and strategies that could make online data visualizations more accessible (Sharif et al. [2021]). Building on this work, I developed a novel interactive solution to support screen-reader users in effectively interacting with online data visualizations called "VoxLens"; I present its core design and functionality in Chapter 4. VoxLens provides screen-reader users with a multi-modal solution that supports three modes of interaction: (1) *Question-and-Answer mode*, where the user verbally interacts with the visualizations on their own; (2) *Summary mode*, where VoxLens describes the summary of the information contained in the visualization; and (3) *Sonification mode*, where VoxLens maps the data in the visualization to a musical scale, enabling listeners to interpret the data trend.

To assess the performance of VoxLens and its three modes, I conducted a task-based experiment with follow-up interviews with screen-reader users. My evaluation of VoxLens shows that with VoxLens, compared to without it, screen-reader users improved their accuracy of extracting information by 122% and reduced their overall interaction time by 36%.

5.2 Study Design

I evaluated the performance of VoxLens using a mixed-methods approach. Specifically, I conducted an online mixed-factorial experiment with screen-reader users to assess VoxLens quantitatively. Additionally, I conducted follow-up interviews with the participants for a qualitative assessment of VoxLens. I direct the users to Chapter 4 for the core design and functionalities of VoxLens, which I evaluated in this study.

Table 5.1: Screen-reader participants, their gender identification, age, screen reader, vision-loss level, and diagnosis. Under the "G" (*Gender*) column, M = *Man*, W = *Woman*, NB = *Non-binary*, and "-" means preferred not to disclose.

	G	**Age**	**Screen Reader**	**Vision-Loss Level**	**Diagnosis**
P1	M	28	NVDA	Blind since birth, Complete blindness	Optic Nerve Hypoplasia
P2	M	61	JAWS	Complete blindness, Lost vision gradually	Optic Neuropathy
P3	M	48	JAWS	Complete blindness, Lost vision gradually	Leber Congenital Amaurosis
P4	F	29	NVDA	Blind since birth, Complete blindness	Optic Nerve Hypoplasia and Glaucoma
P5	F	37	JAWS	Blind since birth, Complete blindness	Leber Congenital Amaurosis
P6	F	51	JAWS	Blind since birth, Complete blindness	Retinopathy of Prematurity
P7	M	58	JAWS	Complete blindness, Lost vision gradually	Glaucoma
P8	M	30	NVDA	Blind since birth, Complete blindness	Leber Congenital Amaurosis
P9	F	64	JAWS	Complete blindness, Lost vision gradually	Retinitis Pigmentosa
P10	F	68	Fusion	Lost vision gradually, Partial blindness	Stargaart's Maculopathy
P11	F	73	JAWS	Complete blindness, Lost vision gradually	Retinitis Pigmentosa
P12	F	64	JAWS	Complete blindness, Lost vision gradually	Cataracts
P13	M	18	NVDA	Complete blindness	Brain Tumor
P14	M	36	JAWS	Blind since birth, Complete blindness	Leber Congenital Amaurosis
P15	M	25	NVDA	Lost vision gradually, Partial vision	Retinopathy of Prematurity and Subsequent Cataracts

P16	M	42	JAWS	Blind since birth, Complete blindness	Microphthalmia
P17	M	68	JAWS	Complete blindness, Lost vision gradually	Detached Retinas
P18	F	31	NVDA	Blind since birth, Complete blindness	Retinopathy of Prematurity
P19	F	47	JAWS	Complete blindness, Lost vision gradually	Optic Neuropathy
P20	M	48	NVDA	Complete blindness, Lost vision gradually	Retinitis Pigmentosa
P21	M	43	NVDA	Complete blindness, Lost vision gradually	Retinitis Pigmentosa
P22	M	19	NVDA	Blind since birth, Complete blindness	Retinopathy of Prematurity

5.2.1 Participants

The participants (see Table 5.1) were 22 screen-reader users, recruited using word-of-mouth, snowball sampling, and email distribution lists for people with disabilities. Nine participants identified as women and 13 as men. Their average age was 45.3 years (SD=16.8). Twenty participants had complete blindness and two participants had partial blindness; nine participants were blind since birth, 12 lost vision gradually, and one became blind due to a brain tumor. The highest level of education attained or in pursuit was a doctoral degree for two participants, a master's degree for seven participants, a bachelor's degree for eight participants, and a high school diploma for the remaining five participants. Estimated computer usage was more than 5 hours per day for 12 participants, 2–5 hours per day for eight participants, and 1–2 hours per day for two participants. The average frequency of interacting with online data visualizations was over two visualizations per day, usually in the context of news articles, blog posts, and social media.

For the task-based experiment and questionnaire, I compensated participants with a $20 Amazon gift card for 30–45 minutes of their time. For the follow-up interview, they were compensated $10 for 30 minutes of their time. No participant was allowed to partake in the experiment more than once.

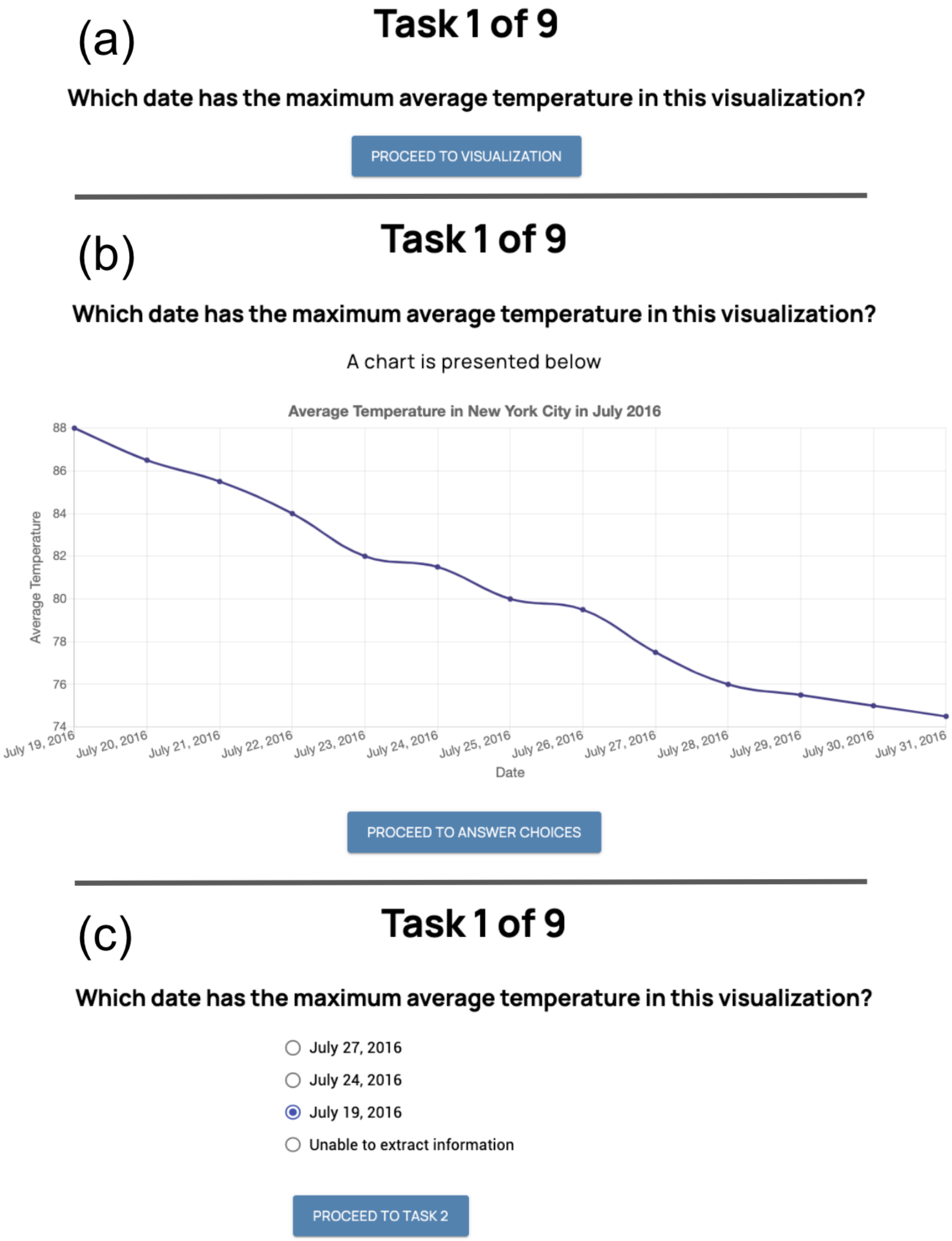
5.2.2 Apparatus

I conducted the task-based experiment online using a user study platform I created using the JavaScript React framework (Inc. [2013]). I tested the platform myself and with screen-reader users, both with and without a screen reader, ensuring maximum and proper accessibility measures. I deployed the experiment platform as a website hosted on my server.

I analyzed the performance of VoxLens comparing the data collected from the task-based experiments with that from my prior work (Sharif et al. [2021]) [see Chapter 3]. For a fair comparison to my prior work, I used the same visualization libraries, visualization data sets, question categories, and complexity levels. I chose the visualization libraries (Google Charts, ChartJS, and D3) based on the variation in their underlying implementations as well as their application of accessibility measures. Google Charts utilizes SVG elements to generate the visualization and appends a tabular representation of the data for screen-reader users by default; D3 also makes use of SVG elements but does not provide a tabular representation; ChartJS uses HTML Canvas to render the visualization as an image and relies on the developers to add alternative text ("alt-text") and Accessible Rich Internet Applications ("ARIA") attributes (W3C [2006]). Therefore, each of these visualization libraries provides a different experience for screen-reader users, as highlighted in my prior work (Sharif et al. [2021]).

I implemented the visualizations following the WCAG 2.0 guidelines (Caldwell et al. [2008]) in combination with the official accessibility recommendations from the visualization libraries. For ChartJS, I added the "role" and "aria-label" attributes to the "canvas" element. The "role" attribute had the value of "img," and the "aria-label" was given the value of the visualization title, as per the official documentation from ChartJS developers (ChartJS [2015]). I did not perform any accessibility scaffolding for Google Charts and D3 visualizations, as these visualizations rely on a combination of internal implementations and the features of SVG for accessibility. My goal was to replicate an accurate representation of how these visualizations currently exist on the Web.

Prior work (Kim et al. [2020]) has reported that the non-visual questions that users ask from graphs mainly comprise compositional questions, similar to the findings from Brehmer and Munzner's task topology (Brehmer and Munzner [2013]). Therefore, the question categories comprised one "Search" action (*lookup and locate*) and two "Query" actions (*identify and compare*), similar to prior work (Brehmer et al.



Task 1 of 9

Which date has the maximum average temperature in this visualization?

Task 1 of 9

Which date has the maximum average temperature in this visualization?

A chart is presented below

Task 1 of 9

Which date has the maximum average temperature in this visualization?

- ○ July 27, 2016
- ○ July 24, 2016
- ● July 19, 2016
- ○ Unable to extract information

Figure 5.1: Participants were shown three pages for each task. **(a)** Page 1 presented the question to explore. **(b)** Page 2 displayed the same question and a visualization. **(c)** Page 3 showed the question again with a set of four multiple choice responses.

[2018]). The categories, in ascending order of difficulty, were: (1) *Order Statistics* (extremum); (2) *Symmetry Comparison* (data points comparison); and (3) *Chart Type-Specific Questions* (value retrieval for bar charts, trend summary for line charts, and correlation for scatter plots). As in prior work (Sharif et al. [2021]), all questions were multiple-choice with four choices: the correct answer, two incorrect answers, and the option for "Unable to extract information." I randomized the order of these choices per trial.

5.2.3 Procedure

The study was conducted online without direct supervision. The study comprised six stages. The first stage displayed the study purpose, eligibility criteria, and the statement of IRB approval. In the second stage, the participants were asked to complete a pre-study questionnaire to record their demographic information, screen-reader software, vision-loss level, and diagnosis (see Table 5.1). Additionally, I asked the participants about their education level, daily computer usage, and frequency of interacting with visualizations.

In the third stage, participants were presented with a step-by-step interactive tutorial to train and familiarize themselves with the modes, features, and commands that VOXLENS offers. Additionally, participants were asked questions at each step to validate their understanding. On average, the tutorial took 12.6 minutes (SD=6.8) to complete. Upon successful completion of the tutorial, participants were taken to the fourth stage, which displayed the instructions for completing the study tasks.

In the fifth stage, each participant was given nine tasks. For each task, participants were shown three web pages: Page 1 contained the question to explore, page 2 displayed the question and visualization, and page 3 presented the question with four multiple-choice responses. Figure 5.1 shows the three pages of an example task. After the completion of the tasks, I asked the participants to fill out the NASA-TLX (Hart and Staveland [1988]) survey in the last stage. An entire study session ranged from 30-45 minutes in duration.

5.2.4 Design & Analysis

The experiment was a $2 \times 3 \times 3 \times 3$ mixed-factorial design with the following factors and levels:

- *VoxLens* (VX), between-Ss.: {yes, no}
- *Visualization Library* (VL), within-Ss.: {ChartJS, D3, Google Charts}

- *Data Complexity* (CMP), within-Ss.: {Low, Medium, High}

- *Question Difficulty* (DF), within-Ss.: {Low, Medium, High}

For the screen-reader users who did not use VoxLens (VX=*no*), I used my prior work's data (Sharif et al. [2021]) (N=36) as a baseline for comparison. There was no overlap of participants between these two user groups to facilitate between-subjects comparisons.

The two dependent variables were *Accuracy of Extracted Information* (AEI) and *Interaction Time* (IT). I used a dichotomous representation of AEI (i.e., "inaccurate" or 0 if the user was unable to answer the question correctly, and "accurate" or 1 otherwise) for the analysis. I used a mixed logistic regression model (Gilmour et al. [1985]) with the above factors, interactions with VoxLens, and a covariate to control for Age. I also included $Subject_r$ as a random factor to account for repeated measures. The statistical model was therefore $AEI \leftarrow VX + VX{\times}VL + VX{\times}CMP + VX{\times}DF + Age + Subject_r$. I did not include factors for VL, CMP, or DF because the research questions centered around VoxLens (VX), and my interest in these factors only extended to their possible interactions with VoxLens.

For *Interaction Time* (IT), I used a linear mixed model (Frederick [1999]; Littell et al. [1998]) with the same model terms as for AEI. IT was calculated as the total time of the screen reader's focus on the visualization element. Participants were tested over three *Visualization Library* × *Complexity* ($VL{\times}CMP$) conditions, resulting in $3{\times}3 = 9$ trials per participant. With 21 participants, a total of $21{\times}9 = 189$ trials were produced and analyzed for this study. One participant, who was unable to complete the tutorial, was excluded from the analysis.

5.2.5 Qualitative and Subjective Evaluation

To qualitatively assess the performance of VoxLens, I conducted follow-up interviews with six screen-reader users, randomly selecting them from the pool of participants who completed the task-based experiment. Similar to prior work (Wyche and Grinter [2009]), I ceased recruitment of participants once I reached saturation of insights.

To analyze the interviews, I used thematic analysis (Braun and Clarke [2006]) guided by a semantic approach (Patton [1990]). I used two interviews to develop an initial set of codes, resulting in a total of 23 open codes. Two co-authors and I coded each interview transcript independently, resolving the disagree-

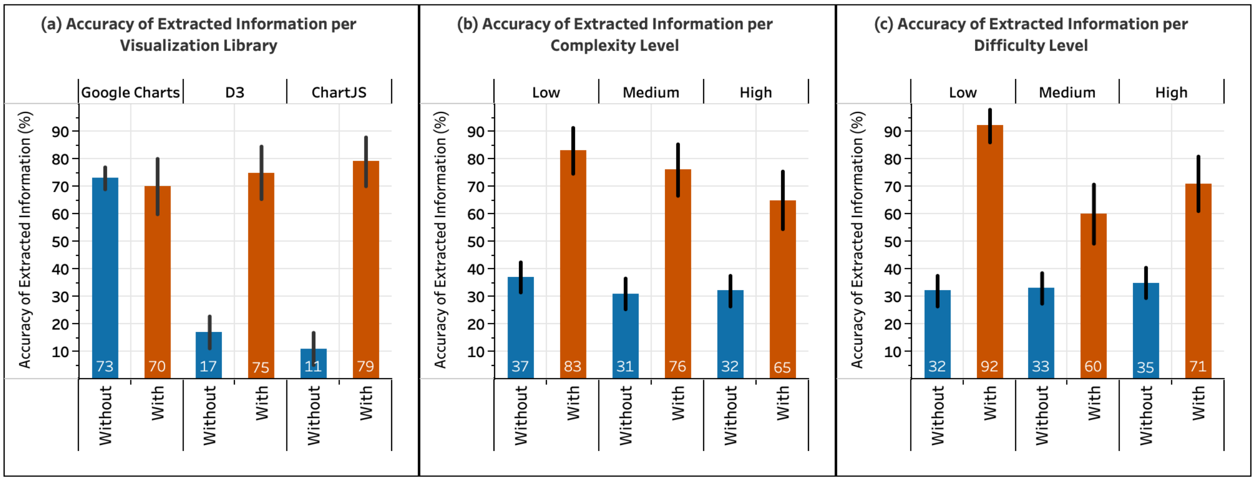


Figure 5.2: Accuracy of Extracted Information (AEI), as a percentage, for screen-reader users without (N=36) and with (N=21) VoxLens, by **(a)** Visualization Library, **(b)** Complexity Level, and **(c)** Difficulty Level. The percentage represents the "accurate" answers. Therefore, higher is better. Error bars represent mean ± standard deviation.

ments through mutual discussions. As suggested by Lombard *et al.* (Landis and Kock [1977]), I calculated inter-rater reliability (IRR) using pairwise percentage agreement together with Krippendorff's α (Krippendorff [2011]). To calculate pairwise percentage agreement, I calculated the average pairwise agreement among the three rater pairs across observations. The pairwise percentage agreement was 94.3%, showing a high agreement between raters. Krippendorff's α was calculated using ReCal (Freelon [2010]) and found to be 0.81, indicating a high level of reliability (Krippendorff [2018]).

In addition to conducting follow-up interviews, I administered the NASA-TLX survey (Hart and Staveland [1988]) with all participants (N=21) to assess the perceived workload of VoxLens.

5.3 Results

In this section, I present the experiment results from the task-based experiment using quantitative measures. I also report the interview results and subjective ratings from the NASA-TLX questionnaire (Hart and Staveland [1988]).

Table 5.2: Numerical results for the $N = 513$ questions asked of screen-reader users with and without VoxLens for each level of *Visualization Library*, *Complexity Level*, and *Difficulty Level*. N is the total number of questions asked, AA is the number of "accurate answers," and $AA(\%)$ is the percentage of "accurate answers."

	Without **VoxLens**			***With*** **VoxLens**		
	N	AA	AA **(%)**	N	AA	AA **(%)**
Overall	324	109	34%	189	141	75%
Visualization Library (VL)						
ChartJS	108	12	11%	63	50	79%
D3	108	18	17%	63	47	75%
Google Charts	108	79	73%	63	44	70%
Complexity Level (CMP)						
Low	108	40	37%	63	52	83%
Medium	108	34	31%	63	48	76%
High	108	35	32%	63	41	65%
Difficulty Level (DF)						
Low	108	35	32%	63	58	92%
Medium	108	36	33%	63	38	60%
High	108	38	35%	63	45	71%

Table 5.3: Summary results from 57 screen-reader users with (N=21) and without (N=36) VoxLens. "VL" is the visualization library, "CMP" is data complexity, and "DF" is question difficulty. Cramer's V is a measure of effect size (Ferguson [2016]). All results are statistically significant ($p < .05$) or marginal ($.05 \leq p < .10$).

	N	χ^2	p	**Cramer's** V
VX(VoxLens)	57	38.16	< .001	.14
$VX \times VL$	57	82.82	< .001	.20
$VX \times CMP$	57	8.90	.064	.07
$VX \times DF$	57	17.95	.001	.09
Age	57	3.58	.058	.04

5.3.1 Quantitative Results

I present the quantitative experiment results using the *Accuracy of Extracted Information* (AEI) and *Interaction Time* (IT) for screen-reader users with and without VoxLens.

5.3.1.1 Accuracy of Extracted Information

The results show a significant main effect of *VoxLens (VX)* on AEI (χ^2(1, N=57)=38.16, p<.001, Cramer's V=.14), with VoxLens users achieving 75% accuracy ($SD = 18.0\%$) and non-VoxLens users achieving only 34% accuracy ($SD = 20.1\%$). This difference constituted a 122% improvement due to VoxLens.

By analyzing the *VoxLens (VX)* × *Visualization Library (VL)* interaction, I investigated whether changes in AEI were proportional across visualization libraries for participants in each VoxLens group. The $VX \times VL$ interaction was indeed statistically significant (χ^2(4, N=57)=82.82, p<.001, Cramer's V=.20). This result indicates that AEI significantly differed among visualization libraries for participants in each VoxLens group. Figure 5.2 and Table 5.2 show AEI percentages for different visualization libraries for each VoxLens group. Additionally, I report these findings in Table 5.3.

Prior work (Sharif et al. [2021]) reported a statistically significant difference between screen-reader users (SRU) and non-screen-reader users (non-SRU) in terms of AEI, attributing the difference to the inaccessibility of online data visualizations. I conducted a second analysis, investigating whether AEI was different between screen-reader users who used VoxLens and non-screen-reader users, to extract information from online data visualizations. Specifically, I investigated the effect of *SRU* on AEI and found a marginally significant effect ($p \approx$ .077). This result itself does not provide evidence in support of VoxLens closing the access gap between the two user groups; further experimentation is necessary to confirm or refute this marginal result. However, taken together with other results, this is an encouraging trend.

5.3.1.2 Interaction Time

Preliminary analysis showed interaction times as conditionally non-normal, determined using Anderson-Darling (Anderson and Darling [1954]) normality tests. To achieve normality, I applied logarithmic transformation before analysis, as is common practice for time measures (Hoyle [1973]; Limpert et al. [2001]; Berry [1987]). For ease of interpretation, interaction times plots show the original non-transformed values.

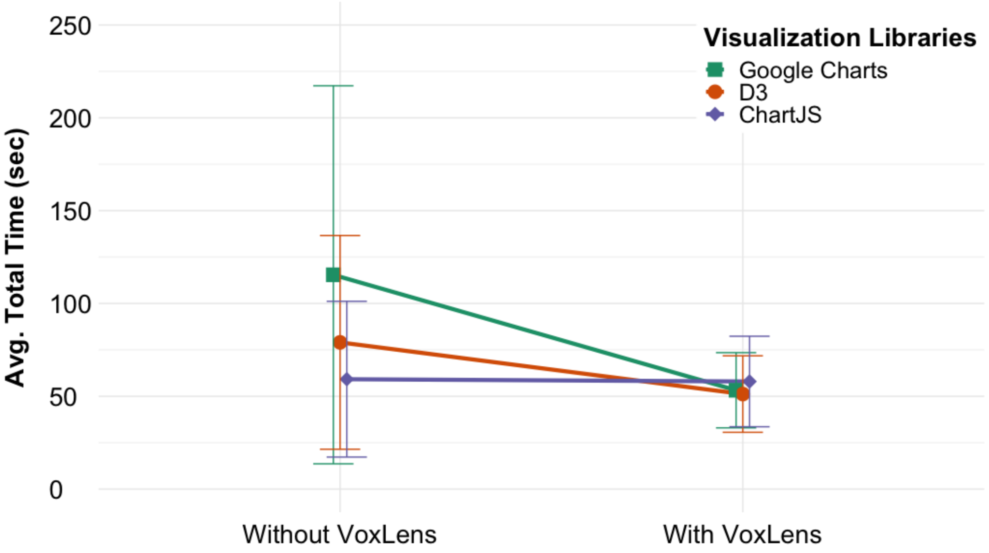
Table 5.4: Summary results from 57 screen-reader participants with (N=21) and without (N=36) VoxLens using a linear mixed model (Frederick [1999]; Littell et al. [1998]). "VL" is the visualization library, "CMP" is data complexity, and "DF" is question difficulty. Partial eta-squared (η_p^2) is a measure of effect size (Cohen [1973]). All results are statistically significant ($p < .05$) except $VX \times CMP$.

	df_n	df_d	F	p	η_p^2
VX (VoxLens)	4	54	12.66	.001	.19
$VX \times VL$	4	444	33.89	< .001	.23
$VX \times CMP$	4	444	1.85	.118	.02
$VX \times DF$	4	444	14.41	< .001	.12
Age	4	54	5.03	.029	.09

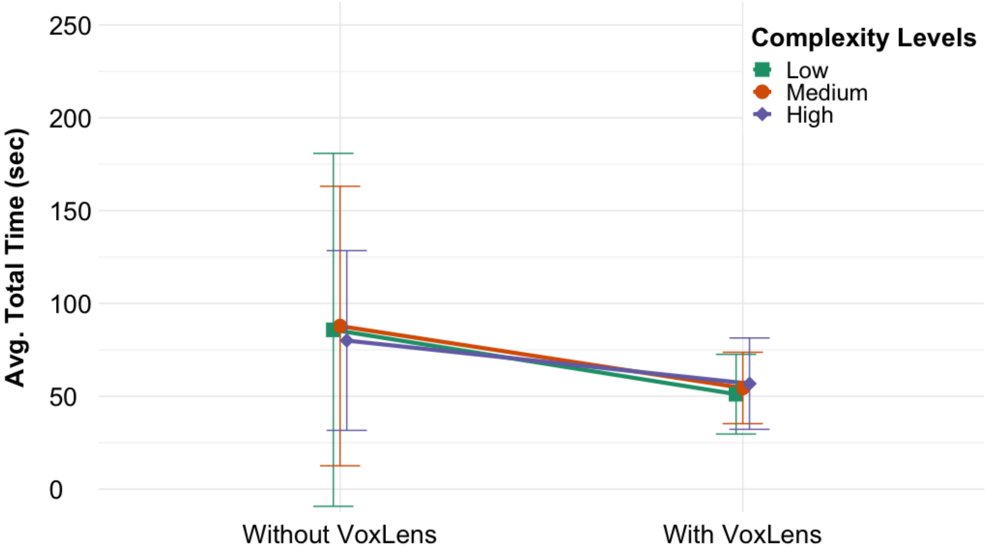
Table 5.5: Summary results from 57 screen-reader participants with (N=21) and without (N=36) VoxLens showing the numerical results for *Interaction Time* (IT), for each age range. N is the total number of participants for the given age range, *Mean* is the average IT in seconds, and SD represents the standard deviation.

	Both Groups			***Without* VoxLens**			***With* VoxLens**		
Age Range	N	$Mean$	SD	N	$Mean$	SD	N	$Mean$	SD
18-19	3	62.5	30.5	1	96.8	-	2	45.3	9.4
20-29	9	40.6	28.0	6	63.7	20.3	3	50.8	7.9
30-39	10	44.2	23.1	7	60.6	16.5	3	43.9	4.5
40-49	15	67.5	78.7	10	106.7	93.1	5	59.8	14.9
50-59	7	47.6	27.3	5	79.1	19.2	2	63.3	3.0
60-69	11	64.8	33.6	6	76.8	33.7	5	55.2	10.8
> 70	2	127.0	127.4	1	217.1	-	1	60.1	-

(a) Interaction Times per Visualization Library



(b) Interaction Times per Complexity Level



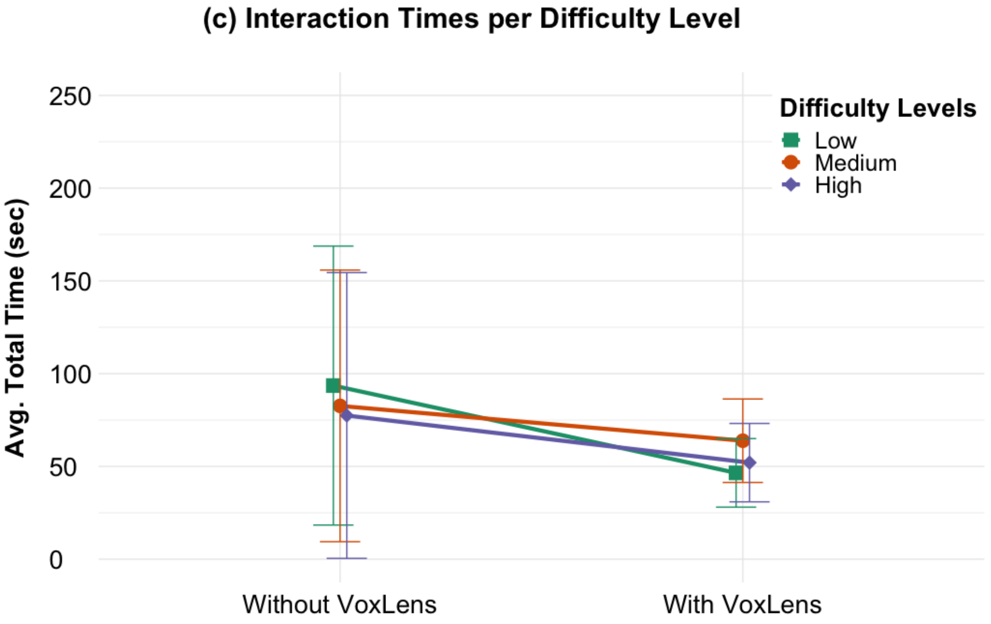


Figure 5.3: Interaction Times (IT), in seconds, for screen-reader users without (N=36) and with (N=21) VOXLENS by **(a)** Visualization Library (VL), **(b)** Data Complexity Level (CMP), and **(c)** Question Difficulty Level (DF). Error bars represent mean ± standard deviation. Lower is better (faster).

VoxLens (VX) had a significant main effect on *Interaction Time* (IT) ($F(4,54)=12.66$, $p<.05$, $\eta_p^2=.19$). Specifically, the average IT for non-VOXLENS users was 84.6 seconds (SD=75.2). For VOXLENS users, it was 54.1 seconds (SD=21.9), 36% lower (faster) than for participants without VOXLENS.

The *VX* × *VL* and *VX* × *DF* interactions were both significant ($F(4,444)=33.89$, $p<.001$, $\eta_p^2=.23$ and $F(444)=14.41$, $p<.001$, $\eta_p^2=.12$, respectively). Figure 5.3 shows interaction times across different visualization libraries, difficulty levels, and complexity levels for VOXLENS group. For VOXLENS users, all three visualization libraries resulted in almost identical interaction times. Figure 5.3 portrays larger variations in interaction times for users who did not use VOXLENS (data used from prior work (Sharif et al. [2021])) compared to VOXLENS users. I attribute these observed differences to the different underlying implementations of the visualization libraries.

I investigated the effects of *Age* on IT. *Age* had a significant effect on IT ($F(1,54)=5.03$, $p<.05$, $\eta_p^2=.09$), indicating that IT differed significantly across the ages of the participants, with participants aged 50 years or older showing higher interaction times by about 7% compared to participants under the age of

50. Table 5.5 shows the average IT for each age range by VoxLens group. Additionally, I report these findings in Table 5.4.

Similar to my investigation of the effect of screen-reader users (SRU) on AEI, I examined the main effect of SRU on IT. The results show that SRU had a significant effect on IT (F(4,54)=48.84, p<.001, η_p^2=.48), with non-screen-reader users performing 99% faster than VoxLens users.

5.3.2 Qualitative Results

To assess VoxLens qualitatively, I investigated the overall experiences of the participants with VoxLens, the features they found helpful, the challenges they faced during the interaction, and the improvements and future features that could enhance the performance of VoxLens. I identified five main results from analyzing the participants' feedback about VoxLens: (1) a positive step forward in making online data visualizations accessible, (2) interactive dialogue is one of the "top" features, (3) sonification helps in "visualizing" data, (4) data summary is a good starting point, and (5) one-size-fits-all is not the optimal solution. I present each of these in turn.

5.3.2.1 A Positive Step Forward in Making Online Data Visualizations Accessible

All participants found VoxLens an overall helpful tool to interact with and quickly extract information from online data visualizations. For example, S1 and S3 expressed their excitement about VoxLens:

> `I have never been able to really interact with graphs before online. So without the tool, I am not able to have that picture in my head about what the graph looks like. I mean, like, especially when looking up news articles or really any, sort of, like, social media, there's a lot of visual representations and graphs and pictographs that I don't have access to so I could see myself using [VoxLens] a lot. The tool is really great and definitely a positive step forward in creating accessible graphs and data.` (S1)

> `Oh, [VoxLens] was outstanding. It’s definitely a great way to visualize the graphs if you can’t see them in the charts. I mean, it’s just so cool that this is something that allows a blind person to access a graph and a chart and be able to parse data from it.` (S3)

Participants highlighted that VOXLENS contributes to bridging the access gap between screen-reader- and non-screen-reader users. As S4 said:

> `So, as a sighted person looks at a graph and as they can tell where the peak is or which one has the most or whatever, we want to be able to do that quickly as well. And even if there is a text description under the graph, and I’ve not seen that very much, you have to read through everything to find a certain piece of information that you’re looking for. [Using VoxLens], I can find out specific pieces of information without having to read an entire page of text.` (S4)

Additionally, participants identified that VOXLENS enables them to quickly extract information from online data visualizations. S5 shared his experiences:

> `Again, you know, [VoxLens] helps you find data a little bit quicker than navigating with a screen reader, and it’ll give you a brief idea of what the data is about before you start digging deeper into it.` (S5)

The findings from the first result show that VOXLENS contributes to reducing the access gap for screen-reader users, and is a positive step forward, enabling screen-reader users to interact with and explore online data visualizations.

5.3.2.2 Interactive Dialogue is One of the "Top" Features

Similar to the first finding, all the participants found the question-and-answer mode of VOXLENS a fast and efficient way to extract information from online data visualizations. S2 considered the question-and-answer mode as one of the key features of VOXLENS:

> `So I believe that one of the really top features is interactive dialogue.` (S2)

Similarly, S1 found the question-and-answer mode a fast and reliable way to extract information, requiring "a lot less brain power." She said:

> `I especially liked the part of the tool where you can ask it a question and it would give you the information back. I thought it was brilliant actually. I felt like being able to ask it a question made everything go a lot faster and it took a lot less brain power I think. I felt really confident about the answers that it was giving back to me.` (S1)

S3 noted the broader utility and applicability of the question-and-answer mode:

> `The voice activation was very, very neat. I'm sure it could come in handy for a variety of uses too. I definitely enjoyed that feature.` (S3)

S5 faced some challenges in activating the right command but was able to learn the usage of the question-and-answer mode in a few tries:

> `You know, sometimes the word was wrong and I think it says something like, it didn't understand, but basically eventually I got it right.` (S5)

The second finding indicates that VOXLENS' question-and-answer mode is a fast, efficient, and reliable way for screen-reader users to extract information. Additionally, the feedback from the question-and-answer mode assists screen-reader users to resolve the challenges by themselves within a few tries.

5.3.2.3 Sonification Helps in “Visualizing” Data

The third result reveals that the participants found sonification helpful in understanding general trends in the data. Specifically, participants were successful in inferring whether an overall trend was increasing or decreasing, obtaining holistic information about the data. S2 said:

> `The idea of sonification of the graph could give a general understanding of the trends. The way that it could summarize the charts was really nice too. The sonification feature was amazing.` (S2)

S1, who had never used sonification before, expressed her initial struggles interpreting a sonified response but was able to “visualize” the graph through sonification within a few tries. She said:

> `The audio graph... I’d never used one before, so I kind of struggled with that a little bit because I wasn’t sure if the higher pitch meant the bar was higher up in the graph or not. But being able to visualize the graph with this because of the sound was really helpful.` (S1)

Overall, the third result shows that sonification is a helpful feature for screen-reader users to interact with data visualizations, providing them with holistic information about data trends.

5.3.2.4 Data Summary is a Good Starting Point

In keeping with findings from prior work (Sharif et al. [2021]), the fourth finding indicates that screen-reader users first seek to obtain a holistic overview of the data, finding a data summary to be a good starting point for visualization exploration. The summary mode of VoxLens enabled the participants to get a “general picture” of the data quickly. S1 and S4 expressed the benefits of VoxLens’ summary mode:

> `I thought the summary feature was really great just to get, like, a general picture and then diving deeper with the other`

> `features to get a more detailed image in my head about what the graphs look like.` (S1)

> `So, um, the summary option was a good start point to know, okay, what is, kind of, on the graph.` (S4)

The fourth result indicates that VoxLens' summary mode assisted screen-reader users in holistically exploring the data visualizations, helping them in determining if they want to delve deeper into the data.

5.3.2.5 One-Size-Fits-All Is Not the Optimal Solution

To enhance the usability of and interaction experience with VoxLens, the participants identified the need to cater to the individual preferences of the screen-reader users. For example, S3 recognized the need to have multiple options to "play" with the sonified response:

> `So I was just thinking maybe, you know, that could be some sort of option or like an alternate way to sonify it. Perhaps having an option to do it as continuous cause I noticed, like, they were all discrete. 'Cause sometimes, you know, it's just preference or that could be something that could add some usability. It's just some little things to maybe play with or to maybe give an option or something.` (S3)

Similarly, S4 was interested in VoxLens predicting her future queries using artificial intelligence (A.I.). She said:

> `You know, I think that [VoxLens] would need a lot more artificial intelligence. It could be a lot [more] intuitive when it comes to understanding what I'm going to ask.` (S4)

Additionally, S2 suggested adding setting preferences for the summary and the auditory sonified output:

> `[Summary mode] could eventually become a setting preference or something that can be disabled. And you, as a screen-reader`

> `user, could not control the speed of the [sonification] to you. To go faster or to go slower, even as a blind person, would be [helpful].` (S2)

These findings indicate that a one-size-fits-all solution is not optimal and instead, a customizable solution should be provided, a notion supported by ability-based design (Wobbrock et al. [2018]). I present a solution that addresses this feedback and suggestions about customization from the participants in Chapter 9.

5.3.3 Subjective Workload Ratings

I used the NASA Task Load Index (TLX) (Hart and Staveland [1988]) workload questionnaire to collect subjective ratings for VOXLENS. The NASA-TLX instrument asks participants to rate the workload of a task on six scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each scale ranges from low (1) to high (20). I further classified the scale into four categories for a score x: *low* ($x < 6$), *somewhat low* ($6 \leq x < 11$), *somewhat high* ($11 \leq x < 16$), and *high* ($16 \leq x$). The results indicate that VOXLENS requires low physical- (M=3.4, SD=3.3) and temporal demand (M=5.7, SD=3.8), and has high perceived performance (M=5.6, SD=5.6). Mental demand (M=7.8, SD=4.4), effort (M=9.9, SD=6.1), and frustration (M=8.3, SD=6.6) were somewhat low.

My prior work (Sharif et al. [2021]), which is the source of the data for screen-reader users who did not use VOXLENS, did not administer a NASA-TLX survey with their participants. Therefore, a direct workload comparison is not possible. However, the subjective ratings from this study could serve as a control for comparisons in future work attempting to make online visualizations accessible for screen-reader users.

5.4 Summary

In this chapter, I presented an evaluation of the first version of VOXLENS through task-based experiments and interviews with screen-reader users. VOXLENS significantly improved the interaction experiences of screen-reader users with online data visualizations, both in terms of accuracy of extracted information and interaction time, compared to their conventional interaction with online data visualizations. The results show that screen-reader users considered VOXLENS a "game-changer," providing them with "exciting new ways” to interact with online data visualizations and saving time and effort.

Among other findings, the follow-up interviews show that the participants found sonification helpful in “visualizing” data. In the next chapter, I shed light on when sonification is helpful for screen-reader users and when it is not. Additionally, I discuss the factors that improve the usability and user-friendliness of sonification for screen-reader users.

Chapter 6

Improving Data Sonification for Screen-Reader Users*

Chapter 2 discussed the common modalities used to make online data visualizations accessible to screen-reader users, and Chapter 4 illustrated the three interaction modes of `VoxLens`. In this chapter, I focus on data sonification, one of the widely used modalities to explore holistic information from online data visualizations. In particular, I assess the experiences of screen-reader users with data sonification to illuminate the use cases in which sonification offers the most value in interpreting data from online visualizations. Additionally, I investigate the usability and user-friendliness of data sonification for screen-reader users. I present a decision tree for visualization creators to assist them in appropriately using sonification to make data visualizations accessible. Furthermore, I provide the implementation details of my enhancements to `Sonifier`—an open-source JavaScript library that enables developers to sonify online data visualizations.

6.1 Motivation

With the increasing data representation through visualizations comes the critical need to make these visualizations accessible to people who may not be able to extract information using visual means (e.g., screen-reader users) (Marriott et al. [2021]; Sharif et al. [2021]; Lee et al. [2021]; Davis [2002]). Researchers and developers have implemented several approaches and strategies to make online data visualizations accessi-

---

*Parts of this chapter are adapted from Sharif et al. [2022c] and (Sharif et al. [2024a]).

ble (Sharif et al. [2022f]; Giudice et al. [2012]; Sharif and Forouraghi [2018]; Geddes et al. [2022]; Kim and McCoy [2018]; Shi et al. [2016]). Among these techniques is sonification, often referred to as "audio graphs," a widely-used approach in conveying data through auditory channels to screen-reader users.

Sonifications[1] are the non-speech audio representations of data visualizations and used to make data visualizations accessible to screen-reader users (for the Blind [2023]; Walker and Nees [2011]; Barrass and Kramer [1999]). In the past five years (2018 – 2023), at least 518 and 347 papers[2] on ACM Digital Library and IEEE Xplore, respectively, utilized, created, or improved data sonification tools across varying domains. Additionally, commercial products are increasingly and ubiquitously using sonification as an accessibility technique (Software [2023]; Wall et al. [2012]; Studios [2011]). For example, Apple added sonification as a standard feature across all their devices in their 2021 operating system releases (Apple [2021b]).

Several prior works have utilized sonification to improve the accessibility of online data visualizations (Sharif et al. [2022f]; McGookin and Brewster [2006]; Zhao et al. [2008]; Highcharts [2009]; Apple [2021a]; Ahmetovic et al. [2019b]; Siu et al. [2022]; Fan et al. [2022]; Holloway et al. [2022]). Despite it's first usage to represent graphs dating back to 1974 (Chambers et al. [1974]) and recent wide and plausible adoption in research (Sharif et al. [2022f]; Austin and Sorge [2023]; Fan et al. [2022]; Holloway et al. [2022]; Hoque et al. [2023]), the use cases where sonification is—and is not—beneficial for screen-reader users remain unclear. Additionally, these solutions focus on the utility (usefulness) of sonification to screen-reader users and provide limited insights into the quality (usability and user-friendliness).

For shedding light on the usefulness and improving the usability of sonification for screen-reader users, my approach is two-fold. First, I assess the experiences of screen-reader users with sonification when interpreting data from visualizations and explore the factors that make sonification beneficial for them. My empirical results show that the participants rarely encounter sonification online and that their experiences with sonification are often limited to only research studies. I utilized my findings to generate a decision tree for visualization creators to aid them in using sonification appropriately for making online data visualizations accessible to screen-reader users.

Then, I examined and improved the usability and user-friendliness of sonified responses from online data visualizations created using JavaScript libraries (e.g., D3 (Bostock et al. [2011])). The most relevant

---

[1] I use "sonification" synonymously with "audio graphs," limiting its domain of usage to data visualizations.

[2] I combined the search results for "sonification" and "audio graph," accounting for plural terms, too.

research to this investigation is the exploration by Wang *et al.* (Wang et al. [2022a]), in which they examined the impact of various auditory channels (e.g., pitch, volume) on users' perception of data and visualization. I build on their work by investigating the effects of different oscillator waveforms and synthesizers on the pleasantness and users' confidence in interpreting simple and complex sonified responses. Further, using my findings, I extended the functionalities of `Sonifier`—an open-source JavaScript library that generates sonified responses for two-dimensional single-series data (Sharif et al. [2022f,e]).

6.2 Formative Study

As an initial step toward investigating the usefulness of sonification, I surveyed 106 screen-reader users. I examined their responses with multiple methods, handling Likert responses with quantitative methods (McCullagh [1980]; McKelvey and Zavoina [1975]) and open-ended responses with qualitative (Braun and Clarke [2006]; Boyatzis [1998]; Patton [1990]).

6.2.1 Participants

The survey respondents ("participants") were 106 screen-reader users (M=49.8 years, SD=16.1). I advertised the survey through the National Federation of the Blind (National Federation of the Blind [1940]). Sixty-one identified as women, 38 as men, five as non-binary, and two did not disclose their gender identity. Sixty participants used JAWS screen reader (Scientific [1995]), 25 used VoiceOver (Inc. [2005]), 17 used NVDA (Access [2006]), and the remaining four used a combination of multiple screen readers. Fifty-nine participants were blind since birth and 47 had lost vision gradually; 84 participants had complete blindness. The highest level of education was a doctoral degree for 10 participants, a master's degree for 34, a bachelor's degree for 34, an associate's degree for eight, a high school diploma for 18, and pre-high-school for the remaining two participants.

6.2.2 The Survey

Participants filled out the three-step survey online without supervision that I created using Google Forms[3]. The first step included the purpose of the study, eligibility criteria, data anonymity clause, the definition of

---

[3] https://forms.google.com

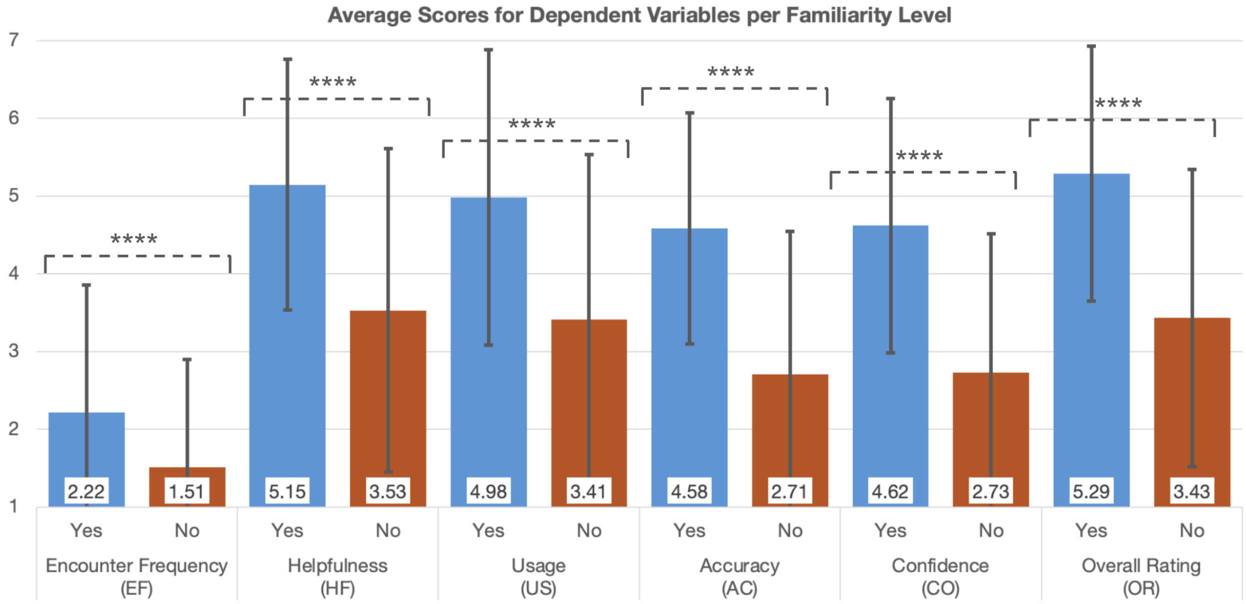


Figure 6.1: Average ratings for EF, HF, US, AC, CO, and OR per each familiarity level (PF). I recognize these are ordinal values; means and SDs are shown for illustrative purposes only, and inferential statistics were performed appropriately using ordinal logistic regression, non-parametric ANOVA. Error bars represent mean ± standard deviation. All results are statistically significant ($p < .05$).

sonification, and an example of sonified output. In step two, I collected demographic information from the participants. I recorded their gender identity, pronouns, age, screen reader usage, vision level, diagnosis, age of diagnosis, and education status. I followed guidelines from Spiel *et al.* (Spiel et al. [2019]) to inquire about their gender identities appropriately. In the last step, I asked about their familiarity with sonification as a binary response ("yes" or "no"). Then, I used a Likert scale ranging from 1 (lowest; e.g., "not at all") to 7 (highest; e.g., "extremely helpful") to get the answers to the following questions on sonification:

1. How often do you encounter sonification? (EF)
2. How helpful do/would you find sonification to extract information from data visualizations? (HF)
3. How often do/would you use sonification in practice? (US)
4. How do/would you perceive the accuracy of information extracted using sonification? (AC)
5. How confident do/would you feel in their overall understanding of the data from sonification? (CO)
6. What is your overall rating for sonification's usefulness? (OR)

Finally, I asked the participants open-ended questions about situations in which they do/would find

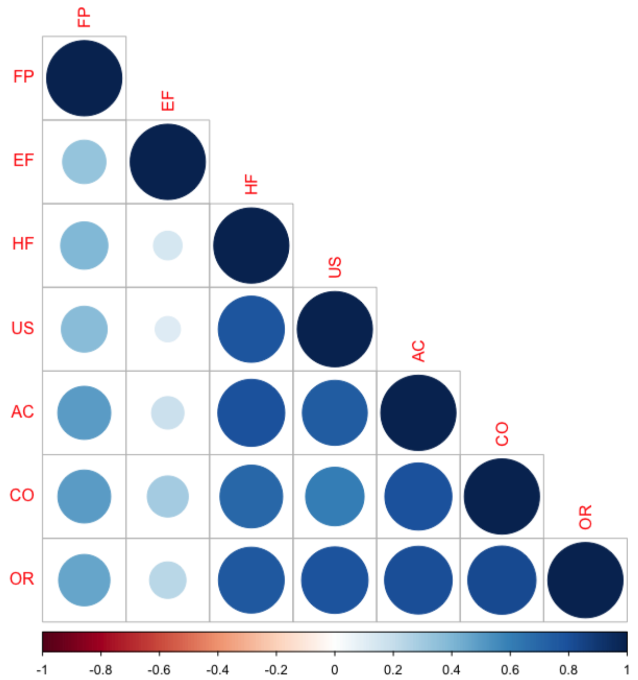


Figure 6.2: Correlogram showing correlations between each pair of EF, HF, US, AC, CO, and OR.

sonification helpful and in which they do/would not. I also asked about their overall opinion on sonification and their interest in participating in a follow-up interview.

6.2.3 Quantitative Evaluation

I investigated the effect of their prior familiarity with sonification on their Likert responses. My goal was to distinguish between ratings from participants who had experienced sonification before and those from participants who had either never experienced it or only encountered it via examples. I note that responses from both these groups are vital to provide insights into the usefulness of sonification. Therefore, prior familiarity was of particular interest to me in this exploration. The independent variable was *Prior Familiarity* (PF), having dichotomous levels ("yes" and "no"). I used the questions identified in the subsection above as the dependent variables (EF, HF, US, AC, CO, and OR), each having an ordinal representation (1-7). To analyze the effect of each of these variables on PF, I used ordinal logistic regression (McCullagh [1980]; McKelvey and Zavoina [1975]), a standard technique used for analyzing ordinal response data.

Table 6.1: Statistical results from the Spearman's rank correlation analysis of all the variables. "V1" means Variable 1 and "V2" means Variable 2. ρ is the Spearman's rank correlation coefficient. All results with p <.05 are statistically significant.

V1	V2	ρ	p
CO	OR	.82	< .001
AC	OR	.81	< .001
AC	CO	.80	< .001
HF	AC	.79	< .001
US	OR	.78	< .001
HF	US	.77	< .001
HF	OR	.77	< .001
US	AC	.75	< .001
HF	CO	.70	< .001
US	CO	.61	< .001
PF	AC	.50	< .001
PF	CO	.49	< .001
PF	OR	.47	< .001
PF	HF	.40	< .001
PF	US	.37	< .001
PF	EF	.33	< .001
EF	CO	.30	< .05
EF	OR	.24	< .05
EF	AC	.19	.057
EF	HF	.15	.139
EF	US	.11	.249

Prior familiarity (PF) had a significant effect on all dependent variables, *i.e.*, the Likert responses (1-7) named above, indicating that each response differs significantly between the two familiarity groups (see Figure 6.1 for average scores for each dependent variable per each familiarity level).

Overall, 74.6% (N=41) of the participants with prior familiarity rated sonification at least "somewhat useful" (5 to 7 on the Likert scale). For participants unfamiliar with sonification, this percentage was only 27.4% (N=14), constituting 47 percentage points (%pt) for OR. Using the same scale for comparison, pp was identically high for HF (32 %pt), AC (33 %pt), CO (41 %pt), and US (30 %pt). For EF, it was 5 %pt (12.7% *vs.* 7.8%), indicating an agreement between the two groups on the low presence of sonification in online data visualizations.

In addition to examining the effect of prior familiarity, as a secondary exploration, I performed Spearman's rank correlation analysis (Spearman [1904]) between all the variables to gather further insights. I found a statistically significant and positive correlation between all variables *except* for EF and HF (p=.139), EF and US (p=.249), and EF and AC (p=.057) (see Table 6.1 and Figure 6.2). I note that correlation does not imply causality but helps determine the linear relationship between the variables. The fact that many of these responses are correlated suggests an interdependent relationship among these variables that future work can explore further. For example, taken together, the results show that with an increase in encountering frequency (EF), participants' confidence (CO) in understanding the data also increased.

6.2.4 Qualitative Evaluation

I qualitatively analyzed the open-ended responses from 106 screen-reader users. Specifically, I conducted a theoretical thematic analysis (Boyatzis [1998]) using a semantic approach (Patton [1990]) and an essentialist paradigm (Potter and Wetherell [1987]; Widdicombe and Wooffitt [1995]), following guidelines from Braun and Clarke (Braun and Clarke [2006]). Two co-authors and I coded each response, reaching a high agreement percentage of 95%. The final analysis converged to a single prominent theme of *"useful but only in the right context."* Overall, the participants accentuated its limited usefulness. For example, P51 and P70, both of whom had experienced sonification before, found it helpful for a high-level overview and unhelpful for identifying granular details, respectively:

> `Better to keep sonification for high-level overview and not to`

> `over-complicate them. The audio equivalent of a first glance. In general, I think this is their best use.` (P51)

> `Where I would need a very specific information, sonification would not be helpful. I see it more for a general overview and less for specific data.` (P70)

On the other hand, P25 and P26, who had only encountered sonification in examples, classified it as "a tool with great potential" and a solution that "needs refinement," respectively.

The findings from the need-finding survey motivated the need to delve further into the use cases in which screen-reader users do and might find sonification useful. I present the details of the in-depth semi-structured interviews in this chapter under its respective section.

6.3 Study Design

In this section, I present the in-depth interviews I conducted to assess the usefulness of sonification for screen-reader users. I also discuss the user study I administered to investigate the usability and user-friendliness of sonified responses.

6.3.1 In-Depth Interviews

To gather further insight into the experiences of screen-reader users with data sonification, I conducted in-depth semi-structured interviews with 12 screen-reader users. I present the methodology of the interviews and discuss the details of the qualitative analysis.

6.3.1.1 Participants

Using the list of respondents interested in participating in a follow-up interview from the survey, I created two participant pools based on their prior familiarity (PF) with sonification. I excluded participants who did not provide meaningful responses to the open-ended questions in the survey (e.g., "none" or "n/a"). I randomly chose six participants from each PF pool, totaling 12 screen-reader users (see Table 6.2; S1–S6 are participants with prior familiarity with sonification and S7–S12 are participants without). Among

Table 6.2: Gender, age, prior familiarity ("PF"), vision level, and diagnosis of the interview participants. Under the "G" (*Gender*) column, M = *Man*, W = *Woman*, NB = *Non-binary*, and "-" means preferred not to disclose.

	G	**Age**	**PF**	**Vision-Loss Level**	**Diagnosis**
S1	M	55	Yes	Complete blindness	Leber Congenital Amaurosis
S2	W	50	Yes	Complete blindness	Retinitis Pigmentosa
S3	W	24	Yes	Complete blindness	Astrocytoma
S4	NB	28	Yes	Complete blindness	Leber Congenital Amaurosis
S5	W	42	Yes	Complete blindness	Glaucoma
S6	-	57	Yes	Complete blindness	Retinopathy of Prematurity
S7	W	41	No	Complete blindness	Functional Neurological Disorder
S8	M	76	No	Complete blindness	Retinitis Pigmentosa
S9	W	36	No	Complete blindness	Retinopathy of Prematurity
S10	W	65	No	Partial blindness	Retinitis Pigmentosa
S11	W	72	No	Partial blindness	Retinitis Pigmentosa
S12	M	32	No	Complete blindness	Retinal Detachment

participants with prior familiarity, four identified as women and two as men. Their average age was 53.7 years (SD=19.5). Four had complete blindness, whereas two were partially blind. Four had attained a master's degree, and the remaining two participants had bachelor's and associate's degrees, respectively.

For participants unfamiliar with sonification, three identified as women, one as a man, one as gender-fluid, and one preferred not to disclose. All participants were blind, and their average age was 42.7 years (SD=14.0). The highest level of education was a doctoral degree for two participants, an associate's degree for three participants, and a high school diploma for the remaining one participant.

I ceased recruitment of participants once I reached saturation of insights and compensated them with a $25 Amazon gift card for 45 minutes of their time.

6.3.1.2 Procedure

I conducted the semi-structured interviews via Zoom using its built-in features to record and transcribe the 45-minute-long sessions with two other co-authors. I took detailed notes during the session. During the interviews, I inquired about the usefulness of data sonification and future efforts to improve the information extraction experiences of screen-reader users with sonification. Specifically, I explored the nuances and intricacies of when sonification is, or might be, helpful in extracting information from online data visualizations for screen-reader users.

6.3.1.3 Analysis

I used inductive thematic analysis(Braun and Clarke [2006]), following a semantic approach (Patton [1990]), guided by an essentialist paradigm (Widdicombe and Wooffitt [1995]; Potter and Wetherell [1987]). Using the first two interviews, I developed an initial set of codes (Ryan et al. [2000]; Braun and Clarke [2006]). Two co-authors and I coded each interview transcript independently, resolving disagreements through mutual discussions. Following Braun and Clarke's guidelines on thematic analysis (Braun and Clarke [2006]), I combined the 41 open codes into 11 axial codes. The final analysis revealed three prominent themes, which I discuss in the results section below. Following the suggestion by Landis *et al.* (Landis and Kock [1977]), I calculated inter-rater reliability (IRR) using pairwise percentage agreement as well as Krippendorff's α (Krippendorff [2011]). The pairwise percentage agreement was 82%, showing a high agreement between raters (Hartmann [1977]; Graham et al. [2012]). Krippendorff's α was 0.81, indicating a high level of reliability (Krippendorff [2018]), computed using ReCal 3.0 (Freelon [2010]).

6.3.2 User Study

To assess the usability and user-friendliness of sonified responses, I conducted a user study with 10 screen-reader users, subsequently interviewing them to gather insights on further improving the sonified responses.

6.3.2.1 Prototypes

I created several sonification prototypes employing different combinations for oscillator waveforms, synthesizers, and partial counts (number of harmonics to generate the waveform). I developed these prototypes

Table 6.3: Screen-reader participants in the user study, their gender identification, age, screen reader, vision-loss level, and diagnosis. Under the *Gender* column, M = *Male*, F = *Female*, and NB = *Non-binary*.

	G	**Age**	**Screen Reader**	**Vision-Loss Level**	**Diagnosis**
P1	M	57	JAWS	Lost vision gradually	Retinitis Pigmentosa
P2	F	38	JAWS	Blind since birth	Leber Congenital Amaurosis
P3	F	65	JAWS	Lost vision gradually	Retinitis Pigmentosa
P4	F	69	Fusion	Lost vision gradually, Partial vision	Juvenile Macular Degeneration
P5	M	33	NVDA	Blind since birth	Peters Anomaly
P6	M	37	JAWS	Blind since birth	Leber Congenital Amaurosis
P7	F	52	JAWS	Blind since birth	Retinopathy
P8	M	58	JAWS	Lost vision gradually	Cataracts and Glaucoma
P9	M	49	JAWS	Lost vision gradually	Leber Congenital Amaurosis
P10	NB	26	VoiceOver	Partial vision	Corneal damage

using the `Tone.js` JavaScript library (Mann [2020])—a widely-used framework to generate sounds in the browser. Then, I eliminated the prototypes that users found undesirable through Wizard-of-Oz (Dahlbäck et al. [1993]; Hajdinjak and Mihelic [2004]) and pilot studies. After elimination, the sound types comprised an oscillator (termed "OmniOscillator" by `Tone.js`) and a synthesizer (termed "MonoSynthesizer" by `Tone.js`), with both using the *square* waveform and a partial count of *8*. I created a continuous (minimal time interval between sounds for each data point) and discrete (clear time interval between sounds for each data point) prototype for each sound type. I used the default settings from the `Sonifier` library as the baseline measure. Since the baseline only supports discrete responses, I developed the fifth prototype that generated a continuous response for the baseline to account for balanced conditions. The final set contained 12 prototypes (including the baseline), six for each *Trend Type* (simple and complex).

6.3.2.2 Participants

The participants were 10 screen-reader users (see Table 6.3). Four identified as women, five as men, and one as non-binary. Their average age was 48.4 (SD=14.4) years. I compensated the participants with a \$10 Amazon gift card for 30 minutes of their time. I supervised the user studies online using Zoom.

6.3.2.3 Procedure

For each participant, I played six sonified responses (five prototypes + baseline) for simple trends, randomizing the order to account for the learning effect. At the end of each sonification, I collected their subjective ratings for *Pleasantness* (timbre), *Clarity* (assessment of the sound to identify the trend clearly), *Confidence* (user's confidence in understanding the overall trend), and *Overall Score* (subjective assessment of the sound overall, including any factors not mentioned above). I used a Likert scale for subjective ratings, ranging from 1 (worst) to 7 (best). Then, I asked follow-up questions from the users to gather insights on the areas of improvement. I followed the same steps for complex trends. The study sessions, on average, took approximately 30 minutes from start to finish.

6.3.2.4 Design & Analysis

The experiment was a $3 \times 2 \times 2$ within-subjects design with the following factors and levels:

- *Sound Type (S)*: {Baseline, OmniOscillator, MonoSynthesizer}
- *Continuity Level (C)*: {discrete, continuous}
- *Trend Type (T)*: {simple, complex}

I used *Pleasantness* (PL), *Clarity* (CL), *Confidence* (CF), and *Overall Score* (OS) as the dependent variables. In the analysis, all the dependent variables were ordinal (on a scale of 1-7). Therefore, I conducted the analysis using an ordinal logistic regression (Gutiérrez et al. [2015]; Winship and Mare [1984]). I also included $Subject_r$ as a random factor to account for repeated measures. I tested the participants over $3 \times 2 \times 2$=12 conditions, resulting in 12×10=120 total trials.

6.4 Results

In this section, I present the results from the in-depth interviews and the user study with screen-reader users.

6.4.1 In-Depth Interviews

I present the findings from the qualitative analysis of the semi-structured interviews with 12 screen-reader users. I identified three themes: (1) "Needle in a Haystack," (2) "Keep it Simple, Sonifier," and (3) "Suggestions to Improve Sonification's Effectiveness." I discuss each of these themes in turn. Additionally, I append "+" and "-" to the participants' identifiers for ease in recognizing which participants had prior familiarity with sonification and which did not, respectively. These results form the basis for the decision tree I present in the next section.

6.4.1.1 Needle in a Haystack

In the survey, 63% of participants answered "never" when asked about how often they encounter sonification in online data visualizations. This finding was also a prevalent theme in the interviews. The findings show that the participants had seldom encountered sonification before, mostly only in research studies. S3+ and S4+ expressed that they have not experienced sonification often in online data visualizations, attributing to it being a "newer technology" (S3+) and not "mainstream" (S4+):

> `It seems like it's a newer technology, so it's not in practice quite yet. Because I think the first time I heard of sonification was through a podcast.` (S3+)

> `I don't really think sonification is at all mainstream. Like, you know, data visualizations are super common, super used, but sonification, I find not very often.` (S4+)

On the other hand, S5+ and S2+ shared that their familiarity with sonification was only through participating in research studies:

> `Actually, a few years ago when the pandemic was starting, I did participate in a focus group where sonification was used in a chart for COVID-19, and I found that really helpful and interesting.  So I guess the last time I interacted with was maybe three years ago.` (S5+)

> `I encounter it maybe once every few months in a couple of studies that were investigating sonification for different purposes.` (S2+)

6.4.1.2 Keep it Simple, Sonifier

The title of the second theme is inspired by the KISS principle ("keep it simple, stupid") (Alwin and Beattie [2016]). In interviews with participants, I found that sonification helps interpret the data in online visualizations when used to provide an *overview* of the data, particularly data trends. For example, S4+ and S12- expressed the usefulness of sonification to obtain an overview of the data and the data trend, respectively:

> `I think it has to stick to, like, if you're just trying to give somebody an overview.  And that's it, that's kind of all you can do with sound.  If you try to really explain lots and lots and lots and then it just becomes too hard and too complicated to understand.` (S4+)

> `It would be generally useful if you could hear a trend. Like if, for example, work gave me income breakdowns for every single week, and instead of me having to look through all of it, if I could sonify that, and it played a sound of where the income's going up or when it's going down, that would be helpful for me.` (S12-)

On the other hand, the participants expressed concerns regarding using sonification for complex[4] data

---

[4] I use the term "complex" based on the factors mentioned in prior work (Kosslyn [1989]; Pinker [1990]; Lewandowsky and Spence [1989]; Trafton et al. [2000]), including data dimensionality, series, cardinality, and users' cognitive load.

and visualizations. For example, S7- and S4+ did not find sonification helpful for understanding the data at a granular level, especially when extracting specific data points:

> `Basically, sometimes it's the visualization that's difficult to understand, and sometimes it's the data. So a lot of things come into play. Sometimes even a simple scatter plot might have data that is really complex to interpret a simple summary out of it. So, I think if anything is going to be super complex, that would make it very challenging to understand sonification.` (S7-)

> `But say you wanna know more than just the general trend, you wanna actually know where all the data points are. But just something that gets harder to just represent with, like, one beam of sound. It's more so how much you personally wanna delve into that data that dictates whether sonification is going to be useful or not.` (S4+)

Interestingly, while the participants considered sonification beneficial for "simple" data (easily interpretable via sound), they did not find it helpful for data that was "too simple" (*e.g.*, of low cardinality). For example, S5+ preferred data to be relayed via text or tables when data points were very few.

> `Probably a simple chart that only had a few points, then that's not very helpful. Because if the alternative text was already there or the table and there wasn't too much to the chart, then there's really no need to have the sonification. I think that information could be relayed through other means.` (S5+)

6.4.1.3 Suggestions to Improve Sonification's Effectiveness

Despite the low encounter frequency of sonification solutions online, the participants overall felt positive about it as a technique to make online data visualizations accessible for screen-reader users. Additionally, they identified avenues to enhance the utility of sonification. For example, several participants indicated that sonification would have increased usefulness when combined with other modalities, such as alternative text or data tables. S9- emphasized the need for other modalities, and S3+ suggested providing a summary of the data and an explanation of the response through other modalities (such as alt-text):

> `So, just as a blind person, you would want to have multiple modalities so you can use sonification when you think it's appropriate and then turn it off and just use something else like alt-text.` (S9-)

> `Basically, you're given a little bit of a summary about the graph and then explained what the sonification would mean, such as the higher pitches mean the higher data, or lower pitches mean the lower values. It gives users a better understanding of the graphs that they're looking at.` (S3+)

Additionally, similar to prior work (Sharif et al. [2022c]), the participants identified the need to make sonification customizable to cater to screen-reader users' individual preferences and discussed the benefits of customization. For example, S9- considered customization important for her interaction needs:

> `Customization's so important to me. I don't know if a sighted interface is like that, but I'd like sonification to be like that, where you can customize it however you want. Like, sounds, speed, all that.` (S9-)

Specifically, the participants identified the need to have the ability to turn the sound on or off and pause it when needed. In addition, they deemed it essential to select the pitch, frequency, and speed of the sound. S11-, S1+, and S7- all discussed the necessity of these features:

> `So, if there were some kind of well-labeled and easy-to-find buttons on the screen that the blind person could either hit or touch on the phone screen, they could pause the sonification`

> `and then hit the button to give, maybe the specific data point number or more information or whatever, you know.` (S11-)

> `We would like to set the pitch to whatever we want. Like a variable frequency oscillator, right? I tune the dials a bit different than my sighted colleagues. But every man is different. Every guy has a different ear tolerance.` (S1+)

> `I think it would be really cool because like even with VoiceOver and JAWS, we have ways that we can do different voices and we can make it all at different speeds and stuff, and we can speed it up or slow it down or change it in between. I think that would be really neat to be able to make sonification customizable.` (S7-)

Finally, the participants accentuated the importance of learnability, highlighting that learning sonification can sometimes have a big learning curve and may need practice. S8- shared his thoughts:

> `This is the kind of thing that takes practice. See, two things have to change here. The computer has to change, but the human has to change also. So I have to learn a different way of understanding data. I could do it well with practice.` (S8-)

Overall, the findings show that sonification has a limited presence in online data visualizations, would be more helpful in providing overviews of data than granular explorations, and could be improved with customization features and resources to help learnability.

6.4.2 User Study

I present the results of the user studies assessing *Pleasantness* (PL), *Clarity* (CL), *Confidence* (CF), and *Overall Score* (OS) of sonified responses. Additionally, I offer the findings on areas of improvement for sonified responses from the follow-up interviews.

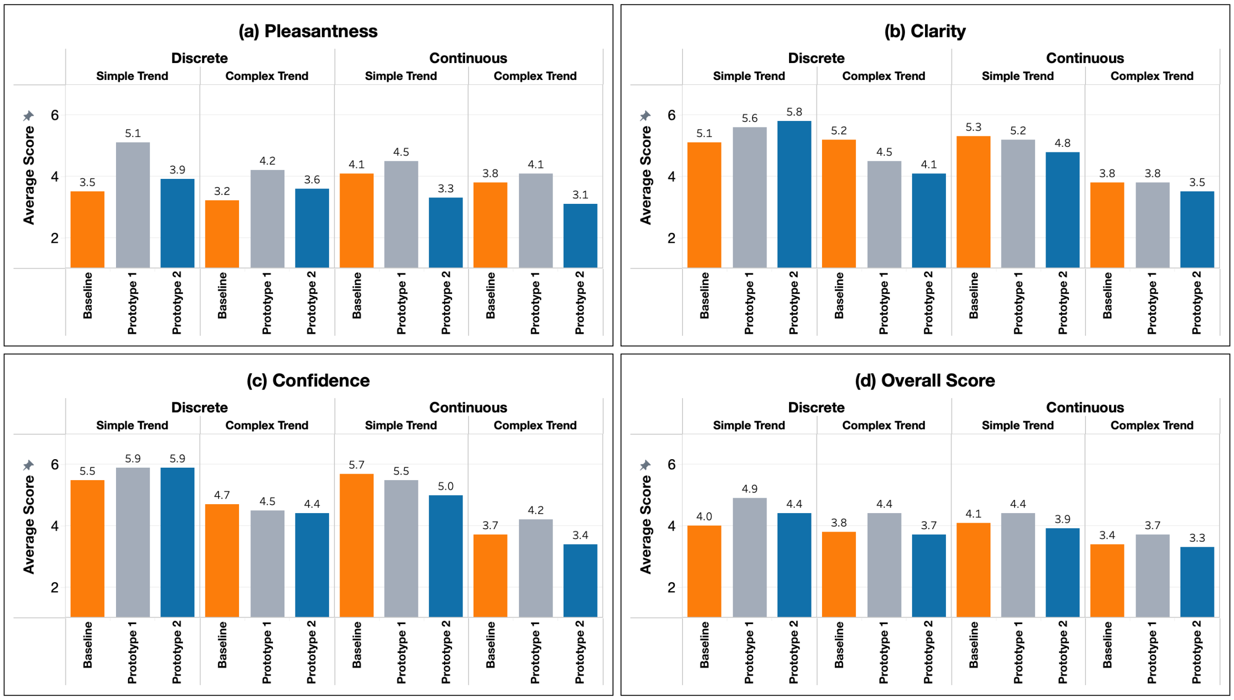


Figure 6.3: Average scores per Continuity Level and Trend Type for each Sound Type for (a) *Pleasantness*, (b) *Clarity*, (c) *Confidence*, and (d) *Overall Scores*. *Baseline* was a *sawtooth* waveform OmniOscillator, *Prototype 1* was a *square* waveform OmniOscillator, and *Prototype 2* was a *square* waveform MonoSynthesizer. Subjective ratings were collected using a Likert scale ranging from 1 (worst) to 7 (best).

6.4.2.1 Pleasantness (PL)

The results show a significant main effect of S (χ^2(2, N=10)=12.15, p<.05, Cramer's V=.42) on PL overall, indicating that PL was significantly different between the three S groups. Overall, the sonified responses from *OmniOscillator* with *square* waveform and discrete continuity level outperformed the other prototypes (M=4.65). C ($p \approx$.844) and T ($p \approx$.122) did not have a significant main effect on PL. Figure 6.3 and Table 6.4 show average PL scores for each independent variable.

6.4.2.2 Clarity (CL)

I found a significant main effect of C (χ^2(1, N=10)=11.03, p<.001, Cramer's V = .40) and T (χ^2(1, N=10)=22.83, p<.001, Cramer's V=.57) on CL overall. This result indicates that CL was significantly different between discrete and continuous sounds and also between simple and complex trends. On average,

Table 6.4: Overall average scores for each sound type per continuity level. PL represents *Pleasantness*, CL represents *Clarity*, CF represents *Confidence*, and OS represents *Overall Score*. Highest average scores for PL, CL, CF, and OS are shown in bold.

		Average Scores			
Sound Type (S)	**Continuous**	PL	CL	CF	OS
OmniOscillator with sawtooth waveform (Baseline)	No	3.35	**5.15**	5.10	3.90
	Yes	3.95	4.55	4.70	3.75
OmniOscillator with square waveform	No	**4.65**	5.05	**5.20**	**4.65**
	Yes	4.30	4.50	4.85	4.05
MonoSynthesizer with square waveform	No	3.75	4.95	5.15	4.20
	Yes	3.20	4.15	4.58	3.80

the *Baseline* had the best scores (M=5.15). S (p ≈.476) did not have a significant main effect on CL. I show the average scores for CL in Figure 6.3 and Table 6.4 for each independent variable.

6.4.2.3 Confidence (CF)

C (χ^2(1, N=10)=6.36, p<.05, Cramer's V=.30) and T (χ^2(1, N=10)=42.34, p<.001, Cramer's V=.78) had a significant main effect on CF, indicating that CF significantly differed between simple and complex trends as well as discrete and continuous sounds. Similar to PL, the sonified responses from *OmniOscillator* with *square* waveform and discrete continuity level had the best overall average scores (M=5.20). I did not find a significant effect of S on CF (p ≈.726). The average scores for CF are shown in Figure 6.3 and Table 6.4 for each independent variable.

6.4.2.4 Overall Score (OS)

The results show a significant main effect of T on OS overall (χ^2(1, N=10)=7.16, p<.05, Cramer's V=.32). This finding indicates that OS significantly varied between the simple and complex trends. *OmniOscillator* with *square* waveform and discrete continuity level performed the best compared to the other prototypes (M=4.65) on average, similar to PL and CF. I only found a marginal effect of S (p ≈.090) and C (p ≈.098) on OS. I display the average scores for OS in Figure 6.3 and Table 6.4 for each variable.

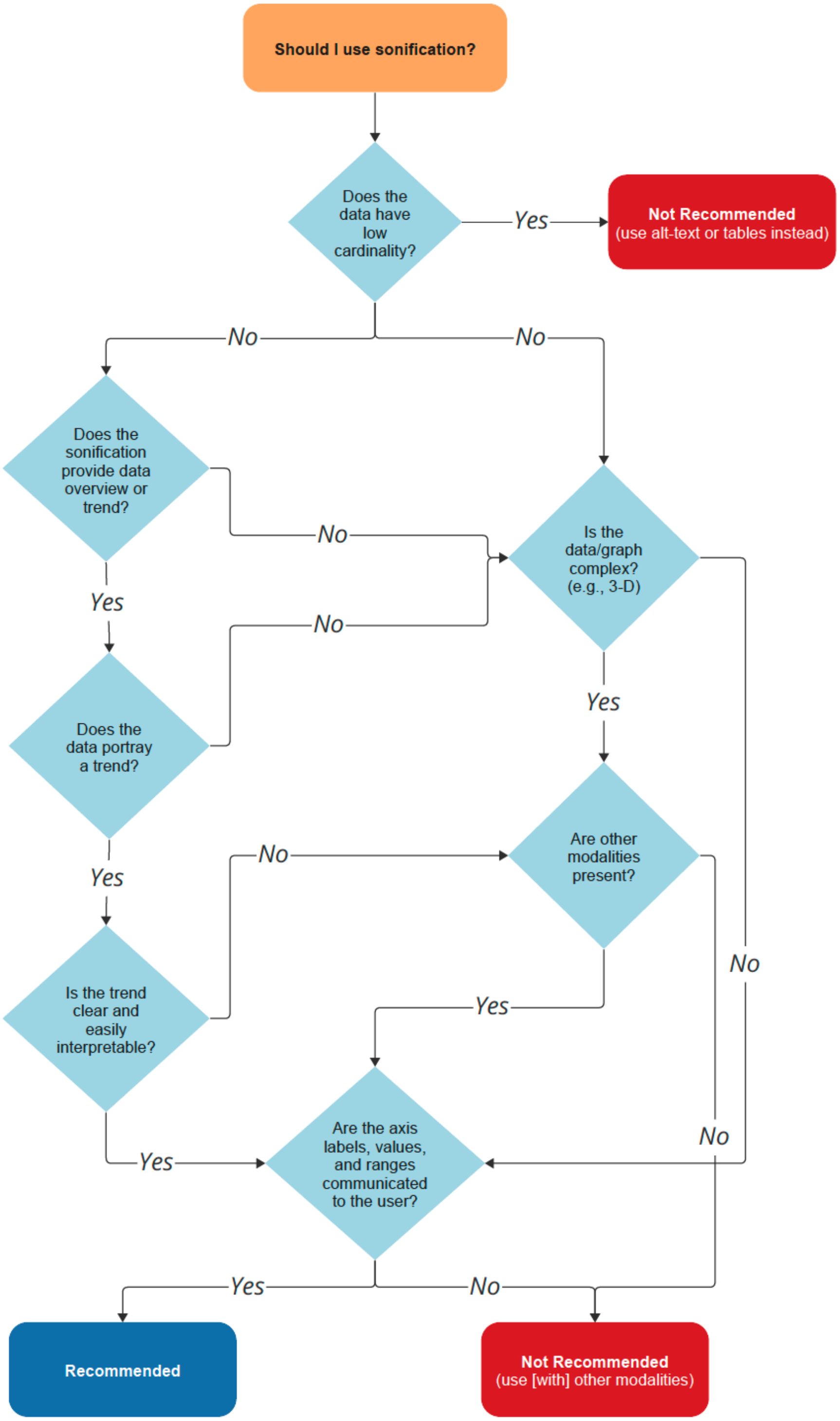


Figure 6.4: Directed acyclic graph (decision tree) to assist visualization creators in deciding to use sonification as a standalone accessibility technique for their data visualizations. Textual description of the decision tree is present in Section 6.5.3.

6.4.2.5 Areas of Improvement

The participants recognized three areas of improvement for sonification: (1) *Personalization* (customizing the auditory output for factors including speed and frequency); (2) *Identification of extrema/outliers* (identifying the maximum and minimum data points); and (3) *Multi-modality* (using a combination of different instrument sounds together with different frequencies to amplify the distinctions between data points). Among these findings, (1) motivates my work on enabling screen-reader users to customize their information consumption from online data visualizations. I discuss this work in Chapter 9.

6.5 Decision Tree

Utilizing the findings from the formative study and semi-structured interviews with screen-reader users, I developed a decision tree for visualization creators to assist them in appropriately choosing sonification to make online data visualizations accessible. Specifically, the decision tree considers the nuances discovered in the studies, serving as a *recommendation* tool for visualization creators. Figure 6.4 shows the tree.

6.5.1 Accessibility Measures

Following WCAG 2.1 (Caldwell et al. [2018]), I chose a color-vision-deficiency-friendly color palette using ColorBrewer 2.0 (Brewer et al. [2021]) and ensured the color contrast ratio was at least 4.5:1 using the WebAIM Contrast Checker (Web Accessibility In Mind [1999]). I also present a textual description of the decision tree below as an effort to make this visual content accessible to screen-reader users.

6.5.2 Usage

Visualization creators can use this decision tree to decide if providing data sonification would be helpful for screen-reader users. For example, for a stock market data visualization, a visualization creator would first determine the data cardinality to decide using sonification. For low cardinality, they can use alternative text or tables instead. Otherwise, they can move to the next step in the decision tree to determine if the visualization shows a trend. If the trend is easily interpretable, and the creator has communicated axis labels, values, and ranges to the user, they can use sonification as a standalone accessibility technique. Otherwise, they can use sonification in conjunction with other modalities.

6.5.3 Textual Description of Decision Tree

The decision tree begins with the Start Terminator symbol containing the text: “Should I use sonification?” A single arrow originates from it, connecting it with the Decision symbol $D1$ containing the text: "Does the data have low cardinality?” The “yes” arrow from $D1$ points to an End Terminator symbol that reads, “Not Recommended (use alt-text or tables instead).” Two “no” arrows originate from $D1$, connecting it with the Decision symbols $D2$ and $D3$, which read, “Does the sonification provide data overview or trend?” and “Is the data/graph complex? (e.g., 3-D),” respectively. The “yes” arrow from $D2$ connects it with the Decision symbol $D4$ that reads, “Does the data portray a trend?” The “no” arrows from $D2$ and $D4$ connect them to $D3$. The “yes” arrow from $D4$ connects it with Decision symbol $D5$, which reads, “Is the trend clear and easily interpretable?” The “no” arrow from $D5$ and “yes” arrow from $D3$ both lead to Decision symbol $D6$, which contains the text “Are other modalities present?” The “yes” arrows from $D5$ and $D6$ and the “no” arrow from $D2$ all connect to the final Decision symbol $D7$, which reads, “Are the axis labels, values, and ranges communicated to the user?” The “yes” arrow from $D7$ leads to the End Terminator that reads “Recommended.” The “no” arrows from $D6$ and $D7$ both connect them with the End Terminator that contains the text “Not Recommended (use [with] other modalities).”

6.6 `Sonifier` Library Enhancements

`Sonifier` is an open-source JavaScript library that I developed to generate a sonified response from two-dimensional single-series data (Sharif et al. [2022f,e]) (discussed in Chapter 5). Utilizing the findings from the user studies with screen-reader users, I enhanced the `Sonifier` library by (1) supporting additional sound types, and (2) modifying its default settings to those from the best-performing prototype.

I refactored and modularized the code to enable developers to create sonified responses using additional sound types, including *MonoSynthesizers* and *Envelopes*. (Previously, the `Sonifier` library only supported the creation of sonification using an OmniOscillator.) I further improved the customization for OmniOscillator by adding more configuration options, including “sourceType” (source of the oscillator; e.g., *am*), “baseType” (waveform of the oscillator; e.g., *sawtooth*), and “partialCount” (number of harmonics to generate the waveform, ranging from 1–32).

Additionally, I modified the default settings of the library based on the findings. Specifically, I changed "sourceType" to *am*, "baseType" to *square*, and "partialCount" to *8*. I made the code publicly available at the `Sonifier` library's open-source repository (Sharif et al. [2022e]).

6.7 Summary

In this chapter, I presented the assessment of the experiences of screen-reader users with sonification through a survey of 106 screen-reader users and semi-structured interviews with 12 screen-reader users. The findings show that sonification is a helpful technique for obtaining an overview of the data but not for exploring data granularly. The results also show that screen-reader users seldom encounter sonification in online data visualizations and usually only encounter it during research studies. The findings indicate that participants who *have* encountered sonification are generally positive and optimistic about its usefulness and potential, suggesting it is underutilized and should be incorporated more widely into online data visualizations of the appropriate type. To aid in understanding when sonification is beneficial as a standalone accessibility measure for data visualizations, I presented a decision tree to assist visualization creators in using sonification to make data visualizations accessible to screen-reader users.

Additionally, I investigated the usability and user-friendliness of sonified responses for screen-reader users via user studies. The results show that screen-reader users preferred distinct and non-continuous sonified responses generated using oscillators with *square* waveforms. Additionally, the follow-up interviews show that the participants identified the need to personalize the sonified responses. I utilized the findings to enhance the capabilities of `Sonifier` (Sharif et al. [2022e])—an open-source library that generates sonified responses from two-dimensional data.

Chapter 7

Drilled-Down Information Extraction*

In contrast to Chapter 5, which presents holistic information extraction for screen-reader users, this chapter discusses drilled-down experiences of screen-reader users with simple and complex online data visualizations. Specifically, in this chapter, I describe the formative studies I conducted to create an enhanced version of VoxLens (version 2.0) to support screen-reader users in extracting information granularly from online data visualizations. I present this VoxLens version's design, implementation, and evaluation.

7.1 Motivation

Data visualizations assist users in extracting information efficiently, helping people make informed life decisions concerning their health, finances, and activities. Recent work in politics (Zhao and Ye [2022]; Kubovics and Bielik [2021]), health (Zheng et al. [2021]; Grosjean et al. [2022]; Patrick and Junaini [2021]), finance (Tuarob et al. [2021]; Er and Sun [2021]), and business analytics (Biagi et al. [2022]; Zhang et al. [2021]) indicate the importance and wide adoption of online data visualizations.

However, the essential visual nature of data visualizations inherently disenfranchises people who cannot see (Marriott et al. [2021]; Lee et al. [2021]; Sharif et al. [2021]; Davis [2002]). These people include people who use screen readers (over 7.6 million people in the United States) to read the contents of their computer screens. In contrast, non-screen-reader users can explore data visualizations by rapidly interpreting visual patterns (Marriott et al. [2021]; Acarturk and Habel [2012]; Gegenfurtner et al. [2011]). Prior work has

---

*Parts of this chapter are adapted from Sharif et al. [2023b].

reported that alternative textual descriptions ("alt-text") for visualizations are often missing (Sharif et al. [2021]; Zong et al. [2022]). In cases when alt-text is present, screen-reader users spend 211% more time and are 61% less accurate in extracting information than their non-screen-reader user counterparts (Sharif et al. [2021]). Therefore, it is essential to find ways to make online data visualizations more accessible, efficient, and usable to screen-reader users.

Several prior works have attempted to improve the accessibility of online data visualizations (Sharif and Forouraghi [2018]; Kim et al. [2021a]; Ahmetovic et al. [2019b]; Giudice et al. [2012]; Shi et al. [2016]; Kim and McCoy [2018]; Sharif et al. [2022f]), such as by auto-generating alt-text (Sharif and Forouraghi [2018]; Mirri et al. [2017]; Kim et al. [2021a]) and enabling verbal information extraction (Sharif et al. [2022f]). While these tools contribute to the accessibility of online visualizations, they either focus on simple graphs (e.g., single-series bar graphs) or the extraction of mainly high-level ("holistic") information, such as extrema and averages. Therefore, their granular ("drilled-down") interactions, besides extracting and comparing individual data points, especially with complex visualizations, remain unexplored.[1]

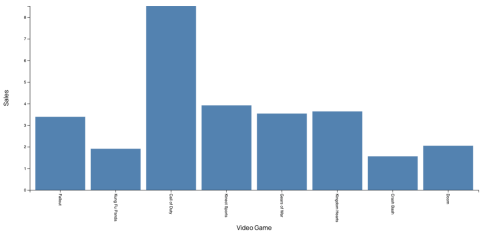
I employed a three-step process to explore drilled-down information extraction by screen-reader users from online data visualizations. First, I aimed to understand the granular information screen-reader users seek from simple and complex online data visualizations.[2] Then, I utilized these findings to develop taxonomies of the information sought by screen-reader users during their holistic and drilled-down explorations. Finally, using the taxonomies, I created an enhanced version of VOXLENS (Sharif et al. [2022f]) by supporting granular information extraction for screen-reader users.

My evaluation of the enhanced version shows that screen-reader users performed 5.6% *more* accurately than non-screen-reader users. (By contrast, using the original version of VOXLENS, screen-reader users performed 15% *less* accurately than non-screen-reader users (Sharif et al. [2022f]).) Additionally, these enhancements made VOXLENS a "promising tool", helping users extract information quickly and accurately from online data visualizations.

---

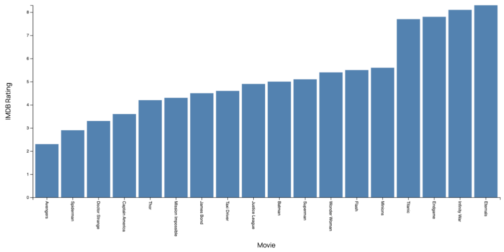
[1] I use the term "drilled-down" in line with its usage in the domain of *accessible* visualizations (Sharif et al. [2021, 2022f]; Zong et al. [2022]), as this term can have different meanings in different domains.

[2] I use the terms "simple visualizations" to refer to single-series bar graphs and "complex visualizations" to refer to geospatial maps and multi-series line graphs.


(a) Video Game Sales in North America (in millions)

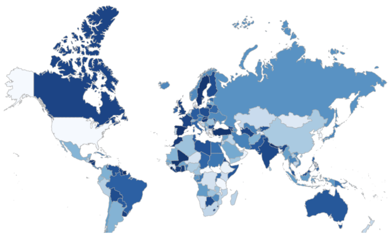
Q: *How many games are in this graph?*

A: **Data is from 8 video games. Say tell me factor levels for video games to hear all video games.**


(b) IMDB Ratings of movies

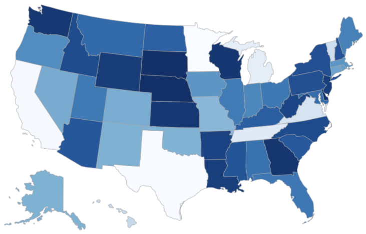
Q: *Tell me the range and the maximum*

A: **Found the following possible results in the data: IMDB Ratings range from 0 to 9. Maximum is 8.3 belonging to Eternals.**


(c) Percentage of fully vaccinated individuals per country

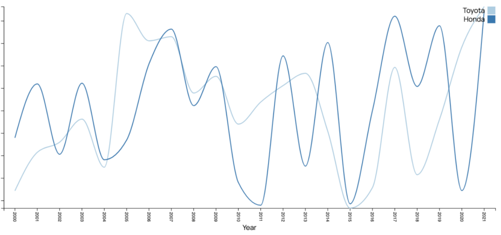
Q: *Compare Africa to North America*

A: **Average vaccination percentage for Africa is 62.23. Average vaccination percentage for North America is 64.53. Vaccination percentage for North America is greater than Africa.**


(d) Public libraries per US State

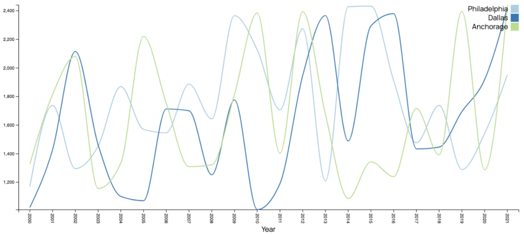
Q: *How many public libraries are in west coast?*

A: **Found the following possible results in the data: Total public libraries in the West Coast are 2,266.**


(e) Average price for Toyota and Honda per year since 2000

Q: *Toyota price in 2015 versus Toyota 2020*

A: **Average price for Toyota 2015 is 15,660. Average price for Toyota 2020 is 22,832. Average Price for Toyota 2020 is greater than Toyota 2015.**


(f) Cropland average value per city per year since 2000

Q: *Average cropland value for Dallas*

A: **Found the following possible results in the data: Average Cropland value for Dallas is 1,645.91.**

Figure 7.1: Sample visualizations shown to participants during our longitudinal study. Plots (a) and (b) show single-series bar graphs, (c) and (d) show geospatial maps, and (e) and (f) show multi-series line graphs. For each visualization, we show a question ("Q") that at least one participant asked during their interaction with it and the answer ("A") that our system, VOXLENS, responded with.

Table 7.1: Screen-reader participants in the role-playing study, their gender identification, age, screen reader, vision level, and diagnosis. Under the "G" (*Gender*) column, M = *Man*, F = *Woman*, and NB = *Non-binary*.

	G	Age	Screen Reader	Vision-Loss Level	Diagnosis
P1	M	36	JAWS	Complete blindness, Blind since birth	Leber Congenital Amaurosis
P2	M	58	JAWS	Complete blindness, Lost vision gradually	Cataracts and Glaucoma
P3	M	49	JAWS	Complete blindness, Lost vision gradually	Leber Congenital Amaurosis
P4	M	32	NVDA	Complete blindness, Blind since birth	Peters Anomaly
P5	W	32	NVDA	Complete blindness, Blind since birth	Retinopathy of Prematurity
P6	W	65	JAWS	Complete blindness, Lost vision gradually	Retinitis Pigmentosa
P7	W	68	Fusion	Partial blindness, Lost vision gradually	Stargaart's Maculopathy
P8	W	69	JAWS	Complete blindness, Blind since birth	Retinopathy of Prematurity
P9	W	52	JAWS	Complete blindness, Blind since birth	Retinopathy of Prematurity
P10	W	38	JAWS	Complete blindness, Blind since birth	Leber Congenital Amaurosis
P11	W	47	JAWS	Complete blindness, Lost vision gradually	Meningitis, Optic Neuropathy
P12	W	57	JAWS	Complete blindness, Lost vision gradually	Retinitis Pigmentosa

7.2 Formative Research

I conducted role-based and longitudinal user studies (Buchenau and Suri [2000]; Lewis-Beck et al. [2003]) with 12 and seven screen-reader users, respectively, to understand their information needs from online data visualizations. Utilizing these findings, I generated taxonomies of the information sought by screen-reader users for their holistic and drilled-down interactions. First, I present the methodology and results from the role-based and longitudinal user studies. Then, I discuss the process of taxonomy development.

7.2.1 Visualizations Selection

I selected three types of data visualizations based on their wide usage online: (1) *single-series bar graphs*; (2) *multi-series line graphs*; and (3) *geospatial maps*. I curated a set of 30 online data visualizations (10 for each type) based on the search results for "most popular data visualizations" and "most popular map visualizations" on Google. Figure 7.1 shows three of the 30 visualizations used in the longitudinal study.

7.2.2 Role-Playing User Study

I conducted a Wizard-of-Oz (Dahlbäck et al. [1993]; Hajdinjak and Mihelic [2004]) role-based user study (Buchenau and Suri [2000]; Lewis-Beck et al. [2003]) with 12 screen-reader users, acting as the "wizard" and simulating responses from a hypothetical screen reader. My goal was to elicit diverse perspectives and motivations for screen-reader users to perform information extraction granularly from online data visualizations. Building on and following recommendations from prior work (Liang and De Graaf [2010]; Hunt [1982]; Martin [1991]), I identified three roles that provided in-depth perspectives: (1) *explorer*; (2) *teacher*; and (3) *news reporter*. I was unsuccessful at finding screen-reader users who were actual teachers or news reporters, as is often the case with recruiting disabled participants, particularly those with specialties (Petrie et al. [2006]). Therefore, I used role-playing in the user study.

7.2.2.1 Participants

The participants were 12 screen-reader users (M=50.3 years, SD=13.6), recruited via word-of-mouth, snowball sampling, and advertisements through email distribution lists for disabled people (see Table 7.1). All but one participant had complete blindness; six participants had been blind since birth, and five had lost their vision gradually. The recruitment of participants ceased after reaching saturation of insights, as in prior work (Wyche and Grinter [2009]; Sharif et al. [2021, 2022f]). Participants received a $20 Amazon gift card for one hour of their time.

7.2.2.2 Procedure

I conducted hour-long user studies with the participants using Zoom videoconferencing and collected their demographic information. I took detailed notes during the sessions. I utilized Zoom's built-in features for recording and transcribing sessions. First, I presented participants with a summary of the visualization generated using the *Summary* mode from VoxLens (Sharif et al. [2022f]) to provide them with a holistic overview of the data. Then, I had the participants explore the data in the visualization by verbally asking questions, to which I responded as a "wizard," replicating the behavior of the *Question-and-Answer* mode of

Table 7.2: Screen-reader participants in the longitudinal study, their gender identification, age, screen reader, vision level, and diagnosis. Under the "G" (*Gender*) column, M = *Man*, W = *Woman*, and NB = *Non-binary*.

	G	Age	Screen Reader	Vision-Loss Level	Diagnosis
L1	M	43	JAWS	Complete blindness, Blind since birth	Microphthalmia
L2	M	57	JAWS	Complete blindness, Lost vision gradually	Retinitis Pigmentosa
L3	M	59	JAWS	Partial blindness, Lost vision gradually	Cataracts and Glaucoma
L4	M	49	JAWS	Partial blindness, Lost vision gradually	Leber Congenital Amaurosis
L5	W	39	JAWS	Complete blindness, Blind since birth	Leber Congenital Amaurosis
L6	W	53	JAWS	Complete blindness, Blind since birth	Retinopathy of Prematurity
L7	W	37	JAWS	Complete blindness, Blind since birth	Leber Congenital Amaurosis

VoxLens. Each participant interacted with nine visualizations randomly selected from the 30 visualizations I curated (three of each type). I randomized the order of the visualizations across participants.

I randomly assigned each participant a unique role (explorer, teacher, news reporter) for each of the three visualizations in each visualization type. As an "explorer," the participants interacted with the visualizations based on their curiosity and interests; as a "teacher" and "news reporter," they had to extract information assuming they were to present it to their students and news audience, respectively. I explained the definitions of these roles to the participants before the study. Hence, each participant took on each role three times. Participants did not portray any difficulty in assuming the roles. I counterbalanced the order of roles across participants. On average, the participants spent approximately five minutes interacting with each visualization.

7.2.3 Longitudinal Study

I conducted a longitudinal study with seven screen-reader users to gain further insights into their holistic and drilled-down information extraction behaviors. The study with each participant lasted 12 days, including a tutorial session and an hour-long follow-up interview.

7.2.3.1 Participants

The participants were seven screen-reader users (M=48.1 years, SD=8.7), recruited similarly to the role-based user study (see Table 7.2). Five had complete blindness, four of whom had been blind since birth. I compensated people with a \$230 Amazon gift card for participating in the longitudinal study of 12 days.

7.2.3.2 Procedure

On the first of 12 days, the participants took part in a tutorial session I administered using Zoom videoconferencing, in which I asked them to share their screen and computer sound. They interacted with a sample visualization using all modes of VOXLENS until they were comfortable. I used these modes because they were commonly used (Sonification and Summary modes) or a new interaction technique (Question-and-Answer mode) to make online data visualizations accessible.

On days 2–11, I asked the participants to interact daily with my curated visualizations using all three VOXLENS modes and extract information from the visualizations. All participants interacted with three visualizations per day (one of each type). Therefore, each participant interacted with 30 different visualizations, spending approximately 10 minutes with each visualization. I logged their interactions, including any queries or commands they issued to VOXLENS and the responses they received. Additionally, at the end of a participant's interaction with each visualization, I asked them to summarize the information they extracted from the visualization. All sessions were unsupervised. On the 12th and final day of the study, I conducted hour-long follow-up interviews. Specifically, I asked participants about their overall experiences with each type of visualization.

7.2.4 Data Analysis

I used a mixed-methods approach to analyze the data. Specifically, I employed both quantitative and qualitative methods to analyze the data from the role-based and longitudinal user studies. The primary sources of data were the interaction logs from VOXLENS and the semi-structured interviews.

7.2.4.1 Quantitative Analysis

My goal was to explore differences between the commands issued and the effects of different visualization types on the information sought by screen-reader users. Therefore, the independent variables were *Command Issued* (CMD)[3], representing the information sought by screen-reader users, and visualization type (VT). *Count* (CNT) was the dependent variable, calculated as the number of times each participant issued a command per chart type. I employed a mixed Poisson regression model (Gilmour et al. [1985]) to analyze CNT, as is standard practice for count data. Additionally, I included $Subject_r$ as a random factor (Frederick [1999]) to account for repeated measures on the same participants.

7.2.4.2 Qualitative Analysis

I used semantic thematic analysis (Patton [1990]; Maguire and Delahunt [2017]) to analyze the interview sessions from the role-playing and longitudinal studies. I employed Braun and Clarke's (Braun and Clarke [2006]) "essentialist" method for the thematic analysis, combining 33 initial codes into 18 axial codes. I separated the axial codes for each exploration type (holistic and drilled-down). Each exploration type contained nine axial codes. Finally, I classified the axial codes into two broader categories within each exploration type. Additionally, I calculated inter-rater reliability, expressed as percentage agreement among raters before resolving disagreements (Hartmann [1977]). The percentage agreement was 90.1%, demonstrating a high level of agreement between raters (Hartmann [1977]; Graham et al. [2012]).

7.2.5 Quantitative Results

Command Issued (CMD) had a significant effect on *Count* (CNT) (χ^2(10, N=1370)=8336.1, p<.001, Cramer's V=0.93). Specifically, the most issued command was the `Value` command, issued 28.6% of the time to extract the value of an individual data point.

I also examined the interaction between *Command Issued* and *Visualization Type* (*CMD* × *VT*), finding a significant effect (χ^2(20, N=1370)=5640.7, p<.001, Cramer's V=0.77). For a *Single-Series Bar Graph*, `Factor` was the most issued command (26.7%); for *Multi-Series Line Graph* and *Geospatial Map*, it was `Average` (22.0%) and `Value` (46.4%), respectively. (The `Factor` command enables users to get

---

[3] Full list of VoxLens commands is present in Table 4.3, Chapter 4

information about the independent and dependent variables.)

7.2.6 Qualitative Results

I present the findings for each exploration type (holistic and drilled-down) and discuss the taxonomy development process.

7.2.6.1 Holistic Exploration

Screen-reader users look for holistic information in data visualizations as an initial step before exploring data in detail (Sharif et al. [2021]). I identified two high-level categories for screen-reader users' holistic explorations: (1) *Summary Statistics* and (2) *Understanding Trends*. To obtain summary statistics, the participants looked for *extrema, averages, axis ranges, factor levels, medians,* and *sums* (in that order of frequency). For example, P9 compared the maximum and minimum data points, the two "extremes":

> `I would definitely write down the minimum and the maximum data point. So then I can compare the two extremes.` (P9)

Similarly, P4 was interested in learning about the average of COVID-19 cases in North America:

> `What's the rough average in North America?` (P4)

The participants also sought data trend information. Specifically, they looked for *overall trend, best-fit line,* and *correlation coefficient*. P12 looked for "visual representation" of the data:

> `What I would do is ask for some type of sonification of that graph that'll let me get a visual representation of it.` (P12)

Overall, these findings show that screen-reader users look for summary statistics and data trends to explore the data holistically. In the role-playing and longitudinal user studies, *extrema* were the most commonly sought information.

7.2.6.2 Drilled-Down Exploration

The user studies revealed that screen-reader users perform drilled-down explorations by extracting and comparing data points. These explorations were straightforward for single-series data (bar graphs, geospatial maps) with at most one independent factor. However, for multi-series data with multiple factors (i.e., multi-line graphs), participants not only performed the extraction and comparison *within* but also *across* different factors. For example, in a graph of housing prices over 10 years per U.S. state, participants looked for:

- *Extraction; within factors*: The data for a given state (e.g., Texas housing price average).
- *Comparison; within factors*: The data for a given state compared to another state (e.g., Texas vs. Oregon).
- *Extraction; across factors*: The data for a given state in a given year (e.g., Texas 2017).
- *Comparison; across factors*: The data for a given state in a given year compared to another year for a different state (e.g., Texas 2017 vs. Oregon 2019).

The participants categorized data as *regional, political, climate-related, population-related,* and *spoken-language-related* to extract information from geospatial data. For example, P1 was interested in finding the United States traffic congestion differences between eastern, western, northern, and southern regions:

> `I mean, is there a difference between the west and the east or north and south?` (P1)

Similarly, P2 categorized the data by political spectrum and P10 was interested in climate categorization:

> `The federal funding... What happened when the Republicans were in power and when the Democrats were in power? That would be interesting to check out.` (P2)

> `Usually, there’s not much information presented about that in maps, but I would like to see how states compare with climate, I mean cold, hot, whatever.` (P10)

Table 7.3: Taxonomy of the holistic information screen-reader users seek when exploring online data visualizations. Information types within each category are in descending order based on their sought frequency. For each information type, the "Query" column shows some of the questions that the participants asked.

Category	**Information Type**	**Query**
Summary Statistics	Extremum	*What country has the highest number?*
		What month was the stock market doing the best?
		Which country had the lowest overall in 2020?
	Average	*What is the national average?*
		What is the rough average of data in North America?
		What is the average of the top ten countries?
	Axis Ranges	*Can you tell me what is the x-axis?*
		What is the y-axis?
		What month does it start at and what month does it end?
	Factor Levels	*What countries are included in this graph?*
		How many companies are in the graph?
		What are the states in this graph?
	Median	*What's the median?*
		Tell me the median score
		What is the exact median?
	Sum	*What is the total for each year?*
		Can you tell me the total amount for 2020?
		How about the sum of all states?
Understanding Trends	Overall Trend	*How many companies have gone down in price?*
		Could I have some type of sonification of the graph?
		Is the United States currently going up?
	Best-Fit Line	*What is the best fit line for the United States?*
		Which line of best fit has the highest slope and lowest slope?
	Correlation Coefficient	*Can you tell me the correlation coefficient?*

Table 7.4: Taxonomy of the drilled-down information screen-reader users seek when exploring online data visualizations to extract and compare data points. Information types within each category are in descending order based on their sought frequency. For each information type, the "Query" column shows some of the questions that the participants asked. Under the *Categorization* category, *Regional, Political, Climate-Related, Population-Related,* and *Spoken-Language-Related* were applicable to geospatial maps, whereas *Factor-Levels-Related* was only applicable to multi-series line graphs.

Category	**Information Type**	**Query**
Categorization	Regional (Geospatial)	*Is there a difference between the east and west, or the north and south?*
		Is there a continent that has a higher life expectancy?
		Tell me the values of the Southern states as opposed to the Northwest.
	Factor-Levels-Related (Multi-series)	*How is California doing now, 10 years ago, and 20 years ago?*
		How was Texas between 2010 and 2015?
		Can we filter out the data to only Apple and Walgreens?
	Political (Geospatial)	*How do the trends during the Democratic presidential campaigns compare to the trends during Republican presidential campaigns?*
		Do socialist countries have higher rates?
		When Republicans were in power, did they increase compared to when Democrats were in power?
	Climate-Related (Geospatial)	*Do warmer climate states have higher values?*
		How are colder places compared to warmer places?
	Population-Related (Geospatial)	*Can you compare these two states by population?*
		Do the states with larger population have higher traffic rates?
	Spoken-Language-Related (Geospatial)	*Can we compare Spanish speaking countries to the English speaking countries?*

Ranking	Top	*Which countries are in the top 5%?*
		What are the five top countries in Western Europe?
		I'd like to see them all in order from the highest to the lowest.
	Bottom	*What are the bottom 10 companies in the graph?*
		What about the second and third lowest?
		Can we organize from lowest to the highest?
	Surrounding Average	*Which ones are in the middle?*
		What are the three countries that are closest to the average?
		What's the distribution, the countries around the average.

I found that the participants performed categorization by factor levels for multi-series data. P2 wanted to categorize the alcohol consumption data for Canada every five years:

> `What is the number in Canada for the past five years? Like, by five-year increments?` (P2)

I also found that screen-reader users ranked data based on the *top*, *bottom*, and *near-average* values. For example, P2 was interested in the average housing prices for U.S. states in 2021:

> `Well, I would like to get an order of the lowest to the highest, so I can maybe organize this in my head.` (P2)

The findings show that screen-reader users categorize and rank the data to extract and compare data points. I found regional and factor-level categorization the most frequently sought information for geospatial and multi-series data, respectively.

7.2.6.3 Taxonomy Development Process

I developed two taxonomies of the information sought by screen-reader users. The taxonomies contain three tiers: (1) *Category* (broader categories); (2) *Information Type* (axial codes); and (3) *Query* (participants questions). Table 7.3 shows the taxonomy for holistic exploration, organizing the categories and information types in the order of their frequency (most to least). To build the taxonomy for drilled-down exploration, I collected the information screen-reader users seek to extract data points. Table 7.4 shows the taxonomy for drilled-down exploration. Through these taxonomies, I produced generalizable knowledge that visualization creators and researchers can use to improve the accessibility of online data visualizations. To demonstrate the utility of the above discoveries and taxonomies, I created an enhanced version of my tool, VOXLENS (Sharif et al. [2022f]).

7.3 Enhanced Version of VOXLENS

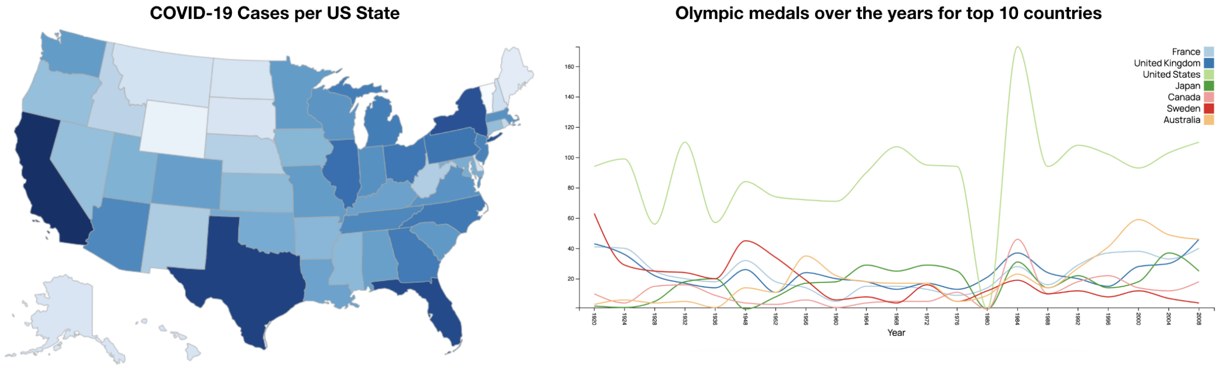
I selected the most frequently sought information types from the taxonomies to implement as additional features for VOXLENS. I encourage the readers to view Chapter 4 for the core design and functionalities of the original version of VOXLENS. From the taxonomy of the holistic information, *extremum, average, axis ranges, factor levels,* and *overall trend* were the most frequently used. Out of these five, *extremum, average* and *overall trend* were already implemented in the original version of VOXLENS. Therefore, I implemented *axis ranges* and *factor levels*. For drilled-down information extraction, the most frequently used information types were *regional categorization for geographic data*, *factor-level categorization for multi-series line graphs*, and *top* and *bottom* for the *ranking* category.

7.3.1 Factor-Level Categorization (Multi-Series)

I used the keyword matching algorithm from VOXLENS to support categorization by factor levels.[4] Specifically, I searched users' queries to find keywords matching the factor levels. For example, if a user said, "Tell me the housing price for Texas," my algorithm would identify "Texas" as the factor level and calculate the average housing prices in Texas for the past 10 years. I used "average" as the default command based

---

[4] In a graph showing average housing prices per U.S. state for the past 10 years, *state* and *year* are factors. For *state*, the levels are the 50 U.S. states; for *year*, they are the last 10 years. Housing prices are the dependent variable.



COVID-19 Cases per US State

Olympic medals over the years for top 10 countries

Q: *How is New England compared to Great Lakes?*

A: **Average cases for New England is 1,382,428.3. Average cases for Great Lakes is 452,551.89. Cases for New England are greater than Great Lakes.**

Q: *How many medals did Japan win in 2004?*

A: **Found the following possible results in the data: Medal Count for Japan in 2004 is 37.**

Figure 7.2: A screen-reader user's interaction with: (*left*) a geospatial map showing COVID-19 cases per US state, and (*right*) a multi-series line graph showing Olympic medals for the top 10 countries over multiple years. For each visualization, the user issues a question ("Q") to my system, VoxLens, which answers the user via their screen reader ("A").

on the findings from the role-playing and longitudinal studies. However, users can specify other statistical measures based on their needs (e.g., "Total housing price for Texas").

However, if a user said, "Tell me the housing price for Texas in 2017," the algorithm would identify "Texas" as the factor level for *state* and "2017" as the factor level for *year*. I employed the same strategy to perform comparisons between data points. To obtain *all* the data points, the participants suggested making the data available through a table, as listening to values from large data sets can induce high cognitive overload. I discuss this further in the subsection below on additional features.

In addition to line graphs, my enhancements are generalizable to other visualizations created using multi-series data. I also note that, currently, these enhancements are limited when categorizing by factor levels if the input query contains prepositions or adverbs of time (e.g., housing price for Texas five years "ago" or housing price for Texas "between" 2017 and 2019). Future work can utilize more advanced Natural Language Processing techniques to understand such queries.

7.3.2 Regional Categorization (Geospatial Maps)

The participants extracted and compared data points from geospatial maps by performing regional categorization. As all of the participants were from the United States, they categorized the data by regions within the United States (e.g., East Coast); for countries of the world, they grouped the data by continents (e.g., Asia). I further expanded the *state* module based on the National Geographic Society's classification of regions in the United States (National Geographic Society. [2023]). Specifically, I added the following regions: *Mountain West*, *Far West*, *Northwest*, *Northeast*, *Southeast*, *Midsouth*, *New England*, *Central*, *Southcentral*, and *Great Lakes*. As the modules are open-source and engineered to be scalable, developers can easily make necessary adjustments to the VOXLENS library to extend the modules to add more regions (and provinces). Figure 7.2 shows an interaction of a screen-reader user with a geospatial map after my enhancements. Similar to multi-series line graphs, the enhancements for geospatial maps are generalizable to other visualizations created using geospatial data.

7.3.3 Ranking Improvements

VOXLENS enables users to obtain the top N and bottom N data points, where N represents the number of data points. However, the existing algorithm only recognizes specific keywords to rank the data (e.g., "top" or "bottom"). Therefore, I extended the vocabulary of VOXLENS in the enhanced version to recognize more keywords (e.g., "most" and "least").

7.3.4 Factor Levels

I employed a two-step interaction design for factor levels. The first step presents the users with the count for the factor levels and the second step enables them to obtain the list. For example, in a visualization showing the stock market prices per company, if a user asks, "How many companies are present?" the response is, "Data is from three companies. Say 'tell me factor levels for companies' to hear all companies." If a user asks for factor levels, the response is, "Data is from three companies: Apple, Microsoft, and Google."

7.3.5 Range for the Dependent Variable

I added the functionality to obtain the range of the dependent variable, providing users with the minimum and maximum values. For example, if a user asks for the stock market price range, the response is, “Stock market price ranges from [minimum value] to [maximum value].”

7.3.6 Visually-Hidden Tables

The participants expressed an interest in obtaining the entire data set. However, they noted that through the Question-and-Answer mode of VoxLens, an entire data set could be cumbersome to process, especially for data sets with large cardinalities. They suggested presenting the data using visually hidden tables, a strategy employed by Google Charts (Developers [2014]). Therefore, I appended a table to the end of the visualization, which was visually hidden but accessible to screen readers.

7.3.7 Updated Configuration Options

I extended the functionality of the *Question-and-Answer* mode, as the other two interaction modes only assist in holistic exploration. I added two more parameters to the existing configuration options: “chartType” and “dataModule.” Five values for “chartType” are possible: (1) *bar* (single-series bar graphs); (2) *line* (single-series line graphs); (3) *scatter* (single-series scatter plots); (4) *map* (geospatial maps); and (5) *multiseries* (multi-series line graphs). The original version of VoxLens supported (1), (2), and (3). In this work, I introduced (4) (maps) and (5) (multi-series line graphs) as significant new features for VoxLens.

7.4 Evaluation

I evaluated the performance of the VoxLens enhancements through a task-based user study with 10 screen-reader users who used VoxLens and 10 non-screen-reader users who did not use any tools to aid in their interaction. I also administered the NASA-TLX questionnaire (Hart and Staveland [1988]).

Table 7.5: Screen-reader participants in the task-based study, their gender identification, age, screen reader, vision level, and diagnosis. SRU stands for Screen-Reader User. Under the "G" (*Gender*) column, M = *Man*, F = *Woman*, and NB = *Non-binary*.

	G	**Age**	**Screen Reader**	**Vision-Loss Level**	**Diagnosis**
S1	M	57	JAWS	Complete blindness, Lost vision gradually	Retinitis Pigmentosa
S2	M	58	JAWS	Complete blindness, Lost vision gradually	Cataracts and Glaucoma
S3	W	52	JAWS	Blind since birth, Complete blindness	Retinopathy of Prematurity
S4	M	49	JAWS	Blind since birth, Complete blindness	Leber Congenital Amaurosis
S5	M	37	JAWS	Blind since birth, Complete blindness	Leber Congenital Amaurosis
S6	W	38	JAWS	Blind since birth, Complete blindness	Leber Congenital Amaurosis
S7	M	33	NVDA	Blind since birth, Complete blindness	Peters Anomaly
S8	W	65	JAWS	Complete blindness, Lost vision gradually	Undiagnosed
S9	M	42	JAWS	Blind since birth, Complete blindness	Microphthalmia
S10	W	30	NVDA	Blind since birth, Complete blindness	Optic Nerve Hypoplasia

7.4.1 Participants

I recruited 10 screen-reader users and 10 non-screen-reader users for the study, advertising via word-of-mouth, snowball sampling, and email distribution lists (see Table 7.5). No participants had partaken in the role-playing or longitudinal user studies. Among screen-reader users, the average age was 46.1 years (SD=11.8). Six identified as men and four as women. All participants had complete blindness; seven were blind since birth, and three had lost their vision gradually. Eight used JAWS as their primary screen reader and two used NVDA. For non-screen-reader users users, the average age was 45.9 (SD=8.3) years. I did not find a statistically significant difference between the ages of the two participant groups (t(18)=0.04, $n.s.$). I compensated screen-reader users with a \$20 Amazon gift card and non-screen-reader users with a \$15 Amazon gift card for 45–60 minutes and 20–35 minutes of their time, respectively.

7.4.2 Apparatus

I conducted the study online using the study platform from my prior work (Sharif et al. [2022f]), created using the React framework (Inc. [2013]). I randomly selected nine data visualizations from my 30 curated visualizations (see Section 7.2.1; three for each visualization type). I implemented these visualizations following the WCAG 2.0 Guidelines (Caldwell et al. [2008]). The performance of users with VOXLENS did not significantly differ between visualization libraries (Sharif et al. [2022f]). Therefore, I generated all visualizations using D3 (Bostock et al. [2011]). All visualizations supported interactive features (e.g., hover and click), as my goal was to accurately replicate the current behavior of these visualizations on the web.

I finalized the question categories for the study based on Brehmer and Munzner's task topology (Brehmer and Munzner [2013]) and prior work on the type of questions users ask of graphs (Kim et al. [2020]; Sharif et al. [2021]; Brehmer et al. [2018]). I intentionally did not base the questions on the findings from the role-playing and longitudinal studies to improve the external validity and avoid bias in the results. Overall, I identified five categories for each visualization:

1. *Order Statistics*: Extraction of the maximum/minimum data point, chosen randomly (e.g., "Which state has the minimum number of cases in this visualization?").
2. *Symmetry Comparison*: Identification of the relationship between two data points (e.g., "Are the cases of state 'Oregon' greater, lesser, or equal compared to 'Michigan'?").
3. *Value Retrieval*: Extraction of the value of an individual data point (e.g., "What is the number of cases for state 'California'?").
4. *Ranking*: Identification of the highest/lowest ranked data points, chosen randomly (e.g., "Which of these states is not in the top three based on cases?").
5. *Chart-Type Specific Questions*:
    - *Factor Levels*: Extraction of the total number of levels for any given independent variable (e.g., "How many countries are shown in this visualization?"). I randomly selected the independent variable for multi-series line graphs.
    - *Symmetry Comparison by Region*: Identification of the relationship between two regions (e.g., "Are the cases on the east coast greater, lesser, or equal compared to the west coast?"). I only

asked this question for geographic map-based visualizations.

All questions were multiple-choice with four choices: the correct answer, two incorrect answers, and the option for "Unable to extract information." Following the study design from prior work (Sharif et al. [2021, 2022f]), I determined the incorrect answer choices programmatically, randomly choosing two data points for categorical values and two integers using a random number generator for numerical values. I randomized the order of the four choices for each question.

7.4.3 Procedure

I conducted an over-the-shoulder-style user study (Twidale [2005]) and asked the participants to share their screens and make their screen reader's audio outputs audible using Zoom videoconferencing. The study sessions were unsupervised for non-screen-reader users. At the beginning of the study, I collected preliminary information from participants. In the next step, I engaged with the screen-reader users in an interactive tutorial session, introducing the operations and functions of the enhancements (e.g., activating the tool and issuing commands). Using a sample visualization of a single-series bar graph, I assisted the screen-reader users in extracting information from the visualization using the *Question-and-Answer* mode. I shared sample questions for them to ask and guided them until they expressed confidence in their familiarity with the system. I did not present a tutorial session to non-screen-reader users.

After the tutorial session, each participant completed five study tasks for each visualization type, totaling 5×3=15 study tasks. Each task comprised three steps. Step 1 contained the question to explore; step 2 displayed the question and visualization; step 3 presented the question with four answer choices. I randomized the order of the study tasks and the order of the multiple-choice responses. I did not interact with the participants while they performed the tasks. Finally, I asked them to fill out the NASA-TLX workload questionnaire (Hart and Staveland [1988]). For screen-reader users, study sessions took 45–60 minutes; for non-screen-reader users, they took 20–35 minutes.

7.4.4 Design and Analysis

The experiment was a 2×3 mixed-factorial design with the following factors and levels: (1) *Screen-Reader User (SRU)*, between-Ss.: {yes, no}, and (2) *Visualization Type (VT)*, within-Ss.: {single-series bar graph, multi-series line graph, geospatial map}. I used the SRU factor to empirically explore the gap in information access between screen-reader- and non-screen-reader users. (I did not compare my enhancements to the original VOXLENS to avoid making a strawman comparison.) Hence, the screen-reader users interacted with visualizations using the enhanced version, and non-screen-reader users interacted with the visualizations as they usually would.

I used *Accuracy of Extracted Information* (AEI) and *Interaction Time* (IT) as the dependent variables. In the analysis, AEI was coded as "1" for "accurate" when a user answered the question correctly and "0" for "inaccurate." Additionally, I calculated IT as the total task completion time. I used a mixed logistic regression model (Gilmour et al. [1985]) and a linear mixed model (Frederick [1999]; Littell et al. [1998]) to analyze AEI and IT, respectively. I used the above factors, their interactions, and a covariate to control for *Age* in my statistical models. I also included $Subject_r$ as a random factor to account for repeated measures (Frederick [1999]). Therefore, the statistical model terms were: $SRU + VT + SRU \times VT + Age + Subject_r$. I tested the participants over three *Visualization Type* (VT) conditions, resulting in a total of 3×5=15 trials per participant. With 20 participants, I had a total of 20×15=300 study trials.

7.4.5 Qualitative and Subjective Evaluation

To qualitatively assess the enhancements, I conducted follow-up interviews with all screen-reader users (N=10). Specifically, I asked them about their overall experience, liked features, areas for improvement, and any issues they encountered during their interactions. To assess the subjective ratings, I administered the NASA-TLX workload questionnaire (Hart and Staveland [1988]) with all participants (N=20).

7.5 Results

In this section, I provide the qualitative, quantitative, and subjective results from my evaluation of the enhanced version of VOXLENS.

7.5.1 Quantitative Results

I present the results of the task-based user study assessing the *Accuracy of Extracted Information* (AEI) and *Interaction Time* (IT) for screen-reader- and non-screen-reader users with online data visualizations.

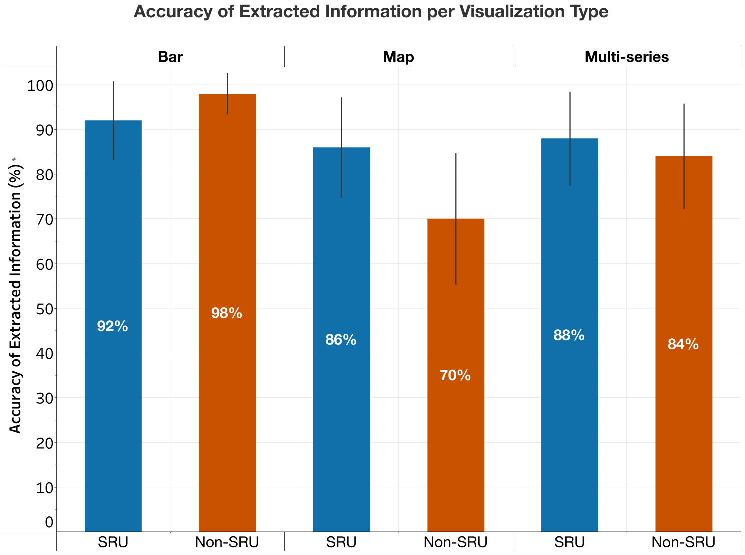


Figure 7.3: Accuracy of Extracted Information (AEI), as a percentage, for SRUs with VoxLens (N=10) and non-SRUs without VoxLens (N=10) by Visualization Type (VT). The percentage represents the "accurate" answers (higher is better). Error bars represent mean ±1 standard deviation.

My goal was to investigate the performance of the enhancements with users and not to assess their cognitive or intellectual abilities.

7.5.1.1 Accuracy of Extracted Information (AEI)

The results do not show a significant effect of *Screen-Reader User* (SRU) on AEI overall ($p \approx .906$), indicating that AEI was not detectably different between the two *Screen-Reader User* groups. In fact, using my enhancements, screen-reader users extracted information 5.6% more accurately on average than non-screen-reader users, although this was not statistically significant. However, in an evaluation before these enhancements (Sharif et al. [2021]), non-screen-reader users *did* outperform screen-reader users by 62%. Therefore, such a non-significant result is noteworthy because it indicated a "closure of the gap."

There was a significant effect of *Visualization Type* (VT) on AEI overall ($\chi^2(2, N{=}300){=}9.35$, $p<.05$, Cramer's V=0.12). This result indicates that AEI differs significantly between different visualization types. Figure 7.3 and Table 7.6 show the AEI percentages across different VT. For both screen-reader- and non-screen-reader users, *Single-Series Bar Graphs* had the best performance (92% and 98% AEI, respectively), followed by *Multi-Series Line Graphs* (88% and 84% AEI, respectively) and *Geographic Maps* (86% and

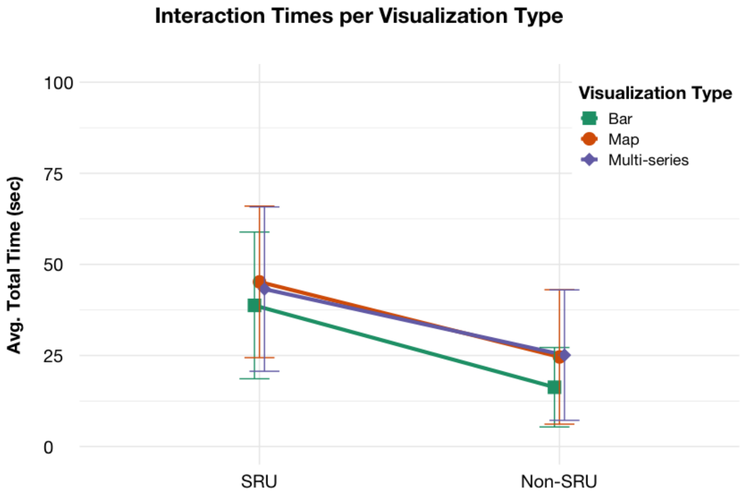


Figure 7.4: Interaction Times (IT), in seconds, for SRUs using the VoxLens enhancements (N=10) and non-SRUs without VoxLens (N=10) by Visualization Type (VT). Lower is better (faster). Error bars represent mean ±1 standard deviation.

Table 7.6: Results for the $N = 150$ questions asked of screen-reader users (SRUs) using VoxLens with the enhancements and non-SRUs without VoxLens for each level of *Visualization Type*. Here, N is the total number of questions asked, AA is the number of "accurate answers," and $AA(\%)$ is the percentage of "accurate answers." Higher is better.

		SRU		***Non-SRU***	
	N	AA	AA **(%)**	AA	AA **(%)**
Overall	150	133	88.7	126	84.0
Single-Series Bar Graph	50	46	92.0	49	98.0
Multi-Series Line Graph	50	44	88.0	42	84.0
Geospatial Map	50	43	86.0	35	70.0

Table 7.7: Statistical test results for *Accuracy of Extracted Information* (AEI) from screen-reader users (SRUs) using VoxLens with the enhancements (N=10) and non-SRUs without VoxLens (N=10). SRU is the screen-reader factor and VT is the visualization type factor. Cramer's V is a measure of effect size (Ferguson [2016]).

	N	χ^2	p	**Cramer's V**
SRU	20	0.01	.903	.00
VT	20	9.33	$< .05$	.12
$SRU \times VT$	20	3.92	.141	.08
Age	20	0.24	.626	.02

Table 7.8: Statistical test results for *Interaction Time* (IT) from screen-reader users (SRUs) using VoxLens with the enhancements (N=10) and non-SRUs without VoxLens (N=10). SRU is the screen-reader factor and VT is the visualization type factor. Partial eta-squared (η_p^2) is a measure of effect size (Cohen [1973]).

	df_n	df_d	F	p	η_p^2
SRU	1	16.98	36.94	$< .001$	0.69
VT	2	275.34	6.14	$< .05$	0.04
$SRU \times VT$	2	275.15	2.46	.087	0.02
Age	1	16.95	8.21	$< .05$	0.33

70% AEI, respectively). I did not investigate the causation of these differences within each visualization type. Additionally, I investigated the effects of *Age* and the *SRU* × *VT* interaction but did not find significant effects on AEI (see Table 7.7).

7.5.1.2 Interaction Time (IT)

Interaction times were initially conditionally lognormal, a common occurrence with temporal measures (Limpert et al. [2001]). Therefore, I applied a logarithmic transformation before conducting the analysis, following standard practice for time measures (Hoyle [1973]; Limpert et al. [2001]; Berry [1987]). Anderson-Darling goodness-of-fit tests of normality (Anderson and Darling [1954]) confirmed that log-transformed interaction times were conditionally normally distributed ($p \approx .117$). For ease of understanding, I display plots of IT using the non-transformed values.

Screen-Reader User (SRU) had a significant main effect on *Interaction Time* (IT) (F(1,18)=36.94, p<.001, η_p^2=0.69). Specifically, the average IT for screen-reader users was 42.4 seconds (SD=21.2). For non-screen-reader users, it was 22.0 seconds (SD=16.5). The average IT for screen-reader users was 92.8% higher than non-screen-reader users. I also found a significant main effect of *Visualization Type (VT)* (F(2,275.3)=6.14, p<.05, η_p^2=0.04) on IT. These results indicate that IT significantly differed among visualization types (VT). I also examined the interaction between *SRU* × *VT*, but did not find a significant effect (see Figure 7.4 and Table 7.8). For screen-reader users, *Geospatial Map* had the maximum interaction time; for non-screen-reader users, it was the *Multi-Series Line Graph*. *Single-Series Bar Graph* had the minimum interaction time for both groups.

Age had a significant effect on IT (F(1,16.9)=8.21, p<.05, η_p^2 =0.33), indicating that IT differed significantly across the ages of the participants. Participants over 50 years old had 44.0% higher interaction times than those under the age of 50.

7.5.2 Qualitative Results

Through follow-up interviews with all the screen-reader users (N=10), I assessed the usability and usefulness of my enhancements. Specifically, I asked them about the features they liked and for any areas of improvement. Additionally, I observed their interactions to identify system errors and user workarounds.

7.5.2.1 Liked Features

Nine out of 10 screen-reader users identified the "interactive dialogue" feature as one of the features they liked. Eight participants appreciated that the enhancements were "intuitive" and "easy to use." Three participants highlighted the adaptiveness of the enhancements, stating that VoxLens "adjusts to your question," and that it is "suitable for people of all ages and backgrounds." Two participants found the enhancements innovative, something they had "never seen before." One participant liked the quick responses.

7.5.2.2 Areas of Improvement

The participants recognized five areas of improvement: (1) *adding a repeat command* (would enable users to re-hear the response from the previous query); (2) *building a "playground" environment* (would enable users to learn more about the tool by trying out different commands and features); (3) *making the response more succinct*; (4) *enabling responses to be explored in text form* (would append the auditory response as text on the web page, enabling them to copy it); and (5) *increasing the query input time* (would enable users to issue longer queries without feeling rushed). I address these suggestions in subsequent versions of VoxLens, discussed in their respective chapters.

7.5.2.3 System Errors

From analyzing participants' interaction logs, I recognized system errors stemming from the limitations of the keyword matching algorithm and voice recognition—two fundamental components of VoxLens. My system does not process users' queries when my algorithm does not find a keyword match—an unfortunate and known limitation of voice assistants (Springer and Cramer [2018]; Cambre et al. [2019]; Myers et al. [2018]). For example, "tell me the three countries doing amazing in vaccinating people" was not recognized by my system, but "tell me the top three countries by vaccination percentages" was correctly processed.

7.5.2.4 User Workarounds

The participants employed workarounds to extract information due to the limitations mentioned above. Specifically, when my system could not process a user's query involving a comparison between data points, the participants asked for the value of each data point separately and computed the comparison mentally. Although the users successfully extracted the information, they issued multiple commands instead of a single command, consequently increasing their interaction time. I address these areas of improvement in subsequent VoxLens versions.

7.5.3 Subjective Workload Ratings

I collected subjective workload ratings for both user groups using the NASA Task Load Index questionnaire (NASA-TLX) (Hart and Staveland [1988]). The NASA-TLX questionnaire records users' perceived task workload on six scales: *mental demand, physical demand, temporal demand, performance, effort,* and *frustration*. For all scales, lower is better, as it corresponds to a lesser perceived workload. I performed the nonparametric Aligned Rank Transform procedure (Wobbrock et al. [2011a]; Elkin et al. [2021]) to statistically analyze the effects of *SRU* (levels: *yes, no*). I did not find a significant effect of *SRU* on any of the six dimensions, suggesting comparable workload levels.

7.6 Summary

In this chapter, I presented the role-playing and longitudinal user studies with screen-reader users I conducted to understand their holistic and drilled-down information extraction needs from simple and complex online data visualizations, including multi-series line graphs and geospatial maps. I provided the taxonomies of the information sought by screen-reader users in their interactions with online data visualizations, generated using the findings from the formative studies. Then, I presented the design and functionalities of an enhanced version of VOXLENS that utilizes these taxonomies to support screen-reader users in extracting information from complex data visualizations in a granular fashion. The evaluation of the enhancements shows improved accuracy of extracted information and interaction times for screen-reader users compared to the original VOXLENS. Additionally, using these enhancements, screen-reader users "closed the gap" compared to non-screen-reader users in extracting information accurately from online data visualizations. Closing this gap such that the accuracy of extracted information from online data visualizations is not detectably different between screen-reader users and non-screen-reader users represents a major advancement in the accessibility of online data visualizations.

Chapter 8

Uncertainty in Data Visualizations*

This chapter presents a version of VoxLens (version 3.0) that further enhances the version introduced in Chapter 7 by enabling screen-reader users to obtain information about uncertainty in online data visualizations. I discuss the formative study I conducted to understand the needs and preferences of screen-reader users in obtaining this information and to make informed design decisions for this version of VoxLens. I further provide its design, functionality, and implementation details.

8.1 Motivation

"*In most cases, visualization is often thought to be an abstract visual representation of the actual information. However, that's not always possible in real-world scenarios, as no data can be sufficiently reliable and complete. There is bound to be some degree of uncertainty associated with any data*" (Kamal et al. [2021]). Due to this reason, communicating the uncertainty in data visualizations is one of the top visualization research problems (Johnson [2004]). The visualization community has widely researched and discussed this topic (Brodlie et al. [2012]; Hullman [2019]; Sanyal et al. [2009]; Shneiderman et al. [2013]; Dimara et al. [2022]; Andrienko et al. [2020]), shedding light on the significance of providing information on data uncertainty in visualizations, as visualizations portraying only the "means" may be misleading and potentially dangerous (Brodlie et al. [2012]; Andrienko et al. [2020]; Hullman [2019]; Shneiderman et al. [2013]; Dimara et al. [2022]). However, communicating uncertainty to screen-reader users via non-visual

---

[*] Parts of this chapter are adapted from Sharif et al. [2023c].

Table 8.1: Screen-reader participants, their gender identification, age, screen reader, vision-loss level, and diagnosis. Under the "G" (*Gender*) column, M = *Man*, W = *Woman*, NB = *Non-binary*, and "-" means preferred not to disclose.

	G	**Age**	**Screen Reader**	**Vision-Loss Level**	**Diagnosis**
P1	M	43	JAWS	Blind since birth	Microphthalmia
P2	M	59	JAWS	Lost vision gradually	Cataracts and Glaucoma
P3	W	53	JAWS	Blind since birth	Retinopathy
P4	M	37	JAWS	Blind since birth	Leber Congenital Amaurosis
P5	M	55	JAWS	Blind since birth	Leber Congenital Amaurosis
P6	W	50	JAWS	Lost vision gradually	Retinitis Pigmentosa
P7	W	36	JAWS	Blind since birth	Retinopathy of Prematurity
P8	GF	28	NVDA	Blind since birth	Leber Congenital Amaurosis
P9	W	40	NVDA	Lost vision gradually	Stargardt Disease
P10	-	57	VoiceOver	Lost vision gradually	Retinopathy of Prematurity
P11	M	32	JAWS	Lost vision gradually	Retinal detachment
P12	W	41	VoiceOver	Lost vision gradually	Functional Neurological Disorder
P13	M	76	JAWS	Lost vision gradually	Retinitis Pigmentosa
P14	W	24	JAWS	Blind since birth	Astrocytoma
P15	W	65	JAWS	Lost vision gradually	Retinitis Pigmentosa
P16	W	72	JAWS	Lost vision gradually	Retinitis Pigmentosa

means remains unexplored, exacerbating the disenfranchisement these users face due to inaccessible data visualizations (Fan et al. [2023]).

The inaccessibility of online data visualizations causes screen-reader users to spend 211% more time interacting with and extracting information 62% less accurately from online data visualizations than non-screen-reader users (Sharif et al. [2021]). Several techniques to make online data visualizations accessible exist, including alternative text (Sharif and Forouraghi [2018]; Mirri et al. [2017]; Kim et al. [2021a]), tables (Developers [2014]; Choi et al. [2019]), audio graphs (Fan et al. [2022]; Siu et al. [2022]; Sharif et al. [2022d]), and verbal question-and-answer-based information extraction (Sharif et al. [2022f]). While Fan *et al.* (Fan et al. [2023]) have briefly discussed this subject, no prior work has conveyed the uncertainty in data visualizations through non-visual means using any of these techniques, which has limited these users in making informed decisions (Weiskopf [2022]). Additionally, no prior work has explored screen-reader users' preferences in obtaining this information. My exploration is the first empirical work to communicate uncertainty in data visualizations to screen-reader users.

My goal was to create generalizable knowledge for visualization creators to communicate uncertainty in data visualizations to screen-reader users. Therefore, I conducted semi-structured interviews with 16 screen-reader users. The findings show that participants were unfamiliar with uncertainty before partaking in the user study. However, after developing a conceptual understanding of uncertainty, they expressed interest in obtaining this information holistically and granularly, in plain language, with an option to include statistical terms for expert users. I utilized these findings to develop an enhanced version of VOXLENS.

8.2 Formative Research

I conducted semi-structured interviews with 16 screen-reader users to understand their needs and preferences for obtaining information on uncertainty in online data visualizations via non-visual means. My research questions were:

1. Are screen-reader users conceptually familiar with data uncertainty in visualizations?
2. Did participating in the study help advance their knowledge of data uncertainty in visualizations?
3. How do they prefer the information on data uncertainty in visualizations conveyed to them?

8.2.1 Participants

I recruited 16 screen-reader users for the interviews via word-of-mouth advertisement and snowball sampling (see Table 8.1). Eight participants identified as women, six as men, one as gender-fluid, and one preferred not to disclose. Their average age was 48.0 years (SD=15.4). Twelve participants used JAWS screen reader (Scientific [1995]), two used NVDA (Access [2006]), and two used VoiceOver (Inc. [2005]). Seven participants were blind since birth, whereas five had lost vision gradually, with nine participants having complete blindness. I compensated the participants with a $20 Amazon gift card for 45 minutes.

8.2.2 Procedure

I conducted the interviews online using Zoom. I used its built-in features to record and transcribe the interviews. As a preliminary step, I collected demographic information from the participants. To create a baseline for comparison, I gathered their subjective ratings for perceived importance of data uncertainty information using a Likert scale ranging from 1 to 7 (1 being "not important at all" and 7 being "extremely important"). Next, I inquired about their conceptual familiarity with uncertainty in data visualizations. After collecting their response, I provided them with the definition of data uncertainty and examples to assist them in understanding the concept and its importance in informed information retrieval. Then, I asked the participants about their preferences in obtaining uncertainty information to answer my last research question. Lastly, I collected their subjective ratings for the perceived importance of data uncertainty information again to examine the difference in their response pre- and post-study.

8.2.3 Analysis

I used thematic analysis (Braun and Clarke [2006]) to analyze the interviews. I followed a semantic approach (identifying themes within the explicit or surface meanings of the data) (Patton [1990]), guided by an essentialist paradigm (focusing on reporting the experiences of the participants) (Potter and Wetherell [1987]; Wooffitt and Widdicombe [1995]). I developed 12 initial codes, transforming them into eight axial codes. The axial codes revealed two themes and one suggestion as answers to the research questions.

8.2.4 Results

I discuss the two themes and one suggestion that emerged as findings from the semi-structured interviews.

8.2.4.1 Unfamiliarity with Uncertainty in Data Visualizations

The findings showed that the participants lacked understanding of or were unfamiliar with uncertainty in data visualizations. Therefore, during the interviews, I explained the meaning of "uncertainty in data visualizations" using plain language and provided examples to familiarize them with the concept. After gaining familiarity, they expressed interest in learning more than the average values to draw more informed conclusions from the data. For example, P4 said:

> `Now that I understand the term, I think it can be very handy actually.` (P4)

To further analyze the perceived importance of data uncertainty, I collected the participants' subjective ratings using a 1- to 7-point Likert scale before and after the interviews. 1 represented "not important at all" and 7 was "extremely important." The median of the participants' perceived importance before and after partaking in the study was 4.0 (IQR=3.3) and 5.5 (IQR=1.3), respectively. This difference was statistically significant based on the Wilcoxon signed-rank test (Z=3.47, p<.001) (Wilcoxon [1945]), highlighting the significance of introducing users to the meaning and benefits of uncertainty in data visualizations.

8.2.4.2 Summary First, Details Later

The participants shared their preference for exploring the data holistically first and then in a drilled-down manner, validating the findings from my prior work (Sharif et al. [2021, 2023b]). Specifically, the participants suggested a twofold solution. For example, P16 advised noting in the summary (or "alt-text") that users should cautiously interpret the data when high variation exists in the aggregated means and then provide opportunities for users to dig deeper into the data:

> `Let's say the description says, this is the chart about, and this is the axis, these are the ranges, stuff like that. But also, use this data with caution because this data might not be representative of the population. And um, and then further, people are given the opportunity to, kind of, dig deeper and actually see all of those data points if they wanna.` (P16)

To delve deeper into the data, the participants suggested using measures, including tabular representation of the entire dataset and textual details using the "longdesc" attribute. For example, P11 expressed their opinions on using a data table:

> `I'd use a table, a simple standard table with rows and cells and put the values in rows and columns. And then try to find`

> `any data points and then manually go through them to pull out what I need.` (P11)

Similarly, P2 shared a possible solution involving the "longdesc" attribute:

> `Longdesc will bring up a page for that only, usually only screen readers can grab. It's like, the chart is actually a link to a description, which would describe all that stuff, like uncertainty and stuff.` (P2)

To convey data uncertainty, the participants showed interest in knowing statistical information, including the margin of errors and confidence interval for each aggregated data point. For example, P1 expressed the benefit of such information for accurately understanding polling data:

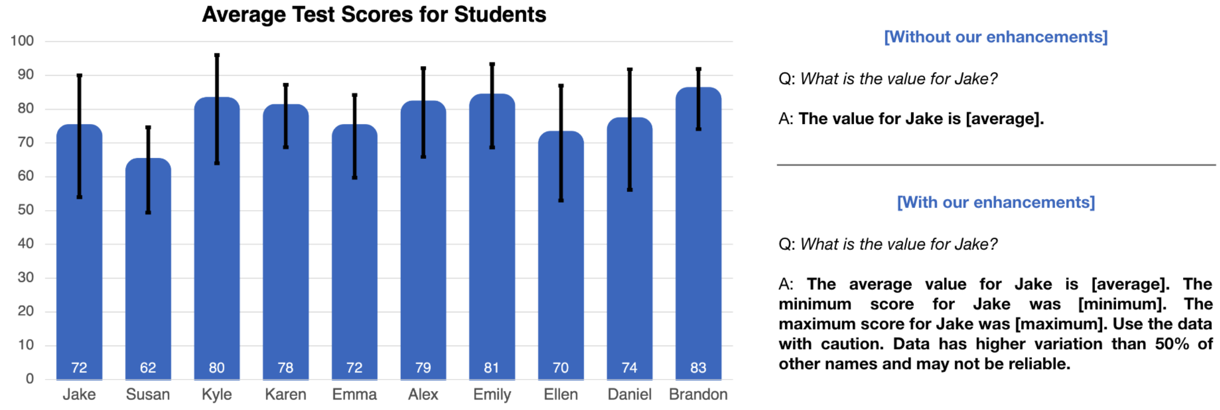
> `I'm really big into the election season. So many times, you know, margin of errors, confidence level, you know, will help me in trying to understand the pool numbers and see how accurate are they.` (P1)

8.2.4.3 Keep Statistical Jargon for the Experts

The participants articulated the need to adapt uncertainty information based on users' expertise and comfort level with statistical terms and jargon. For example, P12 recommended using plain language and common words for easier understanding and an option to get specific statistical information for expert users:

> `I guess a lot of it depends on the sophistication of your user. You know, if they have a sophisticated understanding of statistics, they might wanna really know what is the error bar value or what is the confidence interval or they might really need that level of specificity. If it's a layperson, um, then you might just be able to say, you know, use this with caution, or something like that.` (P12)

I utilized these findings to enhance the functionalities of VOXLENS, which I discuss in the section below.



Q: *What is the value for Jake?*

A: **The value for Jake is [average].**

Q: *What is the value for Jake?*

A: **The average value for Jake is [average]. The minimum score for Jake was [minimum]. The maximum score for Jake was [maximum]. Use the data with caution. Data has higher variation than 50% of other names and may not be reliable.**

Figure 8.1: A screen-reader user's interaction with a data visualization of the average test scores for students. "Q" and "A" represent a question asked by a screen-reader user and the answer issued to them with and without enhancements to VoxLens, respectively. Error bars represent mean ±1 standard deviation.

8.3 Enhanced Version of VoxLens

In this section, I present a new version of VoxLens, created using the findings from the formative study. I encourage readers to peruse the versions of VoxLens presented in Chapters 4 and 7 to understand the design and functionalities of VoxLens that existed before this version.

As a preliminary step, I determine and convey uncertainty in online visualizations by calculating the 50^{th} percentile (median) of the coefficient of variation (CV) value, which is the ratio between the standard deviation (provided by the visualization creator) and the dependent variable value (mean). Motivated by the suggestions from the participants in the interviews, I recommend users cautiously interpret the data for values higher than the median. Additionally, I relay the minimum and maximum values if provided by the visualization creators (see Figure 8.1). I recommend visualization creators to include this information in the VoxLens visualization configuration.

For example, without the enhancements, if the user asks for the data for *Jake* when interacting with the sample visualization in Figure 8.1 using VoxLens' *Question-and-Answer* mode, the response from VoxLens would be:

```
The value for Jake is [average].
```

With my enhancements, it was:

```
The average value for Jake is [average].  The minimum score for
Jake was [minimum].  The maximum score for Jake was [maximum].
Use the data with caution.  Data has higher variation than 50%
of other names and may not be reliable.
```

Following the suggestion from the interview results, I avoided using statistical jargon in the responses from the enhanced version.

Before

```
[
  {
    "name": "Alex",
    "score": 79
  }
  ...
]
```

After

```
[
  {
    "name": "Alex",
    "score": 79,
    "vx_metadata": {
      "min": 56,
      "max": 89,
      "stdev": 13.11,
      "isAverage": true
    }
  }
  ...
]
```

Figure 8.2: Integration code sample comparing original VoxLens to the enhanced version.

8.4 Updated Configuration Options

Visualization creators use a traditional JSON file to generate data visualizations using VoxLens, specifying the independent and dependent variable values. Since this information is insufficient to compute CV, I added the functionality for visualization creators to provide more details about each data point. Specifically, they can supply the minimum, maximum, and standard deviation and specify if the dependent variable values represent an average. Figure 8.2 shows a “before” and “after” code integration example.

8.5 Summary

In this chapter, I discussed a formative study I conducted to understand the needs and preferences of screen-reader users to convey information about data uncertainty in online visualizations. The findings showed that participants were initially unfamiliar with data uncertainty. After gaining familiarity, they suggested a hybrid solution to convey this information using plain language for novice and statistical terms for expert users. I utilized these generalizable findings to develop an enhanced version of VOXLENS and presented its design, functionality, and implementation details.

Chapter 9

User Agency for Screen-Reader Users*

The previous chapters reported a need to build customizable solutions in lieu of one-size-fits-all solutions. Therefore, in this chapter, I present VOXEX[1], a system that provides screen-reader users the ability to customize the information they consume from online data visualizations. I present two formative studies with screen-reader users that informed the design of VOXEX, which comprises a backend server, a browser extension, and a configuration portal for screen-reader users to specify their preferences. I also administered a field deployment of this system by integrating it with a new version of `VoxLens` (version 4.0), introduced in this chapter. To further examine the utility of VOXEX, I offer empirical findings from two case studies with screen-reader users and visualization creators.

9.1 Motivation

Inaccessible online data visualizations exacerbate the inequity and disenfranchisement screen-reader users experience with access to digital content (Marriott et al. [2021]; Kim et al. [2021b]). Given the ubiquity of online data visualizations for conveying information to users, inaccessible visualizations can cause detrimental disruptions in the lives of screen-reader users (Sharif et al. [2021]; Choi et al. [2019]; Zong et al. [2022]; Marriott et al. [2021]; Kim et al. [2021b]). For example, inaccessible financial graphs can restrict people's ability to make critical financial decisions. Furthermore, my prior research shows that due to inaccessible

---

*Parts of this chapter are adapted from (Sharif et al. [2024d])

[1] VOXEX is distinct from VOXLENS, with a broader applicability to other modalities and systems.

visualizations, screen-reader users extract information 61% less accurately and spend 211% more time than non-screen-reader users (Sharif et al. [2021]). To make matters worse, even when accessibility measures are applied, the information provided to screen-reader users by visualization creators is often either insufficient or excessive (Sharif et al. [2021]; Zong et al. [2022]; Kim et al. [2023]). For example, alternative text ("alt-text") too often serves as the single source of accessible information with an expectation to satisfy the needs of over 2.2 billion[2] screen-reader users worldwide. Not having agency over information they consume can further aggravate the disenfranchisement they experience with digital content.

Accessibility experts have created several online tutorials and learning modules for visualization creators to familiarize themselves with digital accessibility (Eisenhuth and Chang [2022]; Onsman [2023]; Roussey [2024]). In addition, researchers and practitioners have explored avenues to enhance data visualization accessibility and devised plausible solutions, including auto-generated alternative descriptions (Sharif and Forouraghi [2018]; Lundgard and Satyanarayan [2021]), audio graphs (Sharif et al. [2022d]; Siu et al. [2022]; Holloway et al. [2022]), haptic feedback (Grabowski and Barner [1998]; Kuber et al. [2007]), 3-D printing (Brown and Hurst [2012]; Götzelmann [2016]), and multi-modality (Thompson et al. [2023]; Sharif et al. [2022f]; Blanco et al. [2022]; Srinivasan et al. [2023]). Further, researchers have examined the utility of customization in their solutions to provide screen-reader users agency over the information they consume (Thompson et al. [2023]; Kulkarni et al. [2016]; Jayant et al. [2011]). However, despite these efforts, how to best provide screen-reader users more agency over their own manner of information consumption from data visualizations, such that their agency isn't limited to a particular software, remains unclear. This work is the first to investigate and improve screen-reader users' agency in consuming information from online data visualizations through both empirical and artifact contributions.

Specifically, I developed VoxEx, a system that provides agency to screen-reader users to customize the information they consume from online data visualizations. To make informed design decisions while creating VoxEx, I conducted two formative studies: (1) a survey of 60 screen-reader users, and (2) a semi-structured interview with 12 screen-reader users. Through these studies, I collected participants' customization needs and preferences from various modalities, including alt-text, data tables, and sonification. I also gathered their thoughts on potential technology interventions. Based on the findings, I created three signifi-

---

[2] https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

cant system components for VoxEx: (1) a backend server, (2) a browser extension, and (3) a configuration portal. The portal enables screen-reader users to provide their preferences, and the extension makes these preferences available to visualization creators. Visualization creators can utilize these preferences to cater their content to the individual needs of screen-reader users using logical if-else conditions. Additionally, I administered a field deployment of VoxEx through integration with `VoxLens`.

To examine the utility of VoxEx, I conducted two case studies: (1) a five-day diary study with three screen-reader users, and (2) a single-session study with three visualization creators. The findings show that VoxEx improves the participants' efficiency in extracting information from online data visualizations and requires minimal development effort from visualization creators. The results also shed light on the importance of user agency for screen-reader users and identified areas of improvement VoxEx, including providing definitions for configuration options and modifying verbosity level by domains of interest. I enhanced VoxEx by implementing solutions for these areas of improvement.

9.2 Formative Research

To better understand opportunities for enhancing user agency for screen-reader users consuming information from online data visualizations, I administered two formative studies: (1) a survey of 60 screen-reader users, and (2) semi-structured interviews with 12 screen-reader users. I conducted the interviews to solicit a more in-depth understanding of the survey results. Altogether, the results from these two studies provide complementary insights. I analyzed the data using a mixed-methods approach involving qualitative analysis and descriptive statistics.

9.2.1 Participants

I recruited participants for both formative studies through collaboration with the National Federation of the Blind (National Federation of the Blind [1940]) and distribution lists for screen-reader users.

I surveyed 60 screen-reader users (M=49.1 years old, SD=14.9). Among the respondents, 38 identified as women, 20 as men, and two as non-binary. Fifty-three had complete blindness and seven were partially blind. The highest level of education was a doctoral degree for four participants, a master's degree for 21, an undergraduate degree for 21, an associate's degree for five, and a high school diploma for nine participants.

Table 9.1: Gender, age, vision level, and diagnosis of the participants in the interview study. Under the "G" (*Gender*) column, M = *Man*, W = *Woman*, NB = *Non-binary*, and "-" means preferred not to disclose.

	G	**Age**	**Vision-Loss Level**	**Diagnosis**
P1	W	41	Complete blindness	Functional Neurological Disorder
P2	M	76	Complete blindness	Retinitis Pigmentosa
P3	M	55	Complete blindness	Leber Congenital Amaurosis
P4	W	50	Complete blindness	Retinitis Pigmentosa
P5	W	36	Complete blindness	Retinopathy of Prematurity
P6	W	24	Complete blindness	Astrocytoma
P7	NB	28	Complete blindness	Leber Congenital Amaurosis
P8	W	65	Partial blindness	Retinitis Pigmentosa
P9	W	72	Partial blindness	Retinitis Pigmentosa
P10	W	42	Complete blindness	Glaucoma
P11	-	57	Complete blindness	Retinopathy of Prematurity
P12	M	32	Complete blindness	Retinal Detachment

Participation in the survey was voluntary and without financial compensation.

I also conducted semi-structured interviews with 12 screen-reader users over Zoom that lasted 45 minutes (see Table 9.1). Seven interviewees identified as women, three as men, one as non-binary, and one did not disclose. Their average age was 48.2 years (SD=17.2). Ten had complete blindness and two were partially blind. Two had attained a doctoral degree, four had a master's degree, four had an associate's degree, and the remaining two had a bachelor's and a high school diploma. I compensated them with a $25 Amazon gift card. I ceased recruitment after reaching saturation of insights (Doan et al. [2023]; Guest et al. [2006]).

9.2.2 Procedure

The online survey included two steps. Participants filled out each step without supervision. In the first step, I displayed the purpose of the study, eligibility criteria, and data anonymity clause. Additionally, I collected participants' demographic information, including their gender (Spiel et al. [2019]), preferred pronouns, age, preferred screen reader, vision level, diagnosis, and education level. Next, I asked what they would customize in alt-text and sonification of data visualizations and tables. I also asked the participants to provide details of their choices via a use case and examples as free-form responses. Additionally, I inquired about any optional comments they have pertaining to user agency.

I conducted the semi-structured interviews on Zoom, using its built-in features to record and transcribe the 45-minute study sessions. I also took detailed notes during the session. During the interview sessions, I gathered participants' thoughts about user agency in extracting information from online data visualizations. In particular, I asked them about *what* they would customize in different modalities, including alt-text, sonification, and data tables, and *how* they would do so (*i.e.*, possible technology interventions). I provided them with examples of alt-text and sonified responses to increase their engagement in the discussion.

9.2.3 Analysis

I calculated descriptive statistics and qualitatively analyzed the participants' responses from both studies. Specifically, I conducted a theoretical thematic analysis (Boyatzis [1998]). I used a semantic approach (Patton [1990]) and an essentialist paradigm (Potter and Wetherell [1987]; Widdicombe and Wooffitt [1995]) for the analysis, following guidelines from Braun and Clarke (Braun and Clarke [2006]). I calculated the inter-rater reliability using pairwise percentage agreement, reaching a high agreement percentage of 97%.

9.2.4 Results

I present the findings from my qualitative analysis of the participants' responses from both formative studies. Specifically, I discuss screen-reader users' customization needs and preferences with various modalities and the technology interventions identified by the participants.

Overall, many of the participants expressed the desire to customize alt-text (75%), data tables (73%), and sonification (48%). Additionally, 8% of the participants identified the necessity of having agency over information extraction using verbal question-and-answer modalities, such as provided by `VoxLens`. I present the findings for each modality in turn below.

9.2.4.1 Alt-Text

The participants communicated their desire to customize the summary content of data visualizations. Some participants were interested in getting a detailed summary, whereas others preferred succinct responses. To cater to the varying needs of users, P1 suggested providing more than the standard options in a summary. She said, "*For a graph I'm particularly interested in, I wanna know more than just the average, the maximum, the minimum. I would want that option [to know more].*"

On the other hand, P4 highlighted the importance of removing excessive information she didn't want to know from the summary. She stated, "*For the description, the problem sometimes that I run into is there's excessive alt-text. I wanna be able to have the flexibility of invoking that because I may not want to know.*"

Additionally, P8 emphasized the need to control the verbosity of the summary. She shared, "*Maybe like 'highest country, blah,' 'lowest thing, blah,' instead of full sentences; like, 'the highest country in the graph is blah.' I want that sentence structure to be shorter.*"

9.2.4.2 Data Tables

A common desire of all participants who stated a need for customizing data tables was a feature to sort data based on their preferences. This finding aligns with those from prior work (Thompson et al. [2023]; Sharif et al. [2022c]) on data visualization accessibility, particularly sonification. S21 shared his interaction and discussed the benefits of customizing sorting: "*I've encountered charts and graphs where the screen reader moves across each row first and then to the next row. This makes it difficult to navigate because sometimes it is hard to remember or keep track of column headings. It happens a lot with financial data. So, being able to customize sorting would be great.*"

Similarly, S45 said, "*I think as a data user I would like to interpret the data based on my own needs. Let's just say I wanted to look at the statistics of a baseball standings chart and wanted to see the rankings based on a secondary statistic instead of just a win-loss percentage. I think it would be very useful to be able to sort all data fields.*"

9.2.4.3 Sonification

Participants emphasized the need to customize the pitch, speed, and sound type of a sonified response to interpret data better. S22 expressed her need to customize the speed of the sound, saying, "*I would like to customize the speed of the sound or the length of the sound. Sometimes audio information moves very quickly, and it would be nice to be able to slow it down and get more details.*"

Further, S33 discussed the need to customize the pitch, frequency, and sound type, finding it critical for people with multiple disabilities. She said, "*Pitch, frequency, and sound type can be important to customize for people with intersectional disabilities that affect sensory processing so as not to trigger any discomfort, as well as to maximize clear understanding for those with any hearing impairments. It is important that any sounds do not interfere with the screen readers.*"

9.2.4.4 Technology Intervention

The participants shared positive views about having a system that enables them to specify their preferences that visualization creators could utilize to provide a customized response. P2 shared the benefits of a centralized system: "*I'm notorious for having, like, three, four, five different machines with different setups and doing the same thing on different machines to install something. And then all of a sudden, something needs to be updated, and all hell breaks loose, and then I have to change it everywhere.*"

P7 discussed the trustworthiness of a system that utilizes their preferences to cater to their experiences of visualizations. They said, "*I would keep my settings the way I wanted, and I would be certain, I would trust that no matter what graph I encountered, it at least attempted to give me the information I needed.*"

Furthermore, P5 considered browser extensions a potential intervention and shared her opinion of an authentication feature. She said, "*You could [develop a] Google extension; sometimes you install extensions and as soon as you launch them, it's ready to go. But for some extensions, you launch them, and then it asks you to sign in. So, I'd rather have it sign in because that way you know your data is somewhat protected.*"

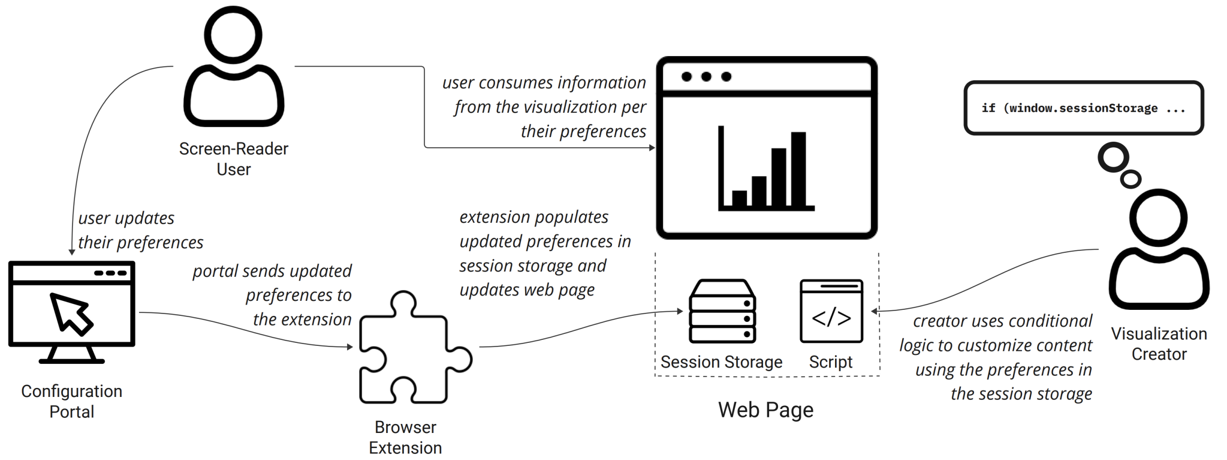


Figure 9.1: The system diagram of VoxEx, showing the interaction flow of a screen-reader user from specifying their preferences to obtaining customized information from an online data visualization. The implementation flow for creators is also shown.

9.3 Study Design

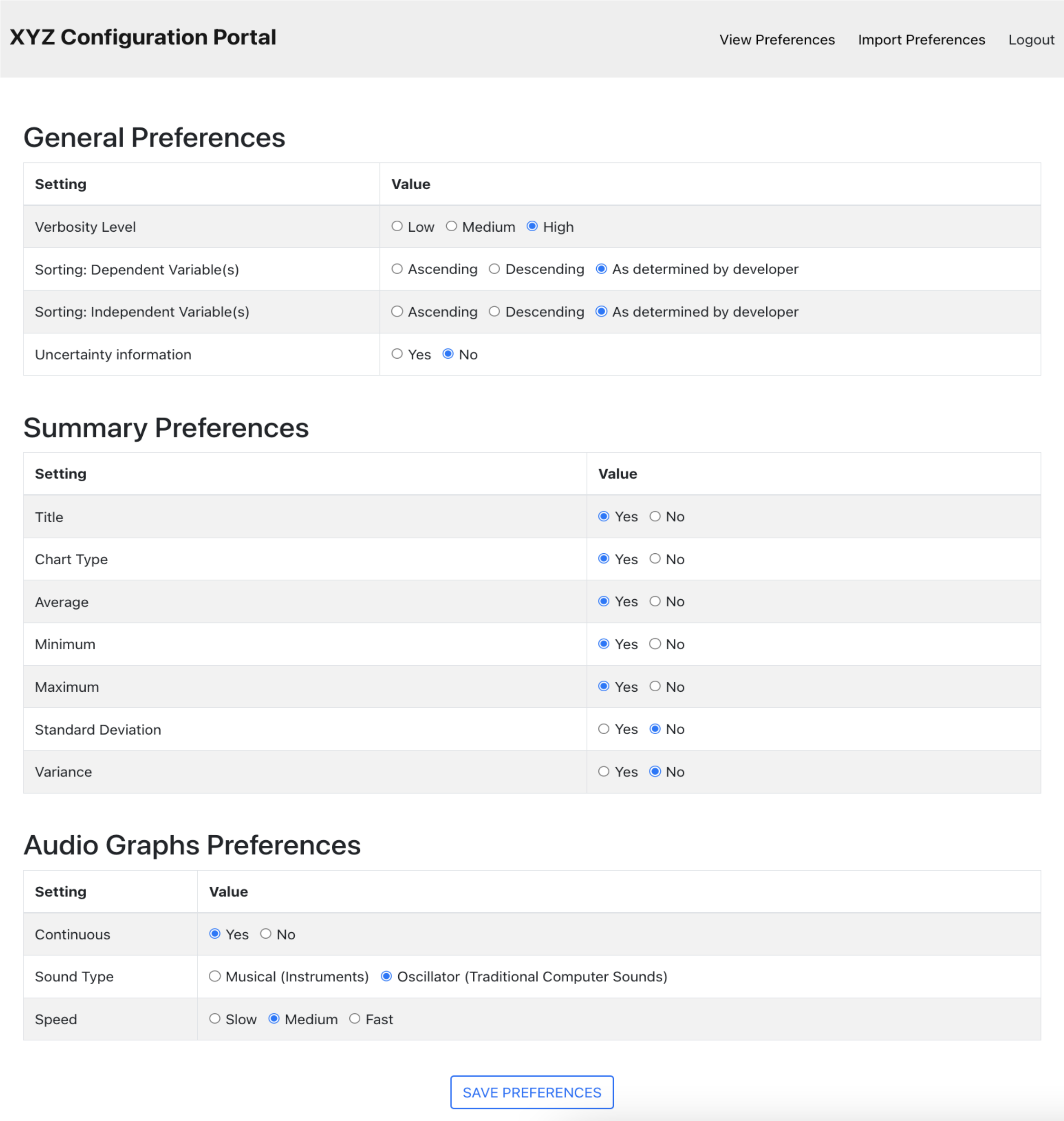
I present the design and implementation of VoxEx, a system comprising a configuration portal, a browser extension, and a backend server. I created VoxEx based on the findings from the formative studies. Specifically, the goal was to enable screen-reader users to specify their information consumption preferences and provide visualization creators with an interface to utilize these preferences to customize the accessibility of their data visualizations. By enabling screen-reader users to receive data visualizations customized to their preferences, VoxEx embodies the principles of ability-based design (Wobbrock et al. [2011b, 2018]).

9.3.1 Functionality Overview

VoxEx is a system that assists screen-reader users and visualization creators. For example, a screen-reader user can add the Chrome extension to their browser through the Chrome Web Store[3] and specify their information consumption preferences by visiting the online portal. For example, a user could modify their preferences for verbosity level, curate summary (alt-text) information, or choose the sound and speed for audio graphs. These preferences are not limited to output modality and style, like verbosity. Users can also specify the content they prefer to hear as well. For example, a user can indicate they wish to include

---

[3] https://chromewebstore.google.com/



XYZ Configuration Portal

General Preferences

Setting	Value
Verbosity Level	○ Low ○ Medium ● High
Sorting: Dependent Variable(s)	○ Ascending ○ Descending ● As determined by developer
Sorting: Independent Variable(s)	○ Ascending ○ Descending ● As determined by developer
Uncertainty information	○ Yes ● No

Summary Preferences

Setting	Value
Title	● Yes ○ No
Chart Type	● Yes ○ No
Average	● Yes ○ No
Minimum	● Yes ○ No
Maximum	● Yes ○ No
Standard Deviation	○ Yes ● No
Variance	○ Yes ● No

Audio Graphs Preferences

Setting	Value
Continuous	● Yes ○ No
Sound Type	○ Musical (Instruments) ● Oscillator (Traditional Computer Sounds)
Speed	○ Slow ● Medium ○ Fast

Figure 9.2: Screenshot of the configuration page of the VoxEx portal, displaying preferences categorized by modalities.

standard deviations. These preferences are auto-populated in session storage[4] of all the existing windows in the user's browser as well as new windows or tabs. Visualization creators can access a web page's session storage and utilize these preferences to implement conditional logic in determining the appropriate way to convey information. For example, if the user specified to include standard deviation alongside means in descriptive statistics, the visualization creators can use if-else conditions to include this information in the alt-text of their visualizations. The visualization creators do not need to install the browser extension or visit the configuration portal.

9.3.2 System Components

I discuss the design and implementation of VOXEX's three main components: (1) configuration portal, (2) browser extension, and (3) backend server (see Figure 9.1).

9.3.2.1 Configuration Portal

I created the configuration portal using `Apollo`[5] and `ReactJS`[6] and hosted it on the server. The configuration portal comprises three subcomponents: (1) Google authentication, (2) configuration page, and (3) import preferences page. To store users' preferences, I require users to log in using their Google account to access the configuration portal and use their email as the primary key in the database. I implemented the log-in feature using the `react-oauth`[7] library.

I display the configuration page once the user has successfully logged in, which shows options for their information consumption preferences in a tabular format organized by different modalities. For example, under "Summary Preferences," users have the options of "Yes" and "No" for various information types, including title, chart type, standard deviation, and variance (see Figure 9.2). I chose these options based on the findings from the formative studies. After selecting their preferences, users can click on the "Save Preferences" button, which updates their preferences in the database and dispatches these preferences to the browser extension via a `MessageEvent`.[8]

---

[4] A separate storage area for each website, available for the duration of the page session.

[5] https://www.apollographql.com/

[6] https://react.dev/

[7] https://www.npmjs.com/package/@react-oauth/google

[8] Events referencing a named message dispatched by a source object.

In addition to the configuration page, I added the functionality for screen-reader users to import their preferences directly from their screen readers. This feature is helpful to save time and reduce burden on users to re-enter their preferences in the portal. Currently, this feature is limited to VoiceOver (Inc. [2005]), and work is underway to extend this feature to JAWS (Scientific [1995]) and NVDA (Access [2006]).

I followed the WCAG 2.1 accessibility guidelines (Caldwell et al. [2018]) in developing the portal. Specifically, I used Color Vision Deficiency (CVD) friendly palettes and checked the contrast ratio using the WebAIM Contrast Checker tool (Web Accessibility In Mind [1999]), ensuring it to be at least 3:1.

9.3.2.2 Browser Extension

I created a VOXEX browser extension and published it on the Chrome Web Store, following the Chrome extension development guidelines and complying with standard privacy and security policies. The browser extension receives the preferences sent by the configuration portal via a `MessageEvent`, verifies the sender's source, and updates the session storage of each tab with these preferences. In other words, the extension enables anonymous sharing of screen-reader users' preferences to visualization creators without requiring additional hardware or software, serving as the middle-tier connection between the two user groups.

9.3.2.3 Backend Server

The backend server receives users' information and preferences and stores this data in the `Postgres`[9] database. I restricted access to the backend server by implementing strict cross-origin resource sharing (CORS) policies to only allow requests from the configuration portal host. I also used the `googleauth-library`[10] to verify the authorization token from the request headers as an additional security measure. I created the backend using `Node.js`[11] and `GraphQL`[12]. Specifically, I implemented a "getPreference" query and a "setPreferences" mutation to fetch and update data, respectively.[13]

---

[9] https://www.postgresql.org/

[10] https://www.npmjs.com/package/google-auth-library

[11] https://nodejs.org/en

[12] https://graphql.org/

[13] In GraphQL, a "query" is a read-only operation, and a "mutation" represents a write operation.

9.3.3 Enhancements to `VoxLens`

I used `VoxLens` (see Chapter 4) to explore VoxEx's utility and usability through a field deployment. I chose `VoxLens` because it is (1) open-source, (2) multi-modal, and (3) suitable for online data visualizations. I emphasize that VoxEx is an independent and generalizable solution, having no dependencies on `VoxLens`. Additionally, `VoxLens` caters to individual needs of users by supporting holistic and drilled-down information extraction through three modalities: (1) Question-and-Answer mode (verbal queries), (2) Summary mode (alt-text), and (3) Sonification (audio graphs). However, without the integration, it does not cater to individual *preferences*, such as verbosity levels, summary content, and sound types.

I implemented conditional if-else logic in `VoxLens` to integrate it with VoxEx. As a preliminary matter, I checked if preferences were present in the session storage. I defaulted to the original `VoxLens` response in cases when these preferences were absent. When the preferences were present, I curated the appropriate response, excluding, including, and modifying information as needed. For example, I added variance to the outputting of descriptive statistics when the user specified that preference. Given that `VoxLens` serves as a proxy for visualization creators to implement multi-modal accessibility in online data visualizations, the enhanced version of `VoxLens` automatically provides information customization and does not need additional development effort from visualization creators. I used the integration of VoxEx with `VoxLens` for a field deployment that I used in the case studies to assess the utility and usefulness of the system. I present the case studies below.

9.4 Evaluation

VoxEx is the first system to provide screen-reader users the ability to customize their information consumption from online data visualizations. Therefore, a baseline comparison was not possible. I reviewed the guidelines on system evaluation strategies (Ledo et al. [2018]; Olsen [2007]; Greenberg and Buxton [2008]; Wobbrock and Kientz [2016]) in choosing assessment methods for VoxEx. Specifically, I conducted two case studies (Shinohara and Tenenberg [2007]): (1) a diary study with three screen-reader users lasting five days, and (2) a single-session study with three visualization creators. I present the methodology and the qualitative results from the case studies below.

Table 9.2: Gender, age, vision level, and diagnosis of the screen-reader users in the case study. Under the "G" (*Gender*) column, M = *Man* and W = *Woman*.

	G	Age	Vision-Loss Level	Diagnosis
C1	W	34	Complete blindness	Glaucoma & Retinopathy of Prematurity
C2	M	59	Partial Vision	Retinitis Pigmentosa
C3	M	60	Complete blindness	Cataracts & Glaucoma

Table 9.3: Gender, age, domain, and visualization implementation frequency of the visualization-creator participants in the case study. Under the "G" (*Gender*) column, W = *Woman*.

	G	Age	Domain	Visualization Implementation Frequency
V1	W	29	Developer	2–3 times per week
V2	W	38	Researcher	> 5 times per week
V3	W	29	Researcher	> 5 times per week

9.4.1 Participants

I recruited three screen-reader users and three visualization creators through social media platforms and distribution lists for screen-reader users (see Tables 9.2 and 9.3). Among the screen-reader participants, two were men and one was a woman. Their average age was 51.0 years old (SD=14.7). Two were completely blind, and one had partial vision. One had attained a bachelor's degree, one had a master's degree, and the other participant had a doctoral degree. I compensated participants with a $130 Amazon gift card.

All three visualization creators identified as women. Their average age was 32.0 years (SD=5.2). None were screen-reader users. Two were researchers and one was a developer; all participants had an industry affiliation. Two had a master's degree, and one had a doctoral degree. All participants implemented data visualizations at least three times per week. I compensated them with a $30 Amazon gift card.

9.4.2 Procedure

Prior to the studies, I performed a field deployment (Siek et al. [2014]; Cherns [1976]) of VoxEx. I published the Chrome extension on the Chrome Web Store and hosted the configuration portal on the server.

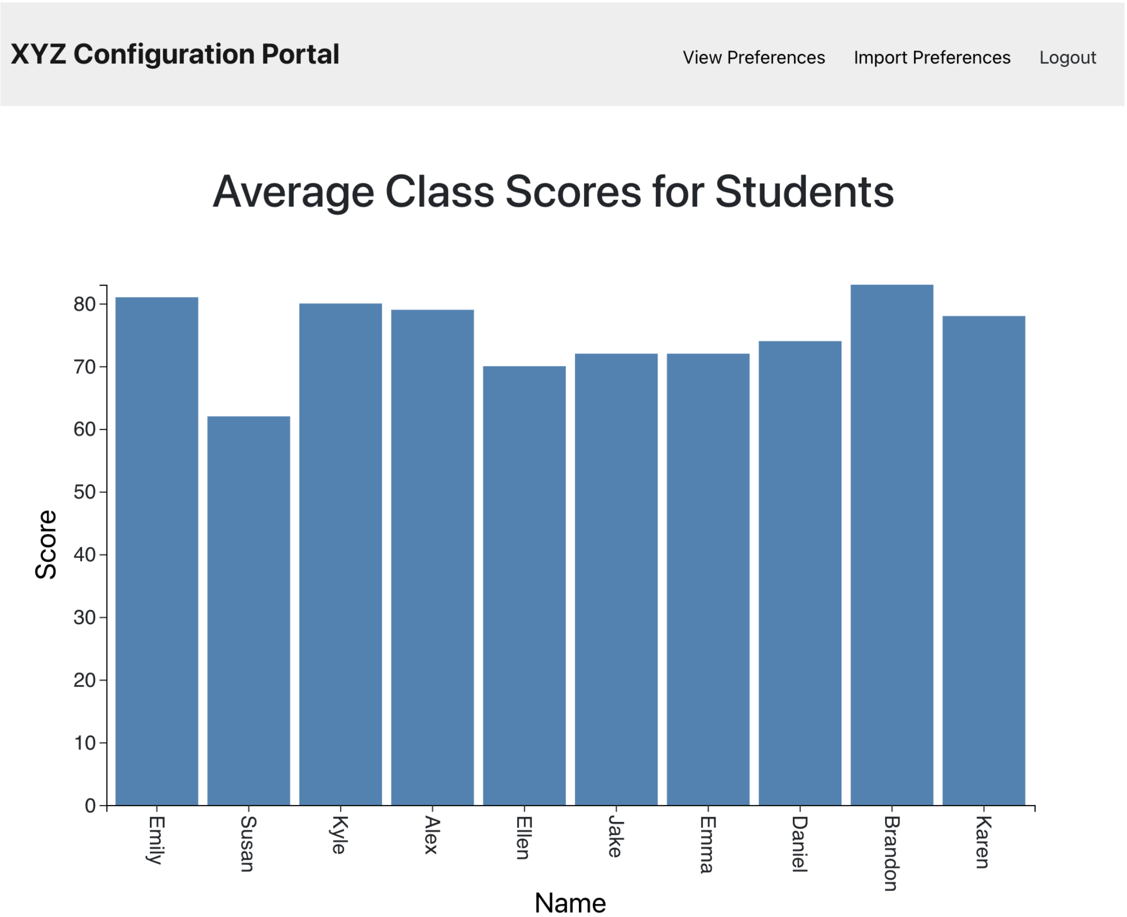


Figure 9.3: Screenshot of the example visualization page used during the case studies.

Additionally, I created an example page containing a sample visualization using D3.js (Bostock et al. [2011]) integrated with the enhanced version of `VoxLens` (see Figure 9.3).

I conducted a diary study (Ohly et al. [2010]; Bolger et al. [2003]; Reis and Gable [2000]) with the screen-reader participants that lasted five days. First, I introduced the participants to VoxEx through a tutorial session. Specifically, I asked the participants to share their screens, assisted them in installing the Chrome extension, and guided them in setting various options for "Verbosity Level" (High, Medium, and Low) to extract information from the sample visualization. For example, participants selected the setting for verbosity level to "High," navigated to the sample visualization page, and extracted information using all three `VoxLens` modalities (Q&A, Summary, and Sonification). They repeated this task using the "Medium" and "Low" settings for verbosity level to experience the difference in responses for each setting.

I finished the tutorial session once the participants expressed comfort in interacting with the system without supervision. During the next three days, participants interacted with the system without supervision and logged a diary entry each day. Their task was to interact with the sample visualization using every setting for each configuration option in the portal. I conducted a semi-structured interview on the fifth day to ask questions about their interactions with the system. I also administered the NASA-TLX perceived workload questionnaire (Hart and Staveland [1988]; Hart [2006]). I used Zoom to administer, record, and transcribe the tutorial session and the follow-up interviews.

I interviewed visualization creators via Zoom, which lasted an hour. I introduced the participants to VoxEx and shared scenarios and examples of using standard logical if-else conditions to cater to users' preferences with various modalities. For example, I demonstrated how a developer could use the system to fetch a user's preferences from the session storage and implement logical conditions to construct an alt-text for an online data visualization. I sought their perception of user agency, thoughts on the utility and usability of the system, feedback on development efforts, and ideas on further improving VoxEx. I recorded and transcribed these interviews using Zoom's built-in features.

9.4.3 Analysis

I qualitatively analyzed the interviews and the diary entries from the screen-reader users. Similar to the formative studies, I used a semantic approach (Patton [1990]) and an essentialist paradigm (Potter and Wetherell [1987]; Widdicombe and Wooffitt [1995]) to conduct a theoretical thematic analysis (Boyatzis [1998]). I followed guidelines from Braun and Clarke (Braun and Clarke [2006]). Additionally, I analyzed the subjective ratings collected from the NASA-TLX questionnaire (Hart and Staveland [1988]; Hart [2006]). I present the findings separately in the section below.

9.5 Results

I present the results from the two case studies with three screen-reader users and three visualization creators. Additionally, I describe further enhancements to VoxEx based on participants' feedback from the studies.

9.5.1 Case Study with Screen-Reader Users

I conducted a five-day diary study with three screen-reader users, which involved a tutorial session, unsupervised interaction with VoxEx, and a follow-up interview. I studied participants' interactions and privacy and security concerns, asked them about areas for improvement, and sought their overall feedback. Additionally, I collected their subjective ratings using the NASA-TLX questionnaire. I discuss these below.

9.5.1.1 Information Extraction Experience

A prevalent theme in the interviews was participants recognizing how VoxEx improved their efficiency in extracting information from online data visualizations, providing them the autonomy to control and choose the information they wanted to extract. C3 and C1 expressed their enthusiasm this way:

> It allows me to do things quickly and efficiently, get the information, glance at it like a sighted person. I was just seeing how well and quickly it responded. And then when I started to define what I wanted, I felt even better. It's a useful way of describing information in the format that I need. It's about damn time. (C3)

C1 said:

> I like that I can get the information I want quicker. And it manipulates the data better than the table would. I could get largest, smallest, or change the order and do things that a table reading with a screen reader does not do. And it's fairly easy. Setting up the stuff was not a problem. Definitely, this is something that I would use. (C1)

Similarly, C2 discussed the importance of autonomy and agency over information she wanted:

> Being able to choose the information I got felt really important and meaningful because I had more autonomy and

> control over the information that I wanted; it would save time not having to listen to information that I didn't need or want or care about. As a blind person, it takes about like 50, maybe even a hundred percent longer than a sighted person to do anything. So any way I can make things more efficient or kind of control how much time something takes is great. (C2)

In particular, C2 highlighted the benefits of VOXEX with sonification:

> Being able to control the type of sound and whether it was continuous or not and then how fast the sound was being presented was really cool. Again, going back to autonomy and choice and being able to pick preferences that worked for me and my learning style was really helpful. (C2)

9.5.1.2 Privacy and Security Concerns

During the follow-up interviews, I asked the participants if they had any privacy or security concerns with installing and using VOXEX. The participants did not report any concerns, regarding the research study as trustworthy. For example, C1 and C2 shared their thoughts:

> I'm not concerned about privacy on this thing simply because I figured you wouldn't give me anything with privacy issues. (C1)

> I think having that call with you and being told more about it was helpful rather than just finding it randomly on the web. (C2)

Naturally, I inquired about participants' concerns if they had encountered VOXEX outside of a research study and what factors might contribute toward resolving any concerns. C1 and C2 stated that having a credible source resolves their concerns, whereas C3 relied on and trusted his security setup:

> I’m very leery about what I put onto my computer. I would have looked if anybody else reviewed it. Like, is this from a credible source? I would look into if there were any articles on the thing, and see if anybody did any research on that. I would trust reputable articles, basically. Which is the way I do everything. (C1)

C2 agreed, saying,

> So really, that’s just kind of my standard practice when I download anything is to make sure it’s [from] a reliable source based on my research and then ask people that I trust who understand these things a lot more than I do. (C2)

And C3 added:

> I have pretty good online security and practices and I also have a protected VPN and everything, so I don’t have any real issues with extensions. (C3)

9.5.1.3 Areas of Improvement

I inquired about areas of improvement for VOXEX from the screen-reader participants. All three participants highlighted the need to provide definitions of the configuration options. In addition to definitions, C2 specified mathematical and scientific concepts as particular use cases for definitions:

> Mathematical and scientific concepts I’m rusty [with], so [I] don’t have a solid understanding of standard deviation and variance and stuff like that. Having the option to get more information about standard deviation in a graph takes the burden off. (C2)

C1 identified the benefits of having a help page:

> `I might not have understood some stuff, so either a little definition or a help page for people who may not know what a maximum [or] minimum is.  I think a decent help page would not be a bad idea.` (C1)

C3 recognized the opportunity for VoxEx to be an educational tool:

> `Describing standard deviation for someone who doesn't know could be a good thing.  I think if a person hasn't had statistics in school, just a quick definition of what those are would help them a lot.  It could be an educational tool as well.` (C3)

9.5.1.4 Overall Comments and Feedback

The participants provided additional feedback on the usefulness of VoxEx by highlighting the import preferences feature. For example, C2 shared:

> `I think it's really helpful in time-saving because if I don't have that ability, then I'm using a lot of my time and mental load on getting the configurations exactly as I want them for general use.  I think it's a very good call on having settings be able to be imported and stuff.` (C2)

Furthermore, the participants discussed how VoxEx would be helpful for non-screen-reader users, including people with non-visual disabilities and those without disabilities. C3 discussed its benefits for people with learning disabilities:

> `Not only would I use this with people that have visual or cognitive disabilities, I'd [also] introduce it to people that have learning disabilities but show them how you can use them in conjunction with other skill sets, and it would enhance it`

> `and also increase the ability to process information, also increase the speed.` (C3)

C2 conveyed VOXEX's advantages for non-screen-reader users:

> `I could see it with anybody with some print-related disability. And people who don't have understanding of graphs and a strong background in statistics. I think this would help bridge those gaps.` (C2)

9.5.1.5 Subjective Workload Assessment

I used the NASA Task Load Index questionnaire (NASA-TLX) (Hart and Staveland [1988]; Hart [2006]) to collect subjective workload ratings from the screen-reader participants. The NASA-TLX questionnaire uses six 21-point scales to determine users' perceived task workload: *mental demand, physical demand, temporal demand, effort, frustration,* and *performance*. To characterize VOXEX's workload, I relied on a characterization from my prior work, since a control condition for comparison does not really exist.

The results indicated that VOXEX requires low mental demand (M=3.0, SD=1.0), low physical demand (M=1.3, SD=.6), low temporal demand (M=5.0, SD=3.0), low effort (M=5.3, SD=2.1), low frustration (M=3.7, SD=2.9), and has high perceived performance (M=17.3, SD=6.4). The subjective ratings can serve as a control for comparisons for future work.

9.5.2 Case Study with Visualization Creators

I interviewed three visualization creators in a semi-structured manner. I introduced them to VOXEX and sought their feedback on the utility of VOXEX and development effort. Additionally, I inquired about areas of improvement and overall comments. I discuss these in turn below.

9.5.2.1 Perceived Utility of VOXEX

I introduced the participants to VOXEX and demonstrated an example use case to aid them in forming an understanding of the system's features. Subsequently, I inquired about their perceived utility of VOXEX.

The participants showed enthusiasm for the system, finding it useful for all users, particularly screen-reader users. V2 classified VOXEX as “remarkable” and surmised that it provides equal access to data visualizations as noted in Section 508 of the Rehabilitation Act (Commission [1998]; Olalere and Lazar [2011]):

> `I think it’s really fantastic. I mean, thinking of laws in the United States, like Section 508 and providing comparable equal access, this [tool] does that. I can’t stress enough this is truly remarkable, to be honest. I think it’s incredibly useful. I do think this is huge in so many ways and many leaps in the right direction toward agency when it comes to screen reader users and data visualization.` (V2)

Similarly, V1 considered VOXEX a win-win situation for screen-reader users and developers:

> `As a designer and developer, I think it alleviates some of the decision-making you need to put in, and at the same time, the user gets what they want. It’s a win-win situation. I mean, I like the idea of customization. With these preferences, it’s a bit more niche.` (V1)

V3 shared her experiences as a developer about the system’s usefulness in catering the content to the preferences of target users:

> `Every time I develop something, I think of my target users first. It’s pretty useful because you get to know them more, what their preferences are, and how you can address those preferences in your design. Because without this [tool], you actually don’t know what your user’s preference is. You can only give them all of the information but then people feel impatient that you’re reading a long paragraph that they are not even interested in. They can get the information straight`

> `ahead.  Like a short answer, you have a setting to control that and the speed and stuff.` (V3)

9.5.2.2 Development Effort

I asked participants their thoughts on the development effort needed to provide customized information from online data visualizations using VOXEX. Specifically, I asked them to rate their perceived difficulty of implementation on a Likert scale of 1 to 7, with "1" representing none-to-minimal difficulty and "7" representing extreme difficulty. All the participants rated the difficulty as "1". V1 elaborated on her choice:

> `It doesn't seem like that much work.  Like, so you know that the person doesn't want a title, for example.  So you then decide, okay, the user doesn't want the title, then I'm not going to include the title in it.  It's pretty simple really.` (V1)

9.5.2.3 Areas of Improvement

Similar to the screen-reader participants, visualization creators also emphasized the need to provide definitions of the configuration options. For example, V2 expressed her opinions as:

> `If I have no idea what a standard deviation is, and especially [what] uncertainty [of] information [is], having a definition here to make this even more accessible would be fantastic.` (V2)

Additionally, the participants noted that domain interest could be a potential factor in determining verbosity levels. V1 suggested to adjust the content based on the domains of interest:

> `You might not have the same preferences if you're going through something you're invested in versus something you're not super invested in.  So, if you asked users for their interests or the domains they're interested in and adjusted the content based on`

> that, then I think that would probably increase the utility. (V1)

V2 shared an example use case of adjusting verbosity:

> Let's say I was trying to learn information about COVID. I think I would want to have higher verbosity levels for visualizations involving health and a lower level for visualizations involving, say, sports. (V2)

9.5.2.4 Overall Comments and Feedback

The participants provided positive and encouraging feedback on VoxEx. In particular, they were excited about the system's central focus on user agency for screen-reader users. For example, V1 conveyed her enthusiasm for the importance of user agency:

> It's pretty important and critical to have agency over the information you consume. Giving users the option to specify what information they want and what information they don't want, and to control that is pretty important. (V1)

V2 agreed, saying:

> I think it's super important. Whether I do a good job in general as visual presenters is the question. Because it could frankly disenfranchise people sometimes from even attending higher education institutions to be able to then participate in a way that feels equal and comparable. (V2)

9.5.3 VoxEx Enhancements

I made enhancements to VoxEx to incorporate the feedback I received from the participants during the case studies. In particular, I focused on the areas of improvement they identified and addressed their concerns by: (1) adding definitions of configuration options, and (2) using domains of interest to determine verbosity levels. I present these enhancements below.

9.5.3.1 Definition of Configuration Options

A unanimous suggestion from the participants was to add a definition for each configuration option to support users in understanding these terms. Therefore, I added an accessible help tooltip icon next to each configuration option, enabling users to obtain the definition of the term in plain language.

9.5.3.2 Domains of Interest

The visualization-creator participants noted the need for an additional feature to determine verbosity levels based on domains of interest. I integrated the `ChatGPT` API[14] to implement this feature in VOXEX as Large-Language Models are a beneficial source of extracting categories from web content (Wei et al. [2023]; Vörös et al. [2023]; Kocoń et al. [2023]). First, I issued `ChatGPT` the query "website categories" to obtain the list of broader website categories. I included this list on the configuration portal and provided users the option to select multiple categories of their interest.

Then, on the client side, I queried `ChatGPT` with the text from the web page's meta tags, title, and body to determine its category. For example, on the sample visualization page, I queried "website category for 'Data visualization showing average class scores for each student'" and received the following response: "The appropriate website category for a 'Data visualization showing average class scores for each student' would likely fall under 'Education' or 'Academic Resources.'" If the identified category of the web page was among the list of user's categories of interest, I set the user's verbosity setting to "High."

More work is underway, beyond what is presented in this dissertation, to improve the performance of VOXEX by minimizing the number of network calls through caching and to find an economically sustainable solution for this feature as the `ChatGPT` API requires paying a subscription to OpenAI.

---

[14] https://chat.openai.com/

9.6 Utility Beyond Online Data Visualizations

I developed VoxEx with a central focus on online data visualizations. However, this work has utility beyond online data visualizations. For example, developers can use this system to generate summaries of information from data tables, especially those with high cardinalities, enabling screen-reader users to extract information more effectively. Furthermore, future work can extend this system to include preferences for customizing one's information consumption concerning images and graphics, allowing content creators to customize the alt-text and cater to the individual needs of screen-reader users. In addition, as the participants identified, this work can benefit non-screen-reader users, including those with non-visual disabilities. For example, this system can support people with color-vision-deficiency (CVD) by enabling them to specify preferences that visualization creators can use to automatically modify color palettes in visualizations and graphics based on their CVD type and preferences. I encourage researchers to explore the utility of this system further, including its potential benefits for visualization literacy.

9.7 Summary

In this chapter, I sought to provide enhanced agency to screen-reader users to customize the information they consume from online data visualizations. To achieve this objective, I surveyed and interviewed screen-reader users and subsequently used the findings from these formative studies to develop VoxEx. Through integration with `VoxLens`, I conducted a field deployment of my system. I conducted a diary study with screen-reader users and single-session interviews with visualization creators to assess the utility and usability of VoxEx. I reported the findings from the studies showing that VoxEx improves screen-reader users' efficiency of information extraction and requires minimal effort from visualization creators. Additionally, I incorporated the suggestions from the participants to improve VoxEx.

Chapter 10

Supporting Visualization Creators in Making Data Visualizations Accessible*

In the prior chapters of this dissertation, I focused on improving the accessibility of online data visualizations for screen-reader users. In this chapter, I set my concentration on improving visualization creators' knowledge of accessibility and supporting them in making data visualizations accessible. I present a formative study and in-depth interviews with visualization creators to identify effective interventions to enhance their experiences with data visualization accessibility. Further, I demonstrate the generalizability of these findings by implementing and assessing each of the five identified interventions, consequently introducing another enhanced version of VoxLens (version 5.0).

10.1 Motivation

Lack of access to the underlying information in online data visualizations can be an inequity issue and severely disenfranchise screen-reader users (Marriott et al. [2021]; Lee et al. [2021]; Sharif et al. [2021]; Davis [2002]), of which there are over 7.6 million in the United States, and who rely on visualization creators to make online data visualizations accessible. Empirical findings show that even the most commonly used technique in data visualization accessibility, alternative text ("ALT-text"), is often missing or insufficient, causing screen-reader users to extract information 61% less accurately and spend 211% more time than

---

*Parts of this chapter are adapted from (Sharif et al. [2024b])

non-screen-reader users (Sharif et al. [2021]; Zong et al. [2022]). Recently, researchers have provided increased accessibility awareness for visualization creators, especially after inaccessible COVID-19 graphs resulted in health concerns for the blind and low-vision community (Fan et al. [2023]; Praharaj et al. [2023]). Consequently, the visualization community has undertaken steps to address these accessibility concerns. For example, *The New York Times* recently created a new job position of an *Accessibility Visuals Editor*, which had never before existed (Cassidy [2023]; Duenes [2022]).

However, despite these efforts, accessibility in data visualizations is often non-existent and significantly varies based on visualization creators' expertise and subject-matter familiarity. Consequently, this inaccessibility results in an inconsistent and cumbersome interaction experience for screen-reader users with online visualizations. Although researchers have proposed solutions to make data visualizations accessible, such as auto-generated ALT-text (Sharif and Forouraghi [2018]; Lundgard and Satyanarayan [2021]), sonification (Austin and Sorge [2023]; Roy and Boppana [2022]), 3-D printing (Brown and Hurst [2012]; Hurst and Kane [2013]), data tables (Developers [2014]), and multi-modality (Sharif et al. [2022f]; Thompson et al. [2023]), investigating and *minimizing* the challenges for visualization creators when making online data visualizations accessible remains unexplored. While Joyner *et al.* (Joyner et al. [2022]) shed preliminary light on developers' challenges with accessibility, this work is the first to reduce those challenges by identifying interventions to: (1) *improve* developers' understanding of the experiences of screen-reader users, (2) *enhance* developers' knowledge of data visualization accessibility, and (3) *assist* developers in making data visualizations accessible to screen-reader users.

To this end, I surveyed 57 visualization creators to understand their perceived importance, knowledge, and prioritization of data visualization accessibility. The findings indicate that their perceived importance of visualization accessibility impacts their knowledge of it, and together, these factors significantly affect their prioritization of it, which influences the challenges they encounter when making visualizations accessible. I also inquired about creators' challenges with accessibility and tools that could reduce those challenges. The analysis revealed five interventions that, when used in combination with each other, elevate creators' understanding of screen-reader users' interactions with visualizations, strengthen their knowledge of data visualization accessibility, and aid them in making data visualizations accessible. These five interventions include one educational intervention (*Workshops*) and four technological interventions (*Emulators*, *Evalua-*

tors, *Feedback Collectors*, and *Multi-Modal Automated Tools*).

To gather further insights into the features that make these interventions effective, I conducted semi-structured interviews with 12 visualization creators, inquiring about their perception of each identified intervention. Specifically, I asked about the distinctive features that would enhance each intervention's effectiveness to improve: (1) their knowledge of accessibility and screen-reader users' experiences, and (2) assist them in creating accessible visualizations. The findings identify the nuanced features, perceived challenges, specifications, and recommendations from visualization creators for effective versions of each intervention.

Then, to assess the generalizability of these findings and the effectiveness of these interventions, I performed a task-based user study with 10 visualization creators. Specifically, I implemented each of the four technological interventions, integrating them into VOXLENS. I compared this enhanced version's performance with the original version of VOXLENS. These enhancements increased participants' understanding of screen-reader users' challenges with data visualizations, knowledge of data visualization accessibility, and perceived usefulness of VOXLENS by 43.8%, 16.7%, and 11.5%, respectively. Additionally, I organized the educational intervention, *Workshop*, finding that it improved participants' knowledge, prioritization, and perceived importance of accessibility by 38%, 4%, and 15%, respectively. Furthermore, participants decreased their rating of challenges they face with accessibility by 17% and increased their desire to conduct studies with screen-reader users by 157%.

10.2 Formative Research

To investigate visualization creators' challenges with data visualization accessibility, identify the factors inducing these challenges, and determine solutions to minimize these challenges, I surveyed 57 visualization creators. I used a mixed-methods approach. I present the details below.

10.2.1 Participants

I surveyed 57 visualization creators (M=35.3 years old, SD=8.4). I advertised the survey through social media platforms and various email distribution lists for visualization creators. Among the respondents, 28 identified as women, 24 as men, four as non-binary, and one did not disclose their gender identity. Thirty-one participants were affiliated with industry, 20 with academia, and six with both. Twenty-eight participants specified themselves as developers, 19 as researchers, and 10 as both. The highest level of education was a doctoral degree for 20 participants, a master's degree for 14, an undergraduate degree for 20, an associate's degree for two, and a high school diploma for one participant. For visualization accessibility knowledge, three participants identified as "beginner," seven as "moderate beginner," 10 as "advanced beginner," 10 as "competent," 19 as "proficient," six as "advanced proficient," and two as "expert."

10.2.2 Procedure

The online survey included five steps. Participants filled out each step without supervision. In the first step of the survey, I displayed the purpose of the study, eligibility criteria, and data anonymity clause. Additionally, I collected participants' demographic information, including their gender (Spiel et al. [2019]), preferred pronouns, age, education level, domain (DO; "industry," "academia," or "both"), and role (RO; "developer," "researcher," or "both"). For clarity, the term "developer" encompassed similar terms, including "programmer" and "designer." Additionally, similar to other studies presented in this dissertation, I collected gender identities to report demographic information; I did not analyze the effect of gender identities on the dependent variables as it was outside the scope of this work.

Next, in the second survey step, I asked about their current practices and knowledge (KR) about visualization accessibility. I also asked them about their accessibility prioritization (PR) and perceived importance (IR), accessibility policy enforcement (PO; "no policies," "policies recommended," or "policies enforced") at their respective organizations, and their strategies to make data visualizations accessible to screen-reader users. Additionally, I asked the participants about their frequency of testing their visualizations with screen-reader users (SF). I used a Likert scale ranging from 1 (lowest; e.g., "not important") to 7 (highest; e.g., "extremely important") for collecting KR, PR, IR, and SF.

In the third step, I inquired about the challenges they experience in incorporating accessibility in visualizations and their Likert-scale rating for the challenge level (CR). Then, in survey step four, I asked the following open-ended questions to gather insights about interventions that could mitigate participants' accessibility challenges:

1. Which resources or tools could enhance creators' understanding of screen-reader users' challenges?
2. Which resources or tools could help enhance creators' knowledge regarding accessible online data

visualizations?

3. Which resources or tools could help make online data visualizations accessible to screen-reader users?

Finally, in survey step five, I asked for any optional comments and their interest in a follow-up interview.

10.2.3 Quantitative Evaluation

To quantitatively analyze the survey responses, I investigated the factors that contribute to creators':

- Challenges with implementing accessibility in data visualizations (RQ1)
- Knowledge of data visualization accessibility (RQ2)
- Prioritization of making data visualizations accessible (RQ3)

Table 10.1: Results from the ordinal logistic regression analysis of N=57 survey responses. "IV" means the Independent Variable. "BIC" is the Bayesian Information Criterion (lower is better). "CR" = *Challenge Rating*, "PR" = *Prioritization Rating*, "DO" = *Domain*, "KR" = *Knowledge Rating*, "SF" = *Studies Frequency*, "IR" = *Importance Rating*, and "PO" = *Policies*. Results with p<.05 are statistically significant.

Model	**IV**	N	χ^2	p
$CR \sim PR + DO + age$ BIC = 190.76	PR	57	19.68	$< .05$
	DO	57	6.62	$< .05$
	age	57	3.18	.075
$KR \sim SF + IR + age$ BIC = 229.49	SF	57	22.45	$< .001$
	IR	57	19.68	$< .05$
	age	57	8.52	$< .05$
$PR \sim KR + IR + PO$ BIC = 232.67	KR	57	12.86	$< .05$
	IR	57	13.32	$< .05$
	PO	57	5.19	.075

I used stepwise ordinal logistic regression (both directions) (Draper and Smith [1981]; Efroymson [1960]; Hocking [1976]) to identify possible outcomes of the independent variables out of the following

Table 10.2: Statistical results from the Spearman's rank correlation analysis of all the variables. "V1" means Variable 1 and "V2" means Variable 2. ρ is the Spearman's rank correlation coefficient. All results with p <.05 are statistically significant.

V1	V2	ρ	p
CR	KR	-.05	.719
CR	IR	.00	.980
CR	PO	-.05	.725
CR	SF	.09	.528
CR	PR	-.08	.561
CR	DO	.16	.247
CR	RO	-.08	.541
KR	IR	.20	.145
KR	PR	.42	< .001
KR	PO	-.39	< .05
KR	SF	.29	< .05
KR	DO	.23	.088
KR	RO	-.08	.548
IR	PR	.49	< .001
IR	PO	-.24	.068
IR	SF	.21	.119
IR	DO	.66	.659
IR	RO	.09	.499
PR	PO	-.43	< .001
PR	SF	.12	.373
PR	DO	.14	.304
PR	RO	-.06	.655
PO	SF	-.34	< .05
PO	DO	-.07	.605
PO	RO	-.03	.803
SF	DO	-.15	.257
SF	RO	-.01	.962
DO	RO	-.25	< .05

candidate variables: SF, KR, IR, PR, DO, RO, PO, and *age*. As a result, for RQ1, the dependent variable (DV) was CR, and the independent variables (IVs) were PR, DO, and *age*. For RQ2, the DV was KR, and the IVs were SF, IR, and *age*. For RQ3, the DV was PR, and the IVs were KR, IR, and PO.

Variables CR, SF, KR, IR, and PR had an ordinal representation (1 to 7 on a Likert scale), whereas DO, RO, and PO were trichotomous. I investigated the effect of the independent variables on the dependent variables using ordinal logistic regression (McCullagh [1980]; McKelvey and Zavoina [1975]), a standard technique for analyzing ordinal responses. Table 10.1 shows the statistical results.

Variables PR and DO had a significant effect on CR. Specifically, 46.7% (N=21) of the respondents who regarded implementing accessibility as at least somewhat challenging (higher than 4 on the Likert scale) rated their prioritization at higher than 4, and 60.0% (N=27) reported their affiliation with "industry." *Age* had a marginal effect on CR ($p \approx$.075).

SF, IR, and *age* had a significant main effect on KR. Specifically, 3.7% (N=1) and 88.9% (N=24) of the respondents who considered themselves at least somewhat knowledgeable in accessibility (higher than 4 on the Likert scale) rated their frequency of conducting studies with screen-reader users and accessibility importance at higher than 4. Additionally, 74.1% (N=20) of these respondents reported being between 30 and 50 years old.

KR and IR had a significant main effect on PR. Specifically, 63.3% (N=19) and 96.7% (N=29) of the respondents who at least somewhat prioritized accessibility (higher than 4 on the Likert scale) rated their accessibility knowledge and importance at higher than 4. PO had a marginal effect on PR ($p \approx$.075).

As a supplementary analysis, I performed Spearman's rank correlation analysis (Spearman [1904]) to investigate the linear relationship between all the variables. I found a positive correlation between KR and PR (p<.001), KR and PO (p<.05), KR and SF (p<.05), IR and PR (p<.001), PR and PO (p<.001), PO and SF (p<.05), and DO and RO (p<.05) (see Table 10.2). This analysis was supplementary to this work and these results do not imply causality. However, these findings provide avenues for future research to determine possible causal relationships among these variables to shed further light on visualization creators' challenges with data visualization accessibility.

10.2.4 Qualitative Evaluation

I qualitatively analyzed the participants' (N=57) open-ended responses to questions in Step 4 of the survey. I conducted a theoretical thematic analysis (Boyatzis [1998]) using a semantic approach (Patton [1990]) and an essentialist paradigm (Potter and Wetherell [1987]; Widdicombe and Wooffitt [1995]), following guidelines from Braun and Clarke (Braun and Clarke [2006]). Three co-authors coded each response. I calculated the inter-rater reliability using pairwise percentage agreement, reaching a high agreement percentage of 97%. The final analysis revealed one educational intervention (*Workshops*) and four technological interventions (*Emulators*, *Evaluators*, *Feedback Collectors*, and *Multi-Modal Automated Tools*) identified by visualization creators that would reduce their challenges in making data visualizations accessible to screen-reader users. I discuss these below. (For a qualitative analysis of visualization designers' experiences with data visualization accessibility, I direct readers to work by Joyner *et al.* (Joyner et al. [2022]).)

10.2.4.1 Workshops

For simplicity, I use the umbrella term *Workshops* to represent synonymous terms that the participants used in their responses, including "course," "training," and "tutorials." The participants expressed an interest in attending workshops to gain "hands-on" experience (P52) to understand challenges that screen-reader users encounter with online data visualizations and to enhance their knowledge of data visualization accessibility. For example, P5 shared their enthusiasm for taking part in and effectively advertising a workshop:

> `I would LOVE to take a class taught by someone who uses these features. An effective way might be to raise awareness of such courses in U.S. federal programs.` (P5)

Additionally, P41 discussed the scarcity of such workshops at her organization:

> `Workshops and tutorials that specifically address the experiences of screen-reader users are helpful but scarce at my institution.` (P41)

10.2.4.2 Emulators

In this dissertation, I define an emulator as software that mimics responses from screen readers to assist visualization creators in hearing the screen reader output without needing to install a screen reader on their computer. The participants stated their desire to understand the interaction experiences of screen-reader users with data visualizations through emulating the output from screen readers. For example, P23 considered an emulator a "middle-ground task":

> `I wish there were a middle-ground task between running a screen reader myself and hiring someone to use a screen reader for me.` (P23)

Similarly, P47 and P51 identified an emulator's benefits as "quickly previewing the experience" and "instantly hearing the visualization from a screen reader perspective," respectively. Additionally, P21 highlighted its use case as a centralized tool that would assist in self-auditing data visualizations:

> `Centralized walk-throughs of different screen reader modalities would be cool. How blind and visually impaired users use the screen reader.` (P21)

10.2.4.3 Evaluators

Unsurprisingly, the participants identified the need for tools that evaluate their data visualizations for accessibility (*Evaluators*). Several tools currently exist that evaluate data visualizations' accessibility (Elavsky et al. [2022]; Alcaraz Martinez et al. [2022]; Takagi et al. [2003]; WebAIM [2016]) based on established guidelines, such as WCAG 2.1 (Caldwell et al. [2018]). However, as the participants identified in their responses, these solutions are not integrated into the visualization libraries and do not provide real-time evaluation of visualizations during the development process. For example, P14 and P49 communicated the need for automated evaluators by providing relatable examples of existing solutions:

> `As someone who likes to create data visualizations but isn't a professional, I'd like to see more tools include accessibility`

> `assessment, like Microsoft has done with PowerPoint, Excel, etc. Or JavaScript libraries.` (P14)

> `Having a tool that checks your visualizations as you're working... Like, a Grammarly-type of resource, but to improve the accessibility of your visualization!` (P49)

Similarly, P48 shared his challenges with generating alternative text and identified a specific use case for evaluators:

> `I've been relying on alt-text and I feel that I have spent a lot of time and effort into this but I don't have a way to say, "okay, what you're doing is sufficient!" So a checker (e.g., alt-text checker) would be great!` (P48)

10.2.4.4 Feedback Collector

A common theme in the survey responses was visualization creators' desire to get feedback on the accessibility of their data visualizations from "real-life users" (P29) and "access to screen-reader users for input" (P46). (In this dissertation, I refer to this utility as *Feedback Collectors*.) For example, P54 and P16 accentuated the need to connect with screen-reader users to receive feedback:

> `I'd like to connect with more people with disabilities to test the designs.` (P54)

> `There is a lot of garbage in the information that is not actually applicable to what I'm trying to do. I would like to be able to have the ability to ask an expert in the matter.` (P36)

Similarly, P35 highlighted the usefulness of a feedback collector, emphasizing protecting users' privacy:

> `I think writing to report a bug or enhancement should be easy. A tool to enable users to report without losing privacy or sensitive information would be valuable.` (P35)

10.2.4.5 Multi-Modal Automated Tool

A *Multi-Modal Automated Tool* in data visualization accessibility is software or a plug-in that provides multiple modalities for users to extract information, such as sonification, ALT-text, and tables. A prevalent mention in the survey responses was that of an "easy to create" (P44) and "auto-implement" (P18, P46) multi-modal solution. Additionally, the visualization creators emphasized the need for such solutions to be "open-source" (P29) and "financially accessible" (P25). Notably, P53 classified an automated multi-modal solution as a "magic wand":

> `A magic wand... a charting tool where a chart is just made accessible in the background without the chart author needing to do anything special.` (P53)

P6 identified ways and specific domains where a multi-modal solution would be beneficial:

> `[It would be useful as] R/ggplot2 and python/seaborn extensions to automatically sonify data and/or generate alt-text, or a web-based library that can be built-in automatically.` (P6)

Additionally, the participants suggested using artificial intelligence (P2, P28, P42), computer vision (P52), and technologies such as ChatGPT (P26) and large-language models (P38) for creating multi-modal solutions to make online data visualizations accessible.

Overall, the findings shed light on the challenges visualization creators experience with making data visualizations accessible, identifying factors contributing to these challenges. In particular, the qualitative results determine an educational intervention and four technological interventions that might assist visualization creators in minimizing these challenges. These findings motivated the need to conduct in-depth interviews with visualization creators to delve further into the specific features that improve the effectiveness of the recognized interventions. I present the methodology and findings from the interviews with visualization creators in the section below.

Table 10.3: Gender, age, domain, role, and education level of the interview participants.

	Gender	Age	Domain	Role	Education Level
S1	Woman	29	Industry	Developer	Master’s
S2	Woman	32	Industry	Developer	Bachelor’s
S3	Woman	35	Academia	Both	Doctorate
S4	Man	40	Industry	Researcher	Doctorate
S5	Woman	27	Academia	Researcher	Master’s
S6	Man	37	Industry	Developer	Bachelor’s
S7	Woman	31	Industry	Developer	Bachelor’s
S8	Man	33	Industry	Researcher	Doctorate
S9	Man	54	Both	Both	Bachelor’s
S10	Man	49	Industry	Developer	Master’s
S11	Man	36	Industry	Developer	Bachelor’s
S12	Man	22	Academia	Researcher	Bachelor’s

10.3 Study Design

To better understand the creators’ perceived effectiveness of features for each intervention (*Workshops*, *Emulators*, *Evaluators*, *Feedback Collectors*, and *Multi-Modal Automated Tools*), I conducted semi-structured interviews with 12 visualization creators. I present the details and results of the qualitative analysis.

10.3.1 Participants

I randomly selected 12 survey respondents from those who indicated they would participate in a follow-up interview (see Table 10.3). Five participants identified their gender as women and seven as men. Their average age was 35.4 years (SD=9.0). Three participants had attained or were pursuing a doctoral degree, three a master’s degree, and the remaining six a bachelor’s degree. I discontinued recruiting participants once I reached saturation of insights. I compensated participants with a $25 Amazon gift card for participating in the hour-long interview.

Table 10.4: Interventions and their effectiveness features identified during the survey of and interviews with visualization creators.

Intervention	**Features**
Workshops	*Synchronous*
	Co-instructed by screen-reader users and practitioners
	Materials available post-workshop
	Step-by-step instructions with examples
Emulators	*Output in textual format*
	Incorporate output from multiple screen readers
	Toggle activation (on/off)
Evaluators	*Score-based evaluation*
	Explanation of the underlying issue
	Evaluation of presence as well as quality of alt-text
	Evaluation of color contrast ratio
	Toggle activation (on/off)
	Post-creation evaluation (CI/CD pipelines, "linters")
Feedback Collectors	*Simple radio buttons (Likert scale)*
	Optional open-ended field for feedback
	Tracking feedback (GitHub tracker)
Multi-Modal Automated Tools	*Customization and personalization options*
	AI techniques for efficiency
	Low-effort avenues to test technical accuracy

10.3.2 Procedure

I interviewed the participants via Zoom and used its built-in features to record the calls, subsequently transcribing the interviews using Descript (Descript [2017]). The study sessions were semi-structured and lasted one hour. At least three authors administered each interview, one taking detailed notes during each session. I engaged the participants in a conversation to identify the utility, challenges, and suggestions to enhance the effectiveness of the interventions. Specifically, I explored each intervention's features that could assist visualization creators in: (1) understanding screen-reader users' challenges with data visualizations, (2) improving creators' knowledge about data visualization accessibility, and (3) making creators' data visualizations accessible to screen-reader users.

10.3.3 Analysis

To analyze the interviews, I used inductive thematic analysis (Braun and Clarke [2006]). Specifically, I followed a semantic approach (Patton [1990]) and an essentialist paradigm (Widdicombe and Wooffitt [1995]; Potter and Wetherell [1987]). As per prior work's guidelines on thematic analysis (Ryan et al. [2000]; Braun and Clarke [2006]), I used the first two interviews to develop an initial set of codes. I added codes to the codebook, when appropriate, during the analysis of the rest of the interviews, identifying 83 total codes. Three co-authors coded each interview transcript independently, resolving disagreements through mutual discussions. The final analysis identified the utility, challenges, and suggestions for enhanced effectiveness for all five interventions, discussed in the section below.

I calculated inter-rater reliability using Krippendorff's α (Krippendorff [2011]) as well as pairwise percentage agreement, following the recommendation by Landis *et al.* (Landis and Kock [1977]). Computed using ReCal 3.0 (Freelon [2010]), the Krippendorff's α was 0.88, indicating a high level of reliability (Krippendorff [2018]). The average pairwise percentage agreement between authors was also high, at 88.2% (Hartmann [1977]; Graham et al. [2012]).

10.4 Results

I present the findings from the qualitative analysis of the semi-structured interviews with 12 visualization creators. Specifically, I discuss each of the five identified interventions in turn. I show these findings in Table 10.4.

10.4.1 Workshops

The participants preferred synchronous over asynchronous workshops, recognizing several benefits of in-person workshops, including the ability to ask questions in real-time, the value of discussions and collaborations with fellow attendees, and a desire to interact with screen-reader users. For example, S7 identified the advantages of real-time interactive workshops:

> I find that to be pretty useful to have a person in real-time. You can ask your questions and have all of that happen in real-time. Because you wanna make sure those charts are accessible, it's nice if you can go ask someone. (S7)

S3 agreed, saying:

> [In] an in-person workshop, there's more motivation to be present and actually building collaborations and general interactions with other people that might also be interested in that space. (S3)

Similarly, S6 discussed the importance of learning from screen-reader users in workshops:

> If there was a workshop given by someone who uses a screen reader and they give that firsthand perspective and show how they use their screen reader, challenges they run into... If I were to learn more about accessibility and understanding the challenges, having that sort of direct instruction is very valuable. (S6)

The participants also highlighted the importance of learning from the practitioners, emphasizing the need for "multiple expertise levels" (S5) and "relatability" (S3). For example:

> The people I'd like to learn from are people who are sighted and trying to do the same things that I'm doing. Like practitioners, who know what the steps are. (S1)

Additionally, some participants specified time prioritization as a challenge for attending in-person workshops, addressable by making the materials available post-workshop. S5 identified this challenge and also offered a solution:

> `I think the struggle is always just finding time. My ideal format is having the materials be available asynchronously. It takes the burden off of having to be somewhere at a specific time and you feel like you're missing out on content. So there's something just really nice about that format that I feel like you kind of get the best of both worlds.` (S5)

Additionally, P11 offered content suggestions and emphasized the need for step-by-step instructions with examples:

> `Step-by-step... explain [it] like I'm five. 'Accessible visualization for dummies.' What is ARIA? What fields are available to me to use for screen readers? How can I actually make Tableau work with a screen reader? Give me viable examples of steps to follow.` (P11)

Overall, the participants found workshops on data visualization accessibility effective in a synchronous format, in which they could collaborate with other attendees in person and ask questions. They preferred the workshops to be co-instructed by screen- and non-screen-reader users to leverage varying knowledge and expertise. Additionally, they suggested making content available after the workshops as a workaround for time prioritization challenges.

10.4.2 Emulators

The participants found emulators essential for understanding the experiences of screen-reader users. They identified several use cases for emulators, including assisting visualization creators in comprehending how screen readers "read" their visualizations, minimizing the need to install and learn several screen readers, and enhancing their accessibility testing experience. For example, S5 said:

> I think there’s only so much reading about an experience that I can do where understanding how the things that I create are interpreted and read by screen reader. I think that would be useful for that reason. And it’s useful in building empathy too, right? (S5)

S9 said:

> The simple fact that there is more than one [screen reader] out there and that they would need to be installed is a barrier to somebody like me. I mean, obviously, I can just go and install lots of different ones, but then I think if I knew I had the option to not install and simply go to one place, which is going to emulate the experience, I would start there for sure. (S9)

S2 expressed a similar sentiment:

> I’ll say testing with screen readers has always just been a little difficult. For somebody just trying to quickly see in general what’s happening with my data through a screen reader, you could use it for that case. (S2)

S12 expressed a suggestion to emulate the output of a screen reader in textual format:

> I would appreciate if it could just generate text so I can just look at the output of a screen reader. So you kind of wanna first take away the narration part of it and [just] have it in a transcript kind of form. (S12)

Unsurprisingly, some of the participants preferred using the actual screen readers for a better understanding of the experiences of screen-reader users while also identifying use cases where emulators could be beneficial. For example, S11 advocated for using the screen reader itself while also recognizing situations where an emulator could assist him:

> `This is really hard because there is a little bit of me that thinks, "Why not enable the screen reader in the first place?" But I could definitely see a use case for having an emulator because at the moment I've [got] VoiceOver running and I wanna see how it works in NVDA. And I can't really do that very easily. So yeah, actually, to be honest, there's definitely a use case.` (S11)

The participants also emphasized that emulators would be increasingly useful by incorporating the ability to emulate the responses from various screen readers, including users' customizations and preferences. For example, S12 shared his enthusiasm for this feature:

> `There are different screen readers on the market, and visually impaired groups, they tend to use different tools according to their preferences.  So if this emulator can account for all different screen readers, I think that would be great.` (S12)

S2 said she also wanted an emulator to combine screen readers:

> `So, obviously if there's a way to kind of combine all of those screen readers into, you know, this emulator, and you were able to flip to different screen readers to test the nuances, that might be cool.` (S2)

In summary, the visualization-creator participants considered emulators a beneficial intervention to understand screen-reader users' experiences. They also noted that learning a screen reader rather than using an emulator could have potential advantages. However, they recognized that due to the technical and learning challenges involved with installing and using several screen readers, emulators would be a plausible alternative to test visualization accessibility.

10.4.3 Evaluators

In the interviews, visualization creators identified the usage of evaluators as an educational toolkit. For example, S1 recognized the benefit of evaluators to improve knowledge of accessibility:

> That's a common mistake that people don't put ALT-texts, right? So, say it starts to catch that and show that, hey, well no ALT-text. So then it's slowly, in a way, teaching you a little bit and improving your knowledge of accessibility. (S1)

S6 said:

> So like, I think is a really useful functionality of guiding you through in a very verbose way because there's so much, you know, to learn and do and check. I think it is a powerful feature. (S6)

Additionally, the participants highlighted essential features to enhance the effectiveness of evaluators, including a score-based evaluation and an explanation of the underlying accessibility issues with visualization elements. S9 discussed the benefits of a score-based evaluation:

> Well, definitely having an overall score is a good thing. Scoring maybe each element or assessing each element so you can make design choices. (S9)

The participants also provided recommendations to include nuanced evaluations of accessibility measures such as alternative text ("ALT-text"), color contrast ratio, and the ability to toggle the activation of automated checkers. For example, S3 shared her thoughts about alternative text:

> Appropriate ALT-text... You have to be able to recognize what it is and for it to be useful, prompt somebody to actually include meaningful elements in the ALT-text. (S3)

S1 shared the importance of toggling features or outputs of the evaluator:

> Depending on where you don't want a certain behavior and if it keeps giving you an error, then you just get annoyed. Then you don't want to use it anymore. So being able toeither dismiss it or [have] a "linter" where you can turn off rules. (S1)

Furthermore, some participants were interested in evaluating visualizations post-creation in addition to *during* visualization development, such as through Continuous Integration/Continuous Delivery ("CI/CD") pipelines and "linters" (code analysis tools that report programming errors):

> I would want it as part of a test system or a "linting" system. Cause I always see those run on most CI/CD projects before it gets to deployment. I think it'd be good to give that red flag before anyone gets a chance to see and encounter it. (S7)

In summary, the participants identified the use case for evaluators as an educational toolkit for visualization creators, particularly novice creators, to learn about the underlying accessibility issues with their visualizations. They also noted several features of an effective evaluator, including evaluating ALT-text quality and color contrast ratio and using score-based methods during and after visualization creation, such as in CI/CD pipelines.

10.4.4 Feedback Collectors

The visualization-creator participants expressed a keen interest in "welcoming feedback" (S5) on the accessibility of their visualizations from screen-reader users. S9 shared the enthusiasm of the visualization community for this idea:

> It's a fantastic idea! I don't know anybody who's working in this field who doesn't welcome feedback. It helps you refine, improve your work. In the visualization world, everybody wants to get better at making their data visualizations accessible. (S9)

As expected, some participants acknowledged and noted the undue burden screen-reader users face in reporting accessibility issues online. For example, S5 said:

> `I think a lot about user burnout and burdens around constantly having to provide feedback.  So, I want to be cognizant of that and not overwhelm screen-reader users who are wanting to provide feedback.` (S5)

To make the feedback collector effective and minimize the burden on screen-reader users providing feedback, the participants suggested using simple radio-button options with an optional open-ended field to collect the accessibility rating of a visualization. S5 shared some solutions to her concerns:

> `Something simple, something easy with options to provide more feedback if they choose to, give them the agency to how much and not put the burden and onus on them to do a lot.  So, as much as they want to provide is effective for you.` (S5)

S11 shared his ideas for capturing feedback efficiently:

> `Maybe you can make it a little easier, like, quick “yes” or “no.”  For example, was the visualization appropriate?  You just want to keep it simple, but at the same time, you want an open-ended [option], like a mini-survey that they can fill out.` (S11)

Additionally, the participants highlighted that tracking feedback would be a beneficial feature for screen-reader users. For example, S8 used GitHub’s issue tracker as an example to identify its benefits:

> `[By] going to GitHub, you can do issue trackers or report a bug and then it gets posted.  So there should potentially be some way for a dialogue with the creator and also notification of resolution of the feedback.` (S8)

Overall, the participants emphasized the potential usefulness of feedback collectors, typified by the enthusiastic quote, "[E]verybody wants to get better at making their data visualizations accessible" (S9). The participants suggested making feedback collectors simple, similar to a mini-survey, with an optional choice to provide open-ended responses. Additionally, they identified feedback tracking as a potentially beneficial feature.

10.4.5 Multi-Modal Automated Tools

The participants expressed enthusiasm for multi-modal automated tools, identifying their inherent characteristic of being an "easy sell" (S11) for decision-makers and helpful for minimizing visualization creators' challenges with making online data visualizations accessible to screen-reader users. As S11 said:

> So the idea that, well, I have a solution, where I can make [visualizations] more accessible rather than rewriting them. I can't even begin to tell you how appealing that is and how much of an easier sell that would be. (S11)

S12 was similarly enthusiastic:

> It would be great. It would be absolutely awesome to have a tool that's sort of in the middle between the developer and screen reader; it would just be awesome. (S12)

As identified by the participants, customization and personalization options for screen-reader users would make multi-modal automated tools more effective. In particular, the participants recommended utilizing the default screen reader settings of the screen-reader users to aid in developing the personalized tool. S11 discussed his concerns with customization options and offered a solution:

> I worry about too many options and sensible defaults; are they gonna be too prescriptive? So instead, what if I took their preferences that they already have and used them? I definitely think that would certainly help. (S11)

The participants noted that multi-modal automated tools offer limited opportunities to learn about accessibility, but also argued that the goal should be about screen-reader users being able to extract the information from visualizations:

> `It’d be great if people knew about accessibility, but if everything was just magically accessible in the background and you didn’t have to know, that’s the best possible outcome for people who need the stuff to be accessible. So, I believe that having an automated tool that just fixes it for you is the most effective way. The outcome of accessibility should be [that] screen-reader users are able to do stuff.` (S7)

Additionally, the participants recommended using research studies with screen-reader users to develop such tools:

> `So, if it’s a person who does accessibility research and they’re the ones building that tool... So, research done with screen-reader users, and something that has been evaluated with them, assessed with them, taken their preferences into account, that would give you more credibility for the tool.` (S1)

S6 suggested providing low-effort avenues for visualization creators to test for technical accuracy to build trust:

> `I think if the tool was research-based and reacting to direct feedback from screen reader users, [and] best practices from developers, that makes it feel much more valuable. You know... research-based, and has been validated and vetted.` (S6)

Altogether, the participants expressed their exhilaration for multi-modal automated tools, classifying them as an “easy sell” for decision-makers. They offered suggestions to customize and personalize these tools to enhance the experiences of screen-reader users. Additionally, they discussed the importance of using research methods and co-design approaches in building such tools with screen-reader users, considering it essential for building trust in the tool's accuracy.

Utilizing the findings, I organized the educational intervention, *Workshops*, and implemented and integrated the four technological interventions into VOXLENS. Subsequently, I conducted a task-based user study to assess the utility of the findings in enhancing creators' work to make data visualizations accessible. I discuss the integration and assessment, in turn, below.

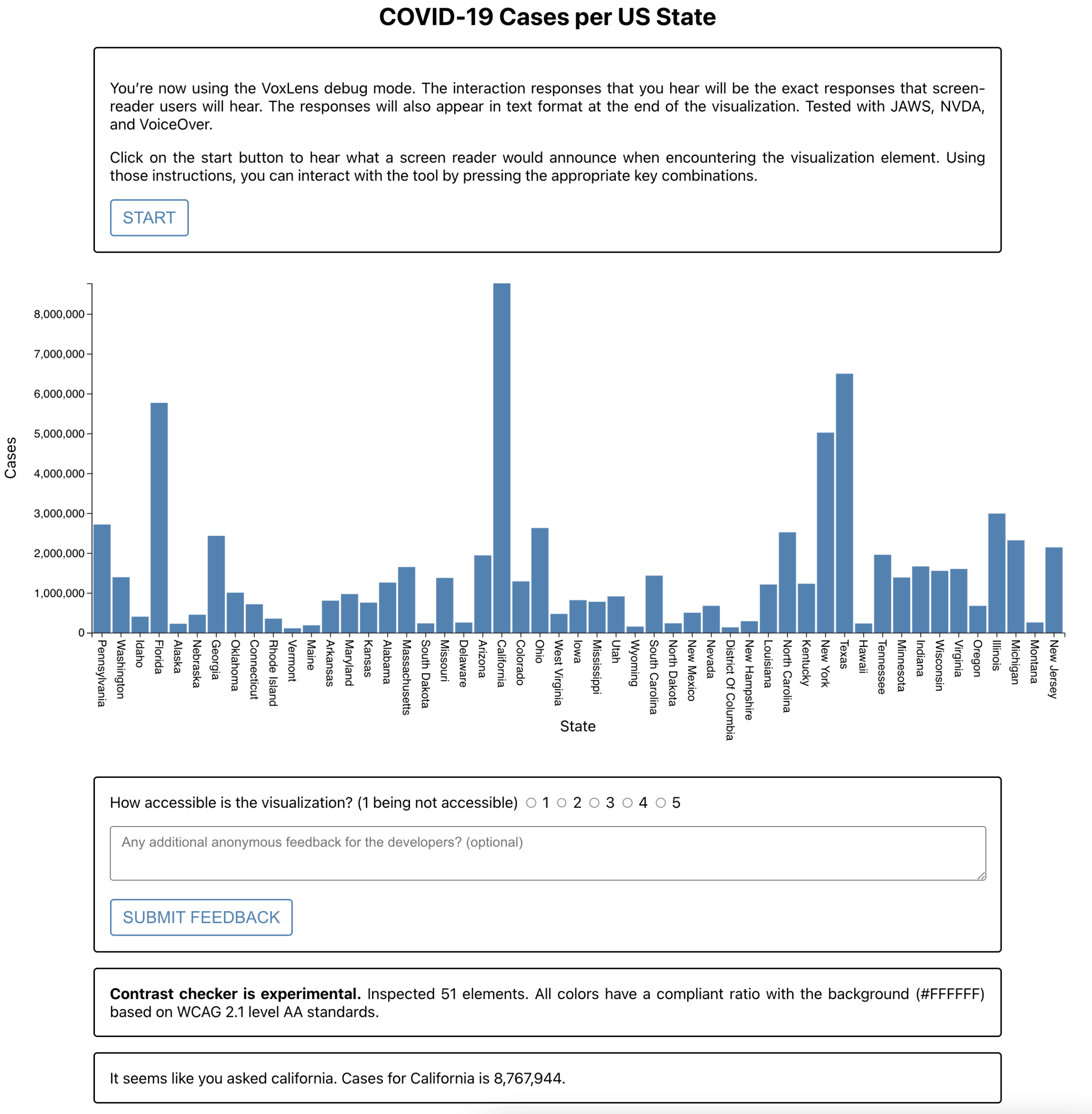
10.5 Integration of Technological Interventions Into VOXLENS

As a secondary contribution of this work, I extended the functionality of VOXLENS by implementing the four technological interventions identified by visualization creators in the studies. The goal in implementing and integrating these interventions into VOXLENS was to assess the generalizability of the findings. Although these enhancements were open source and can assist in creating future solutions, I developed them within the context of my existing system, VOXLENS. Therefore, these enhancements may not be generalizable to other platforms and tools, the exploration of which I leave to future work. Furthermore, as VOXLENS is a technological solution, the implementation and integration focused on the identified technological interventions. I present the educational intervention, namely *Workshops*, in subsequent sections in this chapter. I discuss the four new technological interventions and their implementations below.

I chose VOXLENS for the integration because VOXLENS is (1) a multi-modal solution, (2) open-source, and (3) easy to integrate into existing data visualizations. Additionally, while commercial versions of the interventions exist, VOXLENS provides us with the ability to *collectively* use these interventions *during* the development of online data visualizations via JavaScript libraries.

10.5.1 Video Tutorial

I created a video tutorial for implementing VOXLENS, further explaining the usage of configuration options available in the original and the enhanced version. I uploaded the tutorial to YouTube and provided its link to participants in the step-by-step instructions in the second part of the study. (While a video tutorial can be an educational intervention, I do not consider it a replacement for *Workshops*.)


COVID-19 Cases per US State

You're now using the VoxLens debug mode. The interaction responses that you hear will be the exact responses that screen-reader users will hear. The responses will also appear in text format at the end of the visualization. Tested with JAWS, NVDA, and VoiceOver.

Click on the start button to hear what a screen reader would announce when encountering the visualization element. Using those instructions, you can interact with the tool by pressing the appropriate key combinations.

How accessible is the visualization? (1 being not accessible) ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Any additional anonymous feedback for the developers? (optional)

Contrast checker is experimental. Inspected 51 elements. All colors have a compliant ratio with the background (#FFFFFF) based on WCAG 2.1 level AA standards.

It seems like you asked california. Cases for California is 8,767,944.

Figure 10.1: Screen capture of a sample COVID-19 data visualization with the VoxLens enhancements. Without these enhancements, only the visualization and its title would be shown to the visualization creators.

10.5.2 Debug Mode as Emulator

I introduced a "debug" mode in the VoxLens enhancements to emulate the responses from screen readers. This mode enables visualization creators to hear VoxLens's responses *without* using a screen reader. I relayed the audio output using the Web Speech API, commonly available in modern browsers. As VoxLens responses are consistent across various screen readers, such as JAWS (Scientific [1995]), I implemented the following features from Table 10.4: (1) output in textual format, and (2) toggle activation (on/off). Specifically, visualization creators could deactivate the emulator (speech and text) or turn off the voice output to view the responses only in textual format. With these features, visualization creators can hear and see every piece of information that VoxLens relays to a screen reader, an additional capability that adds emulation and improves multi-modality of VoxLens, which it did not possess before.

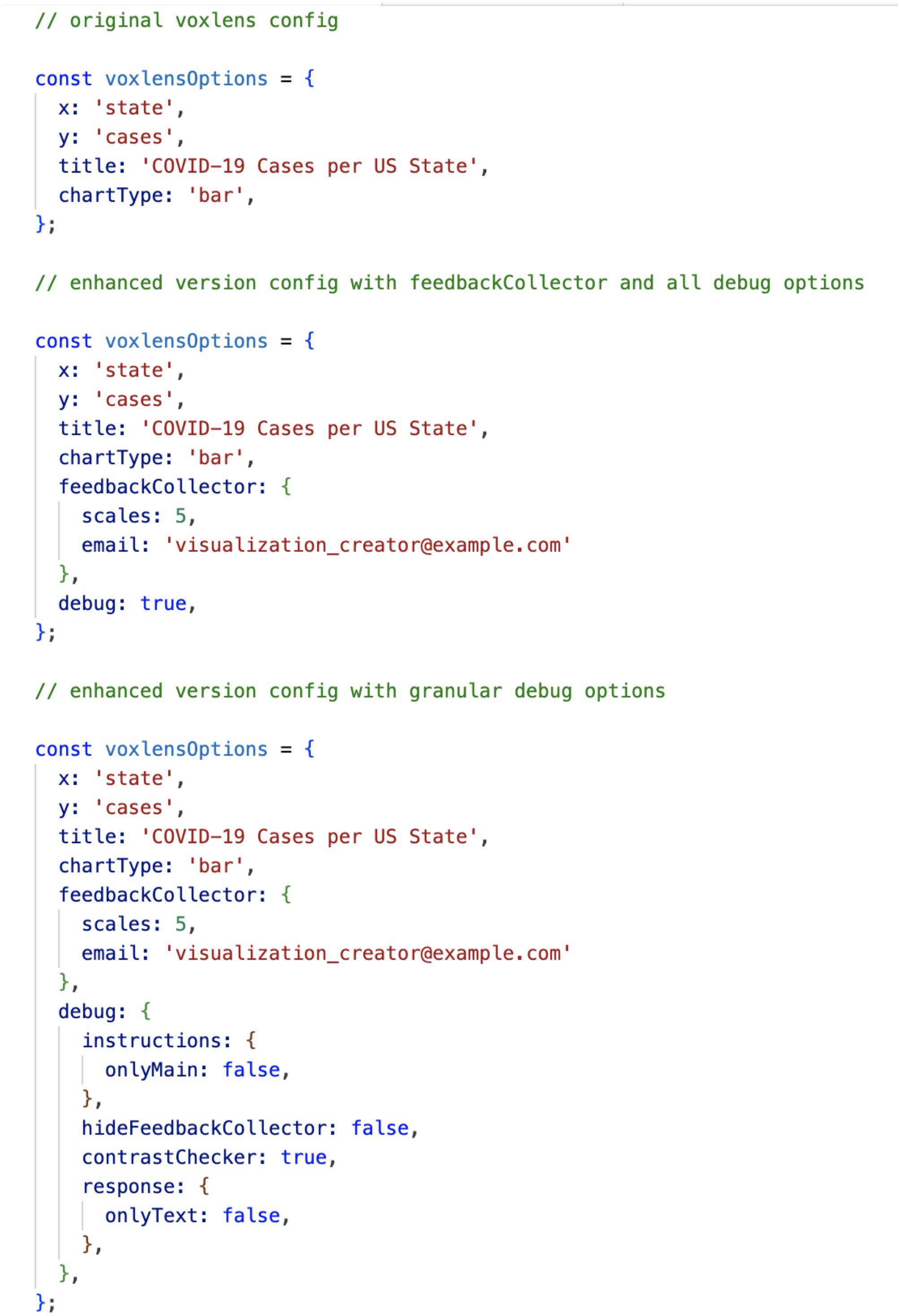
10.5.3 Color Contrast Evaluator

To uphold the interview findings of an evaluator, I built a color contrast evaluator for visualizations that automatically extracts the background color of the visualization and assesses it against the colors of the visualization elements, including the lines, bars, shapes, and text, using the Level AA standards specified in WCAG 2.1 (Caldwell et al. [2018]). To extract the background colors from SVG-based visualizations, I traverse the elements recursively. For visualizations created using HTML Canvases, I convert the canvas element into an image and fetch the color palette using the `Color Thief`[1] library. For background-to-foreground ratios not in compliance with the Level AA standard, I provide the visualization creators with details of each element for easier debugging with an overall score. When in compliance, I show the number of elements inspected and confirmation of compliance (see Figure 10.1). Additionally, per the participants' preferences, I added the ability to disable the evaluator by setting the "contrastChecker" configuration option to "false" (see Figure 10.2), which turns this feature off.

As VoxLens automatically generates a summary of visualizations, I did not implement an evaluator for checking the presence and quality of ALT-text (see Table 10.4). Similarly, I did not implement post-creation tools, as these solutions are internal to the deployment environment of developers and outside of the technical scope of VoxLens.

---

[1] https://lokeshdhakar.com/projects/color-thief/



```
// original voxlens config

const voxlensOptions = {
  x: 'state',
  y: 'cases',
  title: 'COVID-19 Cases per US State',
  chartType: 'bar',
};

// enhanced version config with feedbackCollector and all debug options

const voxlensOptions = {
  x: 'state',
  y: 'cases',
  title: 'COVID-19 Cases per US State',
  chartType: 'bar',
  feedbackCollector: {
    scales: 5,
    email: 'visualization_creator@example.com'
  },
  debug: true,
};

// enhanced version config with granular debug options

const voxlensOptions = {
  x: 'state',
  y: 'cases',
  title: 'COVID-19 Cases per US State',
  chartType: 'bar',
  feedbackCollector: {
    scales: 5,
    email: 'visualization_creator@example.com'
  },
  debug: {
    instructions: {
      onlyMain: false,
    },
    hideFeedbackCollector: false,
    contrastChecker: true,
    response: {
      onlyText: false,
    },
  },
};
```

Figure 10.2: VoxLens configuration for the original version, the enhanced version with default debug options, and the enhanced version with granular debug options.

10.5.4 Feedback Collector

I developed a feedback collector that enables screen-reader users to provide feedback to the visualization creator regarding the accessibility of the visualization. Based on the features identified by the participants (see Table 10.4), I limited the feedback to a simple rating of the visualization accessibility using radio buttons with an optional text box for further comments. By default, the rating is on a 5-point scale ranging from 1 ("not accessible") to 5 ("fully accessible"). Per their preferences, visualization creators can modify the scale and hide the feedback collector in debug mode. A valid email address is required from the visualization creators to activate the feedback collector. I collaborated with and utilized the services from `EmailJS`[2] to implement client-side email sending. I note that screen-reader users provide feedback anonymously to the visualization creators using the feedback collector without any additional risk to their privacy. I did not implement the tracking feature as it requires additional setup from visualization creators, which is outside of the technical scope of VOXLENS.

10.5.5 Algorithm Improvements for Multi-Modality

The survey respondents suggested using advanced and contemporary algorithms for the multi-modal automated tools (see Table 10.4). Specifically, they recommended the use of artificial intelligence. Therefore, I improved the functionality of VOXLENS by using fuzzy logic (Gupta et al. [2018]; Zadeh [1988]; Hellmann [2001]), a natural language processing technique, to replace the current keyword matching algorithm in use by the original VOXLENS. The internal testing revealed that fuzzy logic performed queries faster and more accurately than the keyword matching algorithm, improving the performance of VOXLENS's *Question and Answer* mode. I also did not implement avenues to test technical accuracy, as VOXLENS already provides automated unit and functionality testing with 100% coverage.

Figure 10.1 displays a screen capture of the VOXLENS enhancements, whereas Figure 10.2 shows a comparison of the configuration options between the original VOXLENS and the enhanced version.

---

[2] https://www.emailjs.com/

Table 10.5: Gender, age, domain, role, and education level of the task-based user study participants.

	Gender	Age	Domain	Role	Education Level
V1	Woman	29	Industry	Developer	Master's
V2	Woman	19	Academia	Developer	Bachelor's
V3	Man	20	Academia	Researcher	Bachelor's
V4	Man	19	Both	Developer	Bachelor's
V5	Woman	29	Industry	Developer	Master's
V6	Man	33	Industry	Researcher	Doctorate
V7	Woman	49	Industry	Developer	Master's
V8	Woman	33	Industry	Researcher	Master's
V9	Man	54	Industry	Developer	Bachelor's
V10	Woman	23	Academia	Researcher	Doctorate

10.6 Evaluation

To evaluate the generalizability of the findings, I conducted a task-based user study with 10 visualization creators using the original VOXLENS and the enhanced version. The goal from this study was to investigate the effect of the enhancements on the participants' understanding of screen-reader users' challenges, knowledge of data visualization accessibility, and perceived usefulness of VOXLENS. I present the methodology and results below.

10.6.1 Participants

The participants were 10 visualization creators recruited via social media platforms and email distribution lists for data visualization creators (see Table 10.5). Six participants identified as women and four as men. Their average age was 30.8 years (SD=12.2). The highest level of education was a doctoral degree for two participants; four had a master's degree, and the remaining four had a bachelor's degree. I compensated them with a $100 Amazon gift card for their participation in the task-based user study.

10.6.2 Procedure

The user study was a two-part unsupervised task-based study. Each part included a step-by-step instruction for the participants to complete the task. The first part involved using the original VoxLens (Sharif et al. [2022f]) without any modifications, following the instructions, documentation, and examples from its open-source GitHub repository. In the second part, the participants used the enhanced VoxLens version. At the end of each part, the participants filled out a questionnaire rating their understanding of the challenges screen-reader users experience with online data visualizations (UL), knowledge of data visualizations accessibility (KL), and perceived usefulness of VoxLens (PU) on a scale of 1 to 7 on a Likert scale (with "1" being the lowest and "7" being the highest). The questionnaire also included questions from the NASA-TLX workload questionnaire (Hart and Staveland [1988]; Hart [2006]) and open-ended queries to collect the participants' liked or disliked features, identified areas of improvement, and general comments. I also recorded their demographic information.

10.6.3 Design & Analysis

I conducted a task-based user study to investigate the effect of the VoxLens enhancements on the participants' subjective responses. The independent variable was VoxLens *Version* (VX; within-Ss.), having two levels ("original," "enhanced"). The dependent variables were *Understanding Level* (UL), *Knowledge Level* (KL), and *Perceived Usefulness* (PU); all dependent variables were ordinal (1 to 7 on a Likert scale; "1" being the lowest and "7" being the highest). To analyze the effect of VX on each of these variables, I used mixed ordinal logistic regression (McCullagh [1980]; McKelvey and Zavoina [1975]), identical to the quantitative survey analysis. This model is a type of generalized linear mixed model (GLMM). The focus being a task-based user study and not a controlled experiment, I deliberately had all participants do the original version of VoxLens first, and then the enhanced version, since the latter contains a strict superset of functionality of the first.

To qualitatively assess the open-ended responses from the questionnaires, I followed the same protocol for thematic analysis (Braun and Clarke [2006]) that I used for analyzing the survey responses. To quantitatively assess how VX impacts NASA TLX perceived workload ratings, I performed the nonparametric aligned rank transform (ART) procedure (Wobbrock et al. [2011a]; Elkin et al. [2021]; Higgins and Tashtousch [1994]) (Hart and Staveland [1988]; Hart [2006]); the 1-21 scales were for *mental demand, physical demand, temporal demand, performance, effort,* and *frustration*.

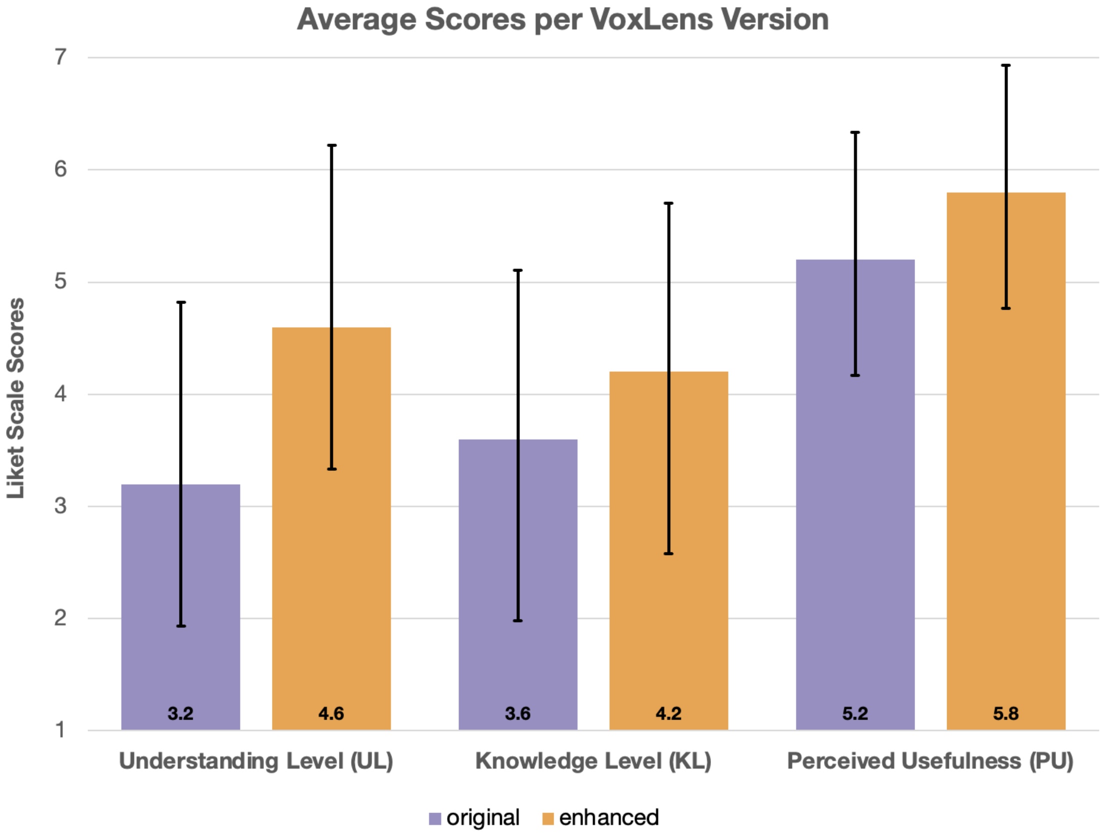


Figure 10.3: Visualization showing the Likert scale scores (with "1" being the lowest and "7" being the highest) for the original and the enhanced VoxLens version. Higher scores are better. Error bars represent mean ±1 standard deviation.

Table 10.6: Statistical results from the ordinal logistic regression analysis from N=10 visualization creators with VoxLens *Version* (VX) as the independent variable ("original" *vs.* "enhanced"). "DV" means dependent variable. All results are statistically significant ($p < .05$).

DV	χ^2	p
Understanding Level (UL)	7.78	< .05
Knowledge Level (KL)	5.88	< .05
Perceived Usefulness (PU)	4.07	< .05

10.6.4 Results

I present the quantitative, qualitative, and workload assessment findings of the VoxLens enhancements.

10.6.4.1 Quantitative Assessment

VoxLens *Version* (VX) had a significant effect on all three ordinal dependent variables, indicating significant differences between the original VoxLens and the enhanced version (see Table 10.6). Overall, there was a 43.8%, 16.7%, and 11.5% increase in *Understanding Level* (UL), *Knowledge Level* (KL), and *Perceived Usefulness* (PU), respectively, for the enhanced version compared to the original VoxLens (see Figure 10.3). These results confirm that the interventions and features revealed by the interviews provided measurable improvements to creators' understanding, knowledge, and utility of visualization accessibility.

10.6.4.2 Qualitative Assessment

I assessed the enhancements' usefulness by analyzing the open-ended questionnaire responses from all participants. Specifically, I asked about their liked and disliked features and improvement areas.

Overall, in comparison to the original VoxLens, participants found the enhanced version to be *valuable for researchers in the field of data visualization; more user-friendly* (V10), *a great inclusion for developers; makes it easier to create accessible visualizations* (V3), and *a good step forward to helping with inclusive data visualization development* (V6). Additionally, participants expressed their excitement by stating *[the enhanced version] provided better accessibility* (V2) and *really useful, can't get over it!* (V5).

The participants also emphasized the usefulness of the emulator feature by classifying it as a good *sanity check that testing was working as expected; an entry point for novice creators* (V1), *definitely enhanced the experience* (V3), *much more convenient than opening and closing a narrator* (V4), and *ease of testing screen reader use without having a screen reader* (V5). Additionally, they found the evaluator feature *handy* (V1) and *a good addition that could save developers time versus having separate tools to do that* (V9). They also shared their liking for the other features, including customization options to toggle these features through configuration options.

For areas of improvement, the participants identified the opportunity to have more working examples and customization settings for the emulator feature to increase and decrease the response speed. At the time of this writing, I have started the process of implementing these suggestions to improve the enhanced VoxLens version.

Table 10.7: Gender, age, and education level of the workshop participants.

	Gender	**Age**	**Education Level**
W1	Woman	31	Master's
W2	Woman	28	Doctorate
W3	Woman	43	Master's
W4	Woman	24	Bachelor's
W5	Woman	27	Bachelor's
W6	Man	43	Doctorate
W7	Woman	38	Master's
W8	Woman	55	Bachelor's
W9	Woman	45	Master's
W10	Woman	29	Bachelor's
W11	Woman	28	Master's

10.6.4.3 Workload Assessment

I collected perceived workload ratings for both VoxLens versions (VX) using the NASA Task Load Index survey (NASA-TLX) (Hart and Staveland [1988]), recording participants' experienced mental demand, physical demand, temporal demand, performance, effort, and frustration. The analysis used the nonparametric aligned rank transform procedure (Wobbrock et al. [2011a]; Elkin et al. [2021]; Higgins and Tashtoush [1994]), but did not reveal a significant effect of *VX* on any of the six workload dimensions, indicating similar workload levels for each version of VoxLens. Put another way, the additional features and tools in the enhanced version of VoxLens did not detectably increase VoxLens's perceived workload.

10.7 Workshop for Visualization Creators

To further assess the generalizability of the findings, I organized a workshop for visualization creators on making online data visualizations accessible. I discuss my findings below.

10.7.1 Participants

I advertised the workshop through collaboration with the Open Scholarship Commons and the eScience Institute at the University of Washington. The participants were 14 visualization creators (see Table 10.7), all affiliated with the University of Washington in varying capacities (*e.g.,* students, teachers, and staff). Twelve identified their gender identity as women and two as men. Their average age was 37.1 years (SD=10.1). Three had attained or were pursuing a doctoral degree, six a master's degree, and the remaining five a bachelor's degree. Participants attended the workshop voluntarily and did not receive financial compensation.

10.7.2 Procedure

I designed the workshop incorporating the effectiveness features identified by visualization creators in the interviews (see Table 10.4). Therefore, the workshop took place in an in-person setting, lasting an hour. The workshop comprised three sections: (1) *The Problem*, (2) *Common Practices*, and (3) *Future Avenues*. Before the beginning of the workshop, the participants filled out the same questionnaire that I presented to the participants in the task-based user study. Specifically, they specified their knowledge, prioritization, and perceived importance of accessibility using a Likert scale ranging from 1–7 ("1" being lowest and "7" being highest). Additionally, using the scale, participants specified their rating for challenges they face with data visualization accessibility and their desire to conduct studies with screen-reader users. To avoid the Hawthorne Effect (Jones [1992]; Sedgwick and Greenwood [2015]), participants were not made aware that they were being studied until the end of the workshop.

In the first section, my goal was to assist participants in building an in-depth understanding of the problem of data visualization accessibility. Therefore, I illustrated the inaccessibility of online data visualizations through a recorded interaction of a screen-reader user with an online data visualization depicting COVID-19 cases per U.S. state. Next, I played a 10-minute-long pre-recorded interview with a screen-reader user, who shared his experiences with digital accessibility in general and with online data visualizations in particular.

Next, I demonstrated four common modalities used to make data visualizations accessible: (1) ALT-text, (2) sonification, (3) 3-D printing, and (4) data tables. Following the demonstration, the attendees partook in an activity to discuss these modalities and their experiences with these modalities with another attendee, subsequently sharing the summary of their discussion with the rest of the attendees. After the activity, I presented the advantages and shortcomings of each modality, building on the discussion from the activity.

Finally, in the third section, I demonstrated the functionalities and integration of VoxLens into existing visualizations to make these data visualizations accessible. Further, I discussed multi-modality and other resources beyond the four modalities. Lastly, I provided attendees with three "homework" topics to ponder with their colleagues and community members in making data visualizations accessible to screen-reader users: (1) equity, (2) user agency, and (3) non-keyboard-based interactions. I aspire to organize workshops on these topics separately in the future. At the end of the workshop, participants filled out the same questionnaire as the one at the beginning of the study.

Table 10.8: Statistical results from the mixed ordinal logistic regression analysis from N=11 visualization creators with *Questionnaire Time* (QT) as the independent variable ("pre-workshop" *vs.* "post-workshop"). "DV" means dependent variable. Results with $p < .05$ are statistically significant.

DV	χ^2	p
Knowledge Level (KL)	7.53	< .05
Prioritization Level (PL)	3.44	.063
Perceived Importance (PI)	9.12	< .05
Challenge Level (CL)	3.51	.061
Studies Frequency (SF)	15.09	< .001

10.7.3 Design & Analysis

I investigated the differences in participants' pre- and post-workshop subjective responses. The independent variable was *Questionnaire Time* (QT; within-Ss.), having two levels ("pre-workshop," "post-workshop"). The dependent variables were *Knowledge Level* (KL), *Priority Level* (PL), *Perceived Importance* (PI), *Challenge Level* (CL), and *Studies Frequency* (SF); all dependent variables were ordinal (1 to 7 on a Likert scale; "1" being the lowest and "7" being the highest). Identical to VoxLens' task-based user study, I used

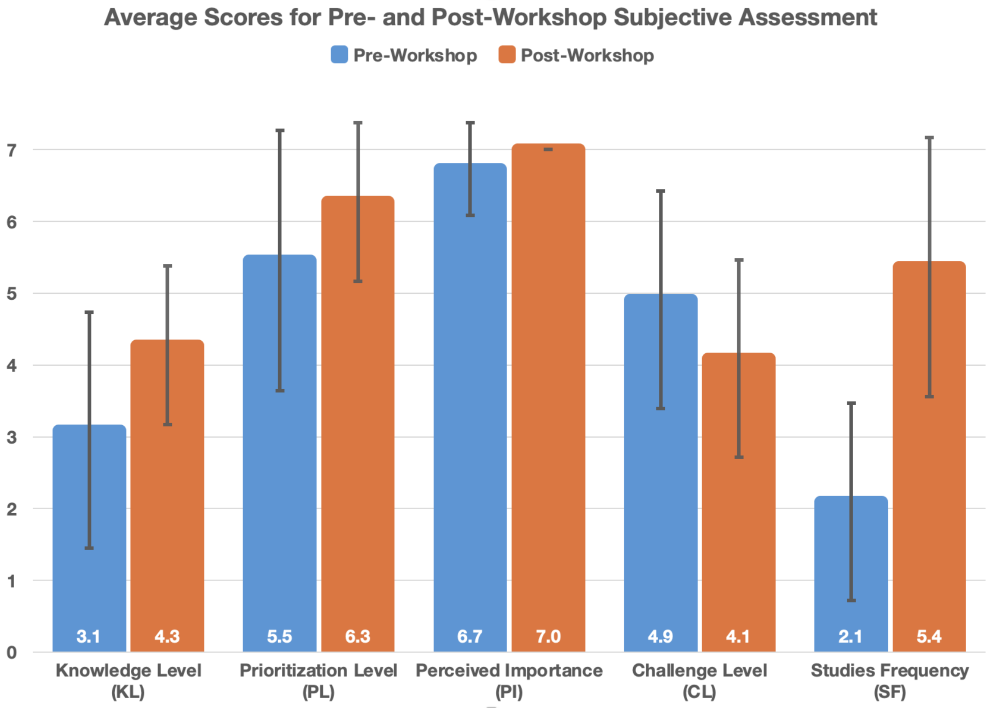


Figure 10.4: Visualization showing the Likert scale scores (with "1" being the lowest and "7" being the highest) for the pre- and post-workshop subjective assessments. Higher scores are better. Error bars represent mean ±1 standard deviation.

mixed ordinal logistic regression (McCullagh [1980]; McKelvey and Zavoina [1975]) to analyze the effect of QT on each of these variables. Three participants did not finish the post-workshop questionnaire and were therefore not included in the analysis.

10.7.4 Results

Questionnaire Time (QT) had a significant effect on *Knowledge Level* (KL), *Perceived Importance* (PI), and *Studies Frequency* (SF). QT had a marginal effect on *Priority Level* (PL) ($p \approx .063$) and *Challenge Level* (CL) ($p \approx .061$). These findings indicate significant differences between the participants' subjective responses pre- and post-workshop (see Table 10.8). Specifically, post-workshop scores show an improvement in participants' knowledge, prioritization, and perceived importance of accessibility by 38%, 4%, and 15%, respectively. Furthermore, participants decreased their ratings of the challenges they face with accessibility by 17% and increased their desire to conduct studies with screen-reader users by 157% (see Figure

10.4). These results confirm that the educational intervention of *Workshop* provided measurable improvements for visualization creators with data visualization accessibility.

Overall, participants shared positive reviews about the workshop (*e.g.,* “Great increase in my knowledge” [W9] and “Really enjoyed the workshop” [W10]). For example, W11 appreciated learning about the advantages and shortcomings of various modalities:

> This was a fantastic workshop. I really appreciated being able to think about each approach to creating more accessible visualizations and the pros/cons. I know I’ve attended a good workshop when I leave with more exciting questions to explore. (W11)

Similarly, W8 said:

> I wasn’t sure what to expect. The information gave attainable ways to present data visualizations accessibly and seemed to make the task less daunting. (W8)

W2 recognized an opportunity for the workshop materials to be included in data visualization courses:

> Expectations were met and exceeded! This is so practical and useful! I recently took the Data Viz class and this should be incorporated there. (W2)

Additionally, W6 liked knowing about the state of accessibility:

> This presentation gave me a good sense of what the state of accessibility is with regard to data viz. (W6)

10.8 Uniqueness of Each Intervention

The participants identified and discussed the effectiveness of five interventions: *Workshops*, *Emulators* (EM), *Evaluators* (EV), *Feedback Collectors* (FC), and *Multi-Modal Automated Tools* (MM). Assessing against the three objectives of enhancing visualization creators' (1) understanding of screen-reader users' challenges with data visualizations, (2) knowledge of data visualization accessibility, and (3) ease of making data visualizations accessible, the participants determined that each intervention was beneficial in fulfilling the three objectives only to a certain degree. For example, MM would drastically improve the ease of making visualizations accessible but would be less effective in understanding the challenges screen-reader users experience with data visualizations. In contrast, these benefits and limitations would be vice versa for EM. Therefore, to achieve all three objectives, these interventions should be utilized conjointly, leveraging the unique benefits they offer visualization creators and screen-reader users.

10.9 Summary

In this chapter, I sought to understand and improve visualization creators' (1) understanding of screen-reader users' challenges with data visualizations, (2) knowledge of visualization accessibility, and (3) ease of making data visualizations accessible to screen-reader users. To this end, I surveyed visualization creators to determine and identify interventions to minimize their challenges with visualization accessibility. These interventions were: *Workshops*, *Emulators*, *Evaluators*, *Feedback Collectors*, and *Multi-Modal Automated Tools*. Then, I conducted semi-structured interviews to examine the effective versions of each of the five interventions and reported the specific features that make these interventions effective. To validate these findings, I implemented and integrated the technological interventions into VoxLens and performed a task-based user study with visualization creators to assess the generalizability of the findings and the usefulness of the open-source enhancements.

Using these enhancements, visualization creators improved their understanding of screen-reader users' challenges with data visualizations, knowledge of visualization accessibility, and perceived usefulness of VoxLens. Additionally, I organized the educational intervention, *Workshop*, finding that it improved creators' knowledge, prioritization, and perceived importance of accessibility while also decreasing their rating of challenges they face with accessibility increasing their desire to conduct studies with screen-reader users. Overall, I presented the effective versions of the interventions I used to achieve the three objectives as generalizable knowledge for researchers and practitioners, particularly in the data visualization community.

Chapter 11

Conclusion

This dissertation has demonstrated a new approach to improve the accessibility of online data visualizations for screen-reader users while reducing the challenges visualization creators experience with data visualization accessibility. At the beginning of this dissertation, I presented my formative research that shed light on the challenges of screen-reader users and their disenfranchisement compared to non-screen-reader users with information extraction from online data visualizations. To minimize these challenges, this dissertation introduced a multi-modal, customizable, and interactive JavaScript plug-in called VOXLENS. Further, this dissertation presented several versions of VOXLENS, VOXEX—a system that enables customization of information consumption from online data visualizations, SONIFIER—an open-source JavaScript library that sonifies two-dimensional data, and five interventions that improve the knowledge and experiences of visualization creators with data visualization accessibility. Altogether, the evaluation of these artifacts showed performance enhancements in the accuracy of information extraction and interaction times for screen-reader users. Additionally, these artifacts improved the knowledge and experiences of visualization creators in making online data visualizations accessible.

11.1 Major Empirical Findings

In this dissertation, I presented findings from empirical evaluations of several VOXLENS versions conducted using the baseline of conventional information extraction from online data visualizations without VOXLENS. These results are highlighted below.

Screen-Reader Users

- *Accuracy of Information Extraction.* With VOXLENS integrated into visualizations, screen-reader users extracted information 164% more accurately compared to those without VOXLENS' integration. Additionally, VOXLENS improved their accuracy of information extraction from being 62% less accurate to being 6% *more* accurate than non-screen-reader users.
- *Interaction Times.* Screen-reader users reduced their interaction times with online data visualizations by 50% using VOXLENS. Compared to non-screen-reader users, their interaction times decreased from spending 211% more time extracting information from online data visualizations to 56%.

Visualization Creators

- *Understanding Level.* Using VOXLENS, visualization creators enhanced their understanding of the challenges screen-reader users experience with online data visualization by 44%.
- *Knowledge Level.* Visualization creators increased their knowledge of making online data visualizations accessible to screen-reader users by 17% using VOXLENS. Additionally, the workshop on making data visualizations accessible increased their knowledge level by 38%.
- *Perceived Usefulness.* VOXLENS enhancements to support visualization creators improved their perceived usefulness of VOXLENS by 12%.
- *Prioritization Level.* The workshop enhanced the prioritization level of visualization creators to make online data visualizations accessible to screen-reader users by 4%.
- *Perceived Importance.* Visualization creators boosted their perceived importance of making data visualizations accessible by 15% after participating in the workshop.
- *Challenge Level.* The workshop assisted the visualization creators by decreasing their rating of challenges they encounter in making data visualization accessible by 17%.
- *Studies Frequency.* After attending the workshop, the participants reported a 157% increase in their desire to conduct studies with screen-reader users to test the accessibility of their data visualizations.

11.2 Reflections and Insights

Today, in 2024, even with the recent bloom of large language models (LLMs), equitable access to digital information for screen-reader users remains sparse. When I began this work in 2018, the recommended measures for developers to make online data visualizations accessible were limited to providing alternative textual descriptions (alt-text) for visualizations. While several other modalities, such as sonification, have gained recognition since, and recommendations now include using multi-modal approaches, alt-text remains the most widespread measure to make online data visualizations accessible to screen-reader users. For example, ChartJS (ChartJS [2015]), a popular open-source and free JavaScript library that enables users to create data visualizations using HTML5, states the following on its web page on accessibility: *Chart.js charts are rendered on user-provided* `canvas` *elements. Thus, it is up to the user to create the* `canvas` *element in a way that is accessible. The* `canvas` *element has support in all browsers and will render on screen but the* `canvas` *content will not be accessible to screen readers. With* `canvas`*, the accessibility has to be added with ARIA attributes on the* `canvas` *element or added using internal fallback content placed within the opening and closing canvas tags.*

However, unlike other digital content that can be sufficiently described using alt-text, data visualizations encapsulate granular information that may not be adequate to represent as a one-size-fits-all alt-text, especially considering that each screen-reader user may have unique needs and preferences. This issue raises concerns of *equality* for screen-reader users to access information from online data visualizations like non-screen-reader users. On the other hand, data tables provide screen-reader users access to the entire dataset, arguably indicating *equality* in information access. However, given the linear navigation characteristic of screen readers, extracting information from data tables could lead to undue cognitive overload, especially for high-cardinality or complex datasets. As opposed to concerns of *equality*, this issue poses concerns of *equity*. I believe that for a solution to be truly accessible, it needs to be *equal* as well as *equitable*. This belief was ingrained in the creation of VOXLENS to provide screen-reader users the agency to consume information from online data visualizations per their preferences while improving their information extraction accuracy and interaction times.

A common perception in providing screen-reader users access to digital information is that their mode of interaction is solely keyboard-based. This dissertation has shown that VOXLENS, a plug-in that utilizes a non-keyboard-based interaction, offers significant enhancements in the information consumption of screen-reader users from online data visualizations. For example, one of the screen-reader participants in my study stated, "*I especially liked the part of the tool where you can ask it a question and it would give you the information back. I thought it was brilliant actually. I felt like being able to ask it a question made everything go a lot faster and it took a lot less brain power I think. I felt really confident about the answers that it was giving back to me*." Ultimately, motivated by my empirical findings and intuition, I believe that utilizing non-keyboard-based interactions, such as the verbal question-and-answering mode of VOXLENS, not only offers the potential to improve the experiences of screen-reader users with consuming information from digital content but also, in turn, assist researchers and practitioners in identifying uncharted benefits and use cases of new technologies for these users and incorporating accessibility from the beginning when building technological solutions.

My formative research (discussed in Chapter 3) reported that the average interaction time for screen-reader users without VOXLENS was 84.6 seconds. VOXLENS improved these interaction times by 50%, reducing the average time to 42.4 seconds. Although this is a significant improvement, screen-reader users spent 56% more time interacting with the visualizations than non-screen-reader users, accentuating the disparity between the two user groups' interaction times. Several factors contribute to the difference in interaction times between screen-reader- and non-screen-reader users (Sharif et al. [2022f]; Bigham et al. [2007]), including the duration of the issuing a verbal query and auditory responses. Thus, this difference is expected. The implications of these findings can help guide existing and future voice assistants for screen-reader users to improve information extraction.

Another interesting observation was the paradox of using multi-modal automated tools (*e.g.,* VOXLENS). Although the participants expressed a need for "*a magic wand charting tool where a chart is just made accessible in the background without the chart author needing to do anything special*," some participants also identified how this tool could restrict visualization creators from fully understanding the challenges of screen-reader users and enhancing their knowledge of accessibility. While I remained impartial during the course of the study, my ultimate opinion on this matter is one that was eloquently stated by one of the participants: "*It'd be great if people knew about accessibility, but the outcome of accessibility should be that people who use screen readers are able to do stuff*." Similarly, one could assume the five interventions presented in this dissertation as “empathy simulations,” a notion criticized by Bennett and Rosner (Bennett and Rosner [2019]). This dissertation does not explore whether these interventions build empathy for screen-reader users. Instead, it examines the utility of these interventions to understand the *technology*.

I believe that effective implementation of accessibility requires consideration of both content consumers and producers. In the case of online data visualizations, I consider these demographics to be screen-reader users and visualization creators. Of course, my foremost objective was to reduce the information access gap between screen-reader- and non-screen-reader users by enabling them to consume information with increased accuracy, reduced interaction times, and in the manner they prefer. But without supporting visualization creators and raising awareness about the importance of accessibility and the disenfranchisement screen-reader users experience with information extraction from data visualizations, this objective would have been only partially fulfilled. For example, providing visualization creators with emulators helps them understand how screen-reader users “see” their content and assists them in making necessary adjustments. Therefore, I supported both demographics in this dissertation and remain of the opinion that accessibility-related work should incorporate studies with both content consumers and producers.

One of the participants expressed in her interview that “*it’s pretty important and critical to have agency over the information you consume*.” Another participant said, “*It allows me to do things quickly and efficiently, get the information, glance at it like a sighted person*.” These quotes align with those from my other studies. Screen-reader users are limited to the content chosen by the visualization creators, unlike non-screen-reader users, who have more freedom and options to interact with and consume information from online data visualizations. Even popular multi-modal tools generate the summary of the data visualization based on the most commonly extracted information, which may not satisfy the needs of some screen-reader users. I believe giving these users the agency to customize the information they consume can enhance their experiences with online data visualizations. I consider user agency as an example of *customization* and view it as distinct from *personalization*, which automatically tailors the content based on the behaviors and patterns determined by user interactions. In this dissertation, I have provided an approach for customization and identified future avenues for personalization.

The need for customization, however, is not just limited to screen-reader users. In this work, I recognized that the same is true for visualization creators, especially given that the participants appreciated the customization options to toggle the features in the enhanced version of VoxLens. For example, one participant conveyed an essential component for the evaluators to *turn off rules* so people wouldn't get *annoyed* and, in turn, *wouldn't wanna use it anymore*. Therefore, I recommend researchers further examine the differences in these options based on granular factors, including their domain (e.g., industry, academia), role (e.g., developer, researcher), or experience with data visualization accessibility.

Overall, since its inception in 2018, this work has posed several challenges over the years that have not only enhanced the rigor and revealed nuances of this work but also assisted in my personal and professional growth as a researcher, making it even more rewarding. One of these challenges was the recruitment of diverse screen-reader users while considering varying factors, including their backgrounds, vision levels and diagnoses, education levels, gender identities, and age groups. Furthermore, studies with these users, which were initially designed to be conducted in an in-person lab environment, had to be appropriately re-calibrated to be carried out virtually, a detour needed to appropriately address the safety concerns of the COVID-19 pandemic. Additionally, technical challenges involved workarounds for limitations of and dependencies on Web Speech API for voice recognition and Natural Language Processing algorithms for query processing.

Finally, as discussed in this dissertation, VoxLens comprised five versions. Each version was a result of a different exploration, targeting a separate problem. For example, Chapter 5 presented the VoxLens version for simple two-dimensional graphs, Chapter 7 extended this version to support drilled-down information extraction from complex visualizations, and Chapter 8 introduced the support for communicating information uncertainty. Then, Chapter 9 contributed the VoxLens version to provide screen-reader users agency in consuming information. Following suit, Chapter 10 offered enhancements to assist visualization creators. I believe that this approach is especially beneficial for open-source accessibility-related solutions, as it enables community contributors and researchers to address newly recognized and previously unaddressed accessibility challenges continually.

11.3 Contributions

This dissertation makes two categories of contributions: (1) *empirical*, and (2) *artifact*. These categories are derived from Wobbrock and Kientz's article on research contributions in human-computer interaction (Wobbrock and Kientz [2016]). I discuss the contributions for each category below.

11.3.1 Empirical Contributions

- Detailed account of the challenges and disenfranchisement screen-reader users experience with data visualizations compared to non-screen-reader users.
- Performance evaluation of screen-reader users when interacting with online data visualizations compared to non-screen-reader users, using the accuracy of information extraction and interaction times as metrics.
- Results from formative studies, including surveys, interviews, and user studies, that provide motivation and design decisions for each version of VoxLens.
- Results from summative studies, including interviews, longitudinal studies, task-based experiments, and case studies, that evaluate the performance of each VoxLens version.
- Results from semi-structured interviews that shed light on the preferences of screen-reader users in obtaining information on uncertainty in data visualizations.
- Results from surveys and interviews that offer screen-reader users' customization needs and preferences from various modalities and potential technological interventions.
- Taxonomies of the information sought by screen-reader users in their holistic and drilled-down explorations of online data visualizations.
- Results from surveys and interviews with visualization creators that provide a detailed account of the challenges visualization creators experience in making online data visualizations accessible and their proposed interventions.
- Design and assessment results from an in-person workshop on making online data visualizations accessible for visualization creators.
- Decision tree for visualization creators to assist them in making informed decisions about using sonification to make data visualizations accessible to screen-reader users.
- Recommendations and guidelines for visualization creators to make online data visualizations accessible to screen-reader users.

11.3.2 Artifact Contributions

- Six versions of VOXLENS, an interactive JavaScript plug-in that improves the accessibility of online data visualizations for screen-reader users. VOXLENS offers a multi-modal solution, enabling screen-reader users to explore online data visualizations, both holistically and in a drilled-down manner, using voice-activated commands.
- SONIFIER, an open-source JavaScript library that generates sonified responses from two-dimensional data, providing visualization creators with the choice of using various waveforms and oscillators.
- VOXEX, a system that provides agency to screen-reader users to customize the information they consume from online data visualizations. VOXEX comprises a browser extension and a configuration portal for screen-reader users to specify their preferences.
- Implementations of four technological interventions (*Emulator*, *Evaluator*, *Feedback Collector*, and *Multi-Modal Automated Tool*) to support visualization creators in making online data visualizations accessible to screen-reader users.

11.4 Future Work

In addition to significantly improving the accuracy of information extraction for screen-reader users while simultaneously reducing their interaction times, this work identifies further improvements for VOXLENS and presents exciting future avenues of research. These research directions include (1) exploring visualization authoring by screen-reader users, (2) investigating the utility of non-keyboard-based interaction techniques with data visualizations, and (3) supporting context-based information extraction from online data visualizations. I discuss each of these avenues below.

11.4.1 Visualization Authoring by Screen-Reader Users

The primary focus of this dissertation is to improve the *consumption* of information contained in online data visualizations by screen-reader users. Extending this research work to assist screen-reader users in *producing* online data visualizations can significantly improve data visualization accessibility. Future work can utilize the epistemologies present in this dissertation to understand and iteratively improve the experiences of screen-reader users through mixed-methods studies in authoring online visualizations. For example, naturalistic observational studies in diverse settings can render invaluable insights. Similarly, longitudinal studies can reveal critical metrics. Compared to consuming data visualizations, producing visualizations is a largely uncharted domain, presenting several open research questions, including how to support screen-reader users in validating visual encodings in visualizations and debugging their creations without external assistance.

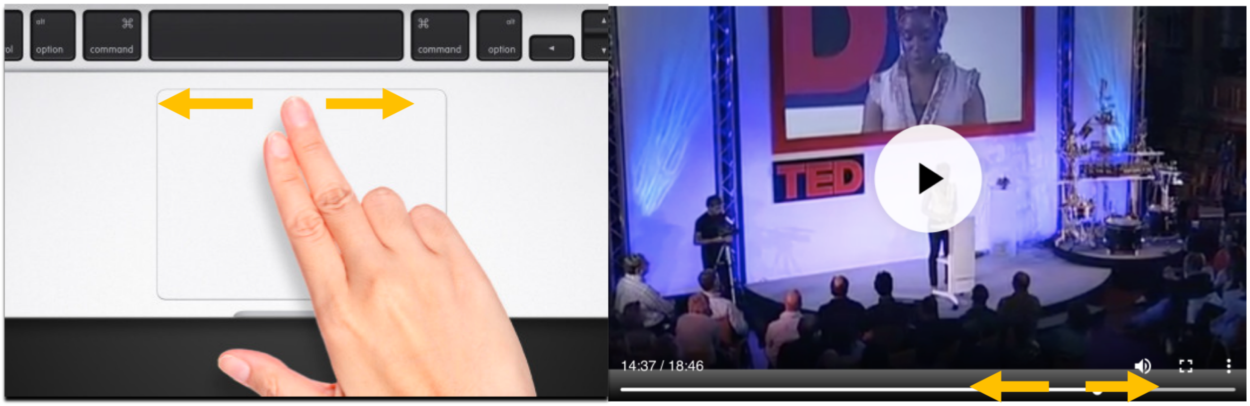


Figure 11.1: Using the touchpad to control a video slider on a web page.

11.4.2 Non-Keyboard-Based Interaction Techniques

VoxLens is the first system to incorporate a non-keyboard-based interaction mechanism for information extraction from data visualizations for screen-reader users. Given its success in bridging the information access gap, future work can leverage other interaction techniques to enhance the experiences of screen-reader users. For example, I explored the utility of touchpads on laptops as input devices for screen-reader users through the development and preliminary evaluation of Touchpad Mapper (Sharif et al. [2024c]), a system that maps digital content (e.g., image and video) to the physical coordinates of a touchpad (see Figure 11.1). My assessment shows that participants expressed they extracted information faster using my system than a conventional keyboard. Similarly, gesture recognition is a well-established interaction technique for people with limited fine motor function and can benefit screen-reader users. Similarly, haptic feedback can also improve their information extraction experiences.

11.4.3 Context-Based Information Extraction

A recurrent theme in this dissertation work is "one size does not fit all," which emphasizes that interaction and information extraction from visualizations differ not only across individuals but also contexts. For example, appropriate methods for extracting information from visualizations containing weather data may differ from those containing financial data. Hence, supporting context-based information extraction can improve the experiences of screen-reader users with data visualizations. Future work can investigate the application of Natural Language Processing and Machine Learning in fostering context-based conversational interactions for screen-reader users with online data visualizations.

11.4.4 Improvements for Sonification

In this work, I performed an *overall* assessment of screen-reader users' experiences with sonification in interpreting data from online visualizations. Therefore, I did not explore its domain-specific usage for expert users (*e.g.*, for screen-reader user brokers who interact with financial data visualizations extensively). I encourage future work to investigate the experiences of screen-reader users with sonification in different disciplines, domains, and professions. Additionally, my findings highlighted the unfamiliarity of screen-reader users with sonification and their emphasis on learning how to interpret sonified responses. Therefore, I did not conduct task-based usability studies to evaluate the use and usefulness of sonification, as subject matter knowledge could have been a significant confound in the analyses. Future work can explore efficient learning methods for screen-reader users to minimize this confound. *In situ* think-aloud usability studies could also assist in understanding how sonification is used in practice by screen reader users, undoubtedly revealing additional insights to what I have presented in this work.

11.4.5 Improvements for VoxLens

The experiments in this dissertation included three different visualization libraries and two-dimensional data to evaluate and compare the visualization interaction experiences of screen-reader users with those of non-screen-reader users. While I chose the most commonly used visualization libraries and all the participants were evaluated using two-dimensional data, different visualization libraries and data dimensionality could extend this work. Therefore, future work could examine the visualization interaction experiences of screen-reader users in comparison to those of non-screen-reader users, using different visualization libraries, such as Highcharts and Recharts, and different data dimensionality, such as three-dimensional data. Additionally, my exploration was limited to single-series bar graphs, multi-series line graphs, and geospatial maps. Future work can utilize this methodology to understand and improve screen-reader users' experiences with other complex data visualizations.

VOXLENS is currently only fully tested on Google Chrome, as support for the Web Speech API's speech recognition is currently limited to Google Chrome. Future work could consider alternatives to the Web Speech API that offers cross-browser support for speech recognition. Furthermore, VOXLENS versions were limited in parsing users' input commands due to restrictions from the keyword-matching algorithm and the Web Speech API's voice recognition. Future work can employ advanced Natural Language Processing and Conversational Question and Answering algorithms to handle complex and nuanced input queries.

11.4.6 Improvements for VOXEX

Based on the suggestions from the participants, I enhanced VOXEX by enabling screen-reader users to specify their domains of interest. However, I did not include the subdomains in this exploration. For example, a user might only be interested in the "basketball" subdomain under the broader domain of "sports." Future work can incorporate these subdomains into VOXEX and investigate strategies to enable users to specify their subdomains of interest with minimal effort overload. Additionally, I did not explore AI-powered personalization in this work. Future work can utilize machine learning paradigms, such as few-shot learning and continuous learning, in conjunction with customization to cater to the individual needs of the screen-reader users. For example, these paradigms can help determine users' subdomains of interest without requiring them to input this information manually.

11.4.7 Extended Support for Visualization Creators

Survey findings revealed a significant correlation between some variables used in my study with visualization creators, such as their knowledge of data visualization accessibility and their prioritization of making data visualizations accessible to screen-reader users. As my objective was to identify tools to support visualization creators, I did not explore the presence of causality between the variables based on these correlations.

I encourage future work to investigate these correlations further, particularly to determine possible causal relationships between these variables.

11.5 Final Remarks

This dissertation has demonstrated the following thesis:

> *A multi-modal, customizable, and interactive JavaScript plug-in called "VoxLens" improves the experiences of screen-reader users in extracting information from simple and complex online data visualizations while also enhancing the knowledge of visualization creators to make online data visualizations accessible.*

This work introduced a novel interaction technique for screen-reader users to interact with online data visualization through five versions of VoxLens, an open-source multi-modal JavaScript plug-in that improves the accessibility of online data visualizations for screen-reader users. Each version was created using findings from formative studies and subsequently evaluated using a mixed-methods approach. These evaluations have indeed demonstrated that VoxLens significantly improved the accuracy of information extraction and interaction times of screen-reader users and enhanced their experiences by providing them agency in customizing their information consumption. Additionally, this dissertation has demonstrated an improvement in the knowledge of visualization creators to make data visualizations accessible.

Data visualization accessibility has gained positive momentum over the course of this work, with several researchers exploring plausible approaches. While this is a promising trend, it remains of pivotal importance that accessibility-related solutions are designed through a proper understanding of the challenges of screen-reader users and their involvement and diverse representation throughout the design and development process. As Eric Eggert, a web accessibility expert, eloquently said, "*People who are tasked with remediating accessibility often have little experience of how people with disabilities actually use the web. This leads to overcomplicated solutions, as they underestimate the capabilities of disabled people*." (Eggert [2022])

Bibliography

Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. 2018. "siri talks at you": An empirical investigation of voice-activated personal assistant (vapa) usage by individuals who are blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '18, page 249–258, New York, NY, USA. Association for Computing Machinery.

Cengiz Acarturk and C. Habel. 2012. Eye tracking in multimodal comprehension of graphs. *CEUR Workshop Proceedings*, 887:11–25.

NV Access. 2006. Nv access | download. `https://www.nvaccess.org/download/`. (Accessed on 08/08/2021).

Dustin Adams, Tory Gallagher, Alexander Ambard, and Sri Kurniawan. 2013. Interviewing blind photographers: Design insights for a smartphone application. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, New York, NY, USA. Association for Computing Machinery.

Dragan Ahmetovic, Federico Avanzini, Adriano Baratè, Cristian Bernareggi, Gabriele Galimberti, Luca A. Ludovico, Sergio Mascetti, and Giorgio Presti. 2019a. Sonification of rotation instructions to support navigation of people with visual impairment. In *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*, pages 1–10, New York, NY, USA. Institute of Electrical and Electronics Engineers.

Dragan Ahmetovic, Niccolò Cantù, Cristian Bernareggi, João Guerreiro, Sergio Mascetti, and Anna Capietto. 2019b. Multimodal exploration of mathematical function graphs with audiofunctions.web. In *Pro-*

ceedings of the 16th International Web for All Conference, W4A '19, New York, NY, USA. Association for Computing Machinery.

Dragan Ahmetovic, João Guerreiro, Eshed Ohn-Bar, Kris M. Kitani, and Chieko Asakawa. 2019c. Impact of expertise on interaction preferences for navigation assistance of visually impaired individuals. In *Proceedings of the 16th International Web for All Conference*, W4A '19, New York, NY, USA. Association for Computing Machinery.

Ruben Alcaraz Martinez, Mireia Ribera Turró, and Toni Granollers Saltiveri. 2022. Methodology for heuristic evaluation of the accessibility of statistical charts for people with low vision and color vision deficiency. *Universal access in the information society*, 21(4):863–894.

C N Alexander and G D Scriven. 1977. Role playing: An essential component of experimentation. *Personality and Social Psychology Bulletin*, 3(3):455–466.

Duane F Alwin and Brett A Beattie. 2016. The kiss principle in survey design: question length and data quality. *Sociological methodology*, 46(1):121–152.

Theodore W Anderson and Donald A Darling. 1954. A test of goodness of fit. *Journal of the American statistical association*, 49(268):765–769.

Gennady Andrienko, Natalia Andrienko, Steven Drucker, Jean Daniel Fekete, Danyel Fisher, Stavros Idreos, Tim Kraska, Guoliang Li, Kwan Liu Ma, Jock D. Mackinlay, et al. 2020. Big data visualization and analytics: Future research challenges and emerging applications.

Apple. 2021a. Audio graphs | apple developer documentation. `https://developer.apple.com/documentation/accessibility/audio_graphs`. (Accessed on 08/01/2021).

Apple. 2021b. Bring accessibility to charts in your app - wwdc21 - videos - apple developer. `https://developer.apple.com/videos/play/wwdc2021/10122/`. (Accessed on 07/03/2023).

Claudia L Aranda, Diane K Levy, and Sierra Stoney. 2015. Role playing. In *Handbook of Practical Program Evaluation*, pages 383–411. Wiley Online Library, Hoboken, NJ, USA.

David Austin and Volker Sorge. 2023. Authoring web-accessible mathematical diagrams. In *Proceedings of the 20th International Web for All Conference*, W4A '23, page 148–152, New York, NY, USA. Association for Computing Machinery.

Ronald J. Baken and Robert F. Orlikoff. 2000. *Clinical Measurement of Speech and Voice*. Speech Science. Singular Thomson Learning, San Diego, California, USA.

Stephen Barrass and Gregory Kramer. 1999. Using sonification. *Multimedia systems*, 7(1):23–31.

David Barter. 2023. Degrees of freedom: Designing information and communication technologies to support enhanced agency for blind and partially sighted individuals through cross-sensory information representation.

Cynthia L Bennett and Daniela K Rosner. 2019. The promise of empathy: Design, disability, and knowing the" other". In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–13.

Donald A Berry. 1987. Logarithmic transformations in anova. *Biometrics*, pages 439–456.

Jacques Bertin. 1983. Semiology of graphics; diagrams networks maps. Technical report, JSTOR.

Vittoria Biagi, Riccardo Patriarca, and Giulio Di Gravio. 2022. Business intelligence for it governance of a technology company. *Data*, 7(1):2.

Jeffrey P. Bigham, Anna C. Cavender, Jeremy T. Brudvik, Jacob O. Wobbrock, and Richard E. Ladner. 2007. Webinsitu: A comparative analysis of blind and sighted browsing behavior. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '07, page 51–58, New York, NY, USA. Association for Computing Machinery.

Syed Masum Billah, Vikas Ashok, Donald E. Porter, and I.V. Ramakrishnan. 2017. Ubiquitous accessibility for people with visual impairments: Are we there yet? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 5862–5868, New York, NY, USA. Association for Computing Machinery.

Matthew Blanco, Jonathan Zong, and Arvind Satyanarayan. 2022. Olli: An extensible visualization library for screen reader accessibility. *IEEE VIS Posters*, 6.

National Federation of the Blind. 2019. Blindness statistics | national federation of the blind. `https://nfb.org/resources/blindness-statistics?q=resources%2Fblindness-statistics`. (Accessed on 04/10/2021).

Perkins School for the Blind. 2023. Sonification summary page – perkins school for the blind. `https://www.perkins.org/resource/sonification-summary-page`. (Accessed on 07/03/2023).

Stella U. Boess. 2008. First steps in role playing. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, page 2017–2024, New York, NY, USA. Association for Computing Machinery.

Niall Bolger, Angelina Davis, and Eshkol Rafaeli. 2003. Diary methods: Capturing life as it is lived. *Annual review of psychology*, 54(1):579–616.

Nicholas Bolten and Anat Caspi. 2019. Accessmap website demonstration: Individualized, accessible pedestrian trip planning at scale. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 676–678.

Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315.

Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309.

Richard E Boyatzis. 1998. *Transforming qualitative information: Thematic analysis and code development*. sage.

Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101.

Matthew Brehmer, Bongshin Lee, Petra Isenberg, and Eun Kyoung Choe. 2018. Visualizing ranges over time on mobile phones: a task-based crowdsourced evaluation. *IEEE transactions on visualization and computer graphics*, 25(1):619–629.

Matthew Brehmer and Tamara Munzner. 2013. A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics*, 19(12):2376–2385.

Cynthia Brewer, Mark Harrower, and The Pennsylvania State University. 2021. Colorbrewer: Color advice for maps. `https://colorbrewer2.org/`. (Accessed on 07/09/2023).

Stephen Brewster. 2002. Visualization tools for blind people using multiple modalities. *Disability and rehabilitation*, 24(11-12):613–621.

Anke M. Brock, Jon E. Froehlich, João Guerreiro, Benjamin Tannert, Anat Caspi, Johannes Schöning, and Steve Landau. 2018. Sig: Making maps accessible and putting accessibility in maps. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, page 1–4, New York, NY, USA. Association for Computing Machinery.

Ken Brodlie, Rodolfo Allendes Osorio, and Adriano Lopes. 2012. A review of uncertainty in data visualization. In *Expanding the frontiers of visual analytics and visualization*, pages 81–109. Springer, New York, USA.

Peter Brophy and Jenny Craven. 2007. Web accessibility. *Library trends*, 55(4):950–972.

Craig Brown and Amy Hurst. 2012. Viztouch: automatically generated tactile visualizations of coordinate spaces. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, pages 131–138.

Lorna M. Brown, Stephen A. Brewster, Ramesh Ramloll, Mike Burton, and Beate Riedel. 2003. Design guidelines for audio presentation of graphs and tables. In *Proceedings of the 9th International Conference on Auditory Display*, pages 284–287, Boston University, USA. Citeseer.

Marion Buchenau and Jane Fulton Suri. 2000. Experience prototyping. In *Proceedings of the 3rd Confer-*

ence on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS '00, page 424–433, New York, NY, USA. Association for Computing Machinery.

Ben Caldwell, Michael Cooper, Loretta Guarino Reid, Gregg Vanderheiden, Wendy Chisholm, John Slatin, and Jason White. 2008. Web content accessibility guidelines (wcag) 2.0. *WWW Consortium (W3C)*.

Ben Caldwell, Loretta Guarino Reid, Gregg Vanderheiden, Wendy Chisholm, John Slatin, and Jason White. 2018. Web content accessibility guidelines (wcag) 2.1.

Julia Cambre, Ying Liu, Rebecca E Taylor, and Chinmay Kulkarni. 2019. Vitro: Designing a voice assistant for the scientific lab workplace. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, pages 1531–1542, New York, NY, USA. Association for Computing Machinery.

Dustin Carroll, Suranjan Chakraborty, and Jonathan Lazar. 2013. Designing accessible visualizations: The case of designing a weather map for blind users. In *International Conference on Universal Access in Human-Computer Interaction*, pages 436–445, Berlin, Heidelberg. Springer, Springer Berlin Heidelberg.

Johny Cassidy. 2023. Blind news audiences are being left behind in the data visualisation revolution: here's how we fix that | reuters institute for the study of journalism. `https://reutersinstitute.politics.ox.ac.uk/blind-news-audiences-are-being-left-behind-data-visualisation-revolution-heres-how-we-fix`. (Accessed on 08/21/2023).

John M Chambers, MV Mathews, and FR Moore. 1974. Auditory data inspection. *Report TM*, pages 74–122.

ChartJS. 2015. Accessibility | chart.js. `https://www.chartjs.org/docs/3.5.1/general/accessibility.html`. (Accessed on 01/08/2022).

Albert Cherns. 1976. The principles of sociotechnical design. *Human relations*, 29(8):783–792.

Jinho Choi, Sanghun Jung, Deok Gun Park, Jaegul Choo, and Niklas Elmqvist. 2019. Visualizing for the non-visual: Enabling the visually impaired to use visualization. *Computer Graphics Forum*, 38(3):249–260.

Pramod Chundury, Biswaksen Patnaik, Yasmin Reyazuddin, Christine Tang, Jonathan Lazar, and Niklas Elmqvist. 2022. Towards understanding sensory substitution for accessible visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1084–1094.

Peter Ciuha, Bojan Klemenc, and Franc Solina. 2010. Visualization of concurrent tones in music with colours. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, page 1677–1680, New York, NY, USA. Association for Computing Machinery.

Mary J Clinkenbeard. 2020. A posthuman approach to agency, disability, and technology in social interactions. *Technical Communication Quarterly*, 29(2):115–135.

Jacob Cohen. 1973. Eta-squared and partial eta-squared in fixed factor anova designs. *Educational and psychological measurement*, 33(1):107–112.

Vanessa Colella, Richard Borovoy, and Mitchel Resnick. 1998. Participatory simulations: Using computational objects to learn about dynamic systems. In *CHI 98 Conference Summary on Human Factors in Computing Systems*, CHI '98, page 9–10, New York, NY, USA. Association for Computing Machinery.

Federal Communications Commission. 1998. Section 508 of the rehabilitation act | federal communications commission. `https://www.fcc.gov/general/section-508-rehabilitation-act`. (Accessed on 03/07/2024).

Joel Cooper. 1976. Deception and role playing: On telling the good guys from the bad guys. *American Psychologist*, 31(8):605.

Milagros Costabel. 2021. Being blind in a digital world. `https://www.lacunavoices.com/explore-world-with-lacuna-voices/being-blind-in-digital-world-social-media-inernet-accessibility`. (Accessed on 04/08/2024).

Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of oz studies—why and how. *Knowledge-based systems*, 6(4):258–266.

Jiamin Dai, John Miedema, Sebastian Hernandez, Alexandra Sutton-Lalani, and Karyn Moffatt. 2023. Cognitive accessibility of digital payments: A literature review. In *Proceedings of the 20th International Web for All Conference*, pages 116–121.

Joel J Davis. 2002. Disenfranchising the disabled: The inaccessibility of internet-based health information. *Journal of Health Communication*, 7(4):355–367.

Patrick Dengler, Anthony Grasso, Chris Lilley, Cameron McCormack, Doug Schepers, and Jonathan Watt. 2011. Scalable vector graphics (svg) 1.1.

Descript. 2017. Descript | all-in-one video & podcast editing, easy as a doc. `https://www.descript.com/`. (Accessed on 08/09/2023).

Google Developers. 2014. Charts.

Andrew Dillon. 2002. Beyond usability: process, outcome and affect in human-computer interactions.

Evanthia Dimara, Harry Zhang, Melanie Tory, and Steven Franconeri. 2022. The unmet data visualization needs of decision makers within organizations. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4101–4112.

Tin Doan, Shelagh Mooney, and Peter B Kim. 2023. The moments of truth: A qualitative exploration of service interactions between employees with disabilities in the food service industry, and their customers. *International Journal of Hospitality Management*, 115:103602.

Norman R Draper and Harry Smith. 1981. Applied regression analysis, john wiley and sons. *New York*, 407.

Roger L. DuBois. 2017. Web audio speech synthesis / recognition for p5.js.

Steve Duenes. 2022. Jaime tanner joins as accessibility visuals editor | the new york times company. `https://www.nytco.com/press/jaime-tanner-joins-as-accessibility-visuals-editor/`. (Accessed on 08/21/2023).

Michael Alin Efroymson. 1960. Multiple regression analysis. *Mathematical methods for digital computers*, pages 191–203.

Eric Eggert. 2022. No accessibility without disabilities · eric eggert. `https://yatil.net/blog/no-accessibility-without-disabilities#:~:text=People%20who%20are%20tasked%20with,the%20capabilities%20of%20disabled%20people`. (Accessed on 04/20/2024).

Kent Eisenhuth and Kai Chang. 2022. An accessibility-first approach to chart visual design — smashing magazine. `https://www.smashingmagazine.com/2022/07/accessibility-first-approach-chart-visual-design/`. (Accessed on 03/08/2024).

Frank Elavsky, Cynthia Bennett, and Dominik Moritz. 2022. How accessible is my visualization? evaluating visualization accessibility with chartability. *Computer Graphics Forum*, 41(3):57–70.

Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An aligned rank transform procedure for multifactor contrast tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, page 754–768, New York, NY, USA. Association for Computing Machinery.

Niklas Elmqvist. 2023. Visualization for the blind. *Interactions*, 30(1):52–56.

Xuanhe Er and Yunqi Sun. 2021. Visualization analysis of stock data and intelligent time series stock price prediction based on extreme gradient boosting. In *2021 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)*, pages 272–279, New York, NY, USA. IEEE, Institute of Electrical and Electronics Engineers.

Danyang Fan, Alexa Fay Siu, Hrishikesh Rao, Gene Sung-Ho Kim, Xavier Vazquez, Lucy Greco, Sile O'Modhrain, and Sean Follmer. 2023. The accessibility of data visualizations on the web for screen reader users: Practices and experiences during covid-19. *ACM Transactions on Accessible Computing*, 16(1):1–29.

Danyang Fan, Alexa Fay Siu, Wing-Sum Adrienne Law, Raymond Ruihong Zhen, Sile O'Modhrain, and Sean Follmer. 2022. Slide-tone and tilt-tone: 1-dof haptic techniques for conveying shape characteristics of graphs to blind users. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery.

Christopher J. Ferguson. 2016. An effect size primer: A guide for clinicians and researchers. In A.E. Kazdin, editor, *Methodological issues and strategies in clinical research*, pages 301—310. American Psychological Association, Washington, DC, USA.

Leo Ferres, Gitte Lindgaard, Livia Sumegi, and Bruce Tsuji. 2013. Evaluating a tool for improving accessibility to charts and graphs. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(5):1–32.

Stefano Ferretti, Silvia Mirri, Catia Prandi, and Paola Salomoni. 2016. Automatic web content personalization through reinforcement learning. *Journal of Systems and Software*, 121:157–169.

Stephen Few and Perceptual Edge. 2007. Data visualization: past, present, and future. *IBM Cognos Innovation Center*, pages 1–12.

John H. Flowers. 2005. Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. In *Proceedings of the 11th Meeting of the International Conference on Auditory Display*, pages 406–409, Limerick, Ireland. Citeseer.

John H Flowers, Dion C Buhman, and Kimberly D Turnage. 1997. Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. *Human Factors*, 39(3):341–351.

OpenJS Foundation. 2009. Node.js. `https://nodejs.org/en/`. (Accessed on 08/08/2021).

The GraphQL Foundation. 2021. Graphql | a query language for your api. `https://graphql.org/`. (Accessed on 08/08/2021).

Brigitte N. Frederick. 1999. Fixed-, random-, and mixed-effects anova models: A user-friendly guide for increasing the generalizability of anova results. In B. Thompson, editor, *Advances in Social Science Methodology*, page 111–122. JAI Press, Stamford, Connecticut.

Deen G. Freelon. 2010. Recal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, 5(1):20–33.

Jon E Froehlich, Anke M Brock, Anat Caspi, João Guerreiro, Kotaro Hara, Reuben Kirkham, Johannes Schöning, and Benjamin Tannert. 2019. Grand challenges in accessible maps. *interactions*, 26(2):78–81.

Connor Geddes, David R Flatla, Garreth W Tigwell, and Roshan L Peiris. 2022. Improving colour patterns to assist people with colour vision deficiency. In *CHI Conference on Human Factors in Computing Systems*, pages 1–17.

Andreas Gegenfurtner, Erno Lehtinen, and Roger Säljö. 2011. Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational psychology review*, 23(4):523–552.

Andrea Gerino, Lorenzo Picinali, Cristian Bernareggi, Nicolò Alabastro, and Sergio Mascetti. 2015. Towards large scale evaluation of novel sonification techniques for non visual shape exploration. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '15, page 13–21, New York, NY, USA. Association for Computing Machinery.

AR Gilmour, RD Anderson, and AL Rae. 1985. The analysis of binomial data by a generalized linear mixed model. *Biometrika*, 72(3):593–599.

Nicholas A. Giudice, Hari Prasath Palani, Eric Brenner, and Kevin M. Kramer. 2012. Learning non-visual graphical information using a touch-based vibro-audio interface. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, page 103–110, New York, NY, USA. Association for Computing Machinery.

Timo Götzelmann. 2016. Lucentmaps: 3d printed audiovisual tactile maps for blind and visually impaired people. In *Proceedings of the 18th international ACM Sigaccess conference on computers and accessibility*, pages 81–90.

Nikolai A Grabowski and Kenneth E Barner. 1998. Data visualization methods for the blind using force feedback and sonification. In *Telemanipulator and Telepresence Technologies V*, volume 3524, pages 131–139. SPIE.

Matthew Graham, Anthony Milanowski, and Jackson Miller. 2012. Measuring and promoting inter-rater agreement of teacher and principal performance ratings. *Online Submission*.

Martin S Greenberg. 1967. Role playing: An alternative to deception? *Journal of Personality and Social Psychology*, 7(2p1):152.

Saul Greenberg and Bill Buxton. 2008. Usability evaluation considered harmful (some of the time). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, page 111–120, New York, NY, USA. Association for Computing Machinery.

Sylvie Grosjean, Jean-Luc Ciocca, Amélie Gauthier-Beaupré, Emely Poitras, David Grimes, and Tiago Mestre. 2022. Co-designing a digital companion with people living with parkinson's to support self-care in a personalized way: The ecare-pd study. *DIGITAL HEALTH*, 8:20552076221081695.

The PostgreSQL Global Development Group. 1996. Postgresql: The world's most advanced open source database. `https://www.postgresql.org/`. (Accessed on 08/08/2021).

William Grussenmeyer, Jesel Garcia, Eelke Folmer, and Fang Jiang. 2017. Evaluating the accessibility of the job search and interview process for people who are blind and visually impaired. In *Proceedings of the 14th International Web for All Conference*, W4A '17, New York, NY, USA. Association for Computing Machinery.

Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How many interviews are enough? an experiment with data saturation and variability. *Field methods*, 18(1):59–82.

Charu Gupta, Amita Jain, and Nisheeth Joshi. 2018. Fuzzy logic in natural language processing–a closer view. *Procedia computer science*, 132:1375–1384.

Pedro Antonio Gutiérrez, Maria Perez-Ortiz, Javier Sanchez-Monedero, Francisco Fernandez-Navarro, and Cesar Hervas-Martinez. 2015. Ordinal regression methods: survey and experimental study. *IEEE Transactions on Knowledge and Data Engineering*, 28(1):127–146.

Melita Hajdinjak and France Mihelic. 2004. Conducting the wizard-of-oz experiment. *Informatica (Slovenia)*, 28(4):425–429.

Patrick A.V. Hall and Geoff R. Dowling. 1980. Approximate string matching. *ACM computing surveys (CSUR)*, 12(4):381–402.

Jordan Harband, Shu-yu Guo, Michael Ficarra, and Kevin Gibbons. 1999. Standard ecma-262.

Sandra G Hart. 2006. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA.

Sandra G. Hart and Lowell E. Staveland. 1988. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, North-Holland, Netherlands.

Donald P Hartmann. 1977. Considerations in the choice of interobserver reliability estimates. *Journal of applied behavior analysis*, 10(1):103–116.

Martin Hellmann. 2001. Fuzzy logic introduction. *Université de Rennes*, 1(1).

Sabine Hennig, Fritz Zobl, and Wolfgang W Wasserburger. 2017. Accessible web maps for visually impaired users: Recommendations and example solutions. *Cartographic Perspectives*, 88:6–27.

Jaylin Herskovitz, Andi Xu, Rahaf Alharbi, and Anhong Guo. 2023. Hacking, switching, combining: Understanding and supporting diy assistive technology design by blind people. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA. Association for Computing Machinery.

James J Higgins and Suleiman Tashtoush. 1994. An aligned rank transform test for interaction. *Nonlinear World*, 1(2):201–211.

Highcharts. 2009. Sonification | highcharts. `https://www.highcharts.com/docs/accessibility/sonification`. (Accessed on 08/01/2021).

Ronald R Hocking. 1976. A biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics*, pages 1–49.

Jake Holland. 2017. New york times' upshot editor discusses data visualization, storytelling. `https://dailynorthwestern.com/2017/05/03/campus/new-york-times-upshot-editor-discusses-data-visualization-storytelling/`. (Accessed on 03/05/2022).

Leona M Holloway, Cagatay Goncu, Alon Ilsar, Matthew Butler, and Kim Marriott. 2022. Infosonics: Accessible infographics for people who are blind using sonification and voice. In *CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Karen Holtzblatt and Sandra Jones. 1995. Conducting and analyzing a contextual interview (excerpt). In *Readings in Human–Computer Interaction*, pages 241–253. Elsevier.

Md Naimul Hoque, Md Ehtesham-Ul-Haque, Niklas Elmqvist, and Syed Masum Billah. 2023. Accessible data representation with natural sound. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA. Association for Computing Machinery.

MH Hoyle. 1973. Transformations: An introduction and a bibliography. *International Statistical Review/Revue Internationale de Statistique*, pages 203–223.

Weijian Hu, Kaiwei Wang, Kailun Yang, Ruiqi Cheng, Yaozu Ye, Lei Sun, and Zhijie Xu. 2020. A comparative study in real-time scene sonification for visually impaired people. *Sensors*, 20(11):3222.

Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. 2014. Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 21(3):420–433.

Jessica Hullman. 2019. Why authors don't visualize uncertainty. *IEEE transactions on visualization and computer graphics*, 26(1):130–139.

Todd Hunt. 1982. Raising the issue of ethics through use of scenarios. *The Journalism Educator*, 37(1):55–58.

Amy Hurst and Shaun Kane. 2013. Making" making" accessible. In *Proceedings of the 12th international conference on interaction design and children*, pages 635–638.

Apple Inc. 2005. Accessibility - vision - apple. `https://www.apple.com/accessibility/vision/`.

Facebook Inc. 2013. React – a javascript library for building user interfaces. `https://reactjs.org/`. (Accessed on 08/08/2021).

Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P Bigham. 2011. Supporting blind photography. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, pages 203–210.

Chris Johnson. 2004. Top scientific visualization research problems. *IEEE Computer Graphics and Applications*, 24(4):13–17.

Shuli Jones, Isabella Pedraza Pineros, Daniel Hajas, Jonathan Zong, and Arvind Satyanarayan. 2023. " customization is key": Four characteristics of textual affordances for accessible data visualization. *arXiv preprint arXiv:2307.08773*.

Stephen RG Jones. 1992. Was there a hawthorne effect? *American Journal of sociology*, 98(3):451–468.

Shakila Cherise S Joyner, Amalia Riegelhuth, Kathleen Garrity, Yea-Seul Kim, and Nam Wook Kim. 2022. Visualization accessibility in the wild: Challenges faced by visualization designers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery.

Hernisa Kacorri. 2017. Teachable machines for accessibility. *SIGACCESS Access. Comput.*, 119:10–18.

Aasim Kamal, Parashar Dhakal, Ahmad Y Javaid, Vijay K Devabhaktuni, Devinder Kaur, Jack Zaientz, and Robert Marinier. 2021. Recent advances and challenges in uncertainty visualization: a survey. *Journal of Visualization*, 24(5):861–890.

Shaun K. Kane, Jeffrey P. Bigham, and Jacob O. Wobbrock. 2008. Slide rule: Making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '08, page 73–80, New York, NY, USA. Association for Computing Machinery.

Chloe Keilers, Garreth W Tigwell, and Roshan L Peiris. 2023. Data visualization accessibility for blind and low vision audiences. In *International Conference on Human-Computer Interaction*, pages 399–413. Springer.

Dae Hyun Kim, Enamul Hoque, and Maneesh Agrawala. 2020. Answering questions about charts and generating visual explanations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, page 1–13, New York, NY, USA. Association for Computing Machinery.

Edward Kim and Kathleen F McCoy. 2018. Multimodal deep learning using images and text for information graphic classification. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 143–148.

Edward Kim, Connor Onweller, and Kathleen F McCoy. 2021a. Information graphic summarization using a collection of multimodal deep neural networks. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 10188–10195. IEEE.

Hee-Woong Kim, Sumeet Gupta, and Joon Koh. 2011. Investigating the intention to purchase digital items in social networking communities: A customer value perspective. *Information & Management*, 48(6):228–234.

Nam Wook Kim, Grace Ataguba, Shakila Cherise Joyner, Chuangdian Zhao, and Hyejin Im. 2023. Beyond alternative text and tables: Comparative analysis of visualization tools and accessibility methods. In *Computer Graphics Forum*, volume 42, pages 323–335. Wiley Online Library.

Nam Wook Kim, Shakila Cherise Joyner, Amalia Riegelhuth, and Y Kim. 2021b. Accessible visualization: Design space, opportunities, and challenges. *Computer Graphics Forum*, 40(3):173–188.

Jan Kocoń, Igor Cichecki, Oliwier Kaszyca, Mateusz Kochanek, Dominika Szydło, Joanna Baran, Julita Bielaniewicz, Marcin Gruza, Arkadiusz Janz, Kamil Kanclerz, et al. 2023. Chatgpt: Jack of all trades, master of none. *Information Fusion*, 99:101861.

Mario Konecki, Charles LaPierre, and Keith Jervis. 2018. Accessible data visualization in higher education. In *2018 41st international convention on information and communication technology, electronics and microelectronics (MIPRO)*, pages 0733–0737, New York, NY, USA. IEEE, Institute of Electrical and Electronics Engineers.

Stephen M Kosslyn. 1989. Understanding charts and graphs. *Applied cognitive psychology*, 3(3):185–225.

Klaus Krippendorff. 2011. Computing krippendorff's alpha-reliability. Retrieved from.

Klaus Krippendorff. 2018. *Content analysis: An introduction to its methodology*. SAGE Publications Inc., Pennsylvania, USA.

Ravi Kuber, Wai Yu, and Graham McAllister. 2007. Towards developing assistive haptic feedback for visually impaired internet users. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 1525–1534.

Michal Kubovics and Pavel Bielik. 2021. Visualization of data and keywords in online journalism. *Marketing Identity*, 9(1):121–133.

Aditi Kulkarni, Allan Wang, Lynn Urbina, Aaron Steinfeld, and Bernardine Dias. 2016. Robotic assistance in indoor navigation for people who are blind. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 461–462. IEEE.

LabSound. 2021. Github - labsound/labsound: :speaker: graph-based audio engine. `https://github.com/LabSound/LabSound`. (Accessed on 07/04/2023).

Pauli PY Lai. 2013. Adapting data table to improve web accessibility. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, pages 1–4.

Ali Lakhani, Donna McDonald, and Heidi Zeeman. 2018. Perspectives of self-direction: a systematic review of key areas contributing to service users' engagement and choice-making in self-directed disability services and supports. *Health & social care in the community*, 26(3):295–313.

Richard J. Landis and Gary G. Kock. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

CM Laney, KS Baker, DPC Peters, and KW Ramsey. 2013. Recommendations for data accessibility. *Long-term trends in ecological systems. a basis for understanding responses to global change. Technical Bulletin*, (1931):216–225.

Julianna Langston. 2022. Github - julianna-langston/chart2music: Turns charts into music so the blind can hear data. `https://github.com/julianna-langston/chart2music`. (Accessed on 07/04/2023).

Megan M Lawrence and Amy K Lobben. 2011. The design of tactile thematic symbols. *Journal of Visual Impairment & Blindness*, 105(10):681–691.

Jonathan Lazar, Alfreda Dudley-Sponaugle, and Kisha-Dawn Greenidge. 2004. Improving web accessibility: a study of webmaster perceptions. *Computers in human behavior*, 20(2):269–288.

David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. 2018. Evaluation strategies for hci toolkit research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–17, New York, NY, USA. Association for Computing Machinery.

Bongshin Lee, Eun Kyoung Choe, Petra Isenberg, Kim Marriott, and John Stasko. 2020. Reaching broader audiences with data visualization. *IEEE Computer Graphics and Applications*, 40(2):82–90.

Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. 2015. More than telling a story: Transforming data into visually shared stories. *IEEE computer graphics and applications*, 35(5):84–90.

Bongshin Lee, Arjun Srinivasan, Petra Isenberg, John Stasko, et al. 2021. Post-wimp interaction for information visualization. *Foundations and Trends® in Human-Computer Interaction*, 14(1):1–95.

Hae-Na Lee and Vikas Ashok. 2020. Towards personalized annotation of webpages for efficient screen-reader interaction. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, HT '20, page 111–116, New York, NY, USA. Association for Computing Machinery.

Stephan Lewandowsky and Ian Spence. 1989. The perception of statistical graphs. *Sociological Methods & Research*, 18(2-3):200–242.

Michael Lewis-Beck, Alan E Bryman, and Tim Futing Liao. 2003. *The Sage encyclopedia of social science research methods*. Sage Publications, Thousand Oaks, California.

Peng Liang and Onno De Graaf. 2010. Experiences of using role playing andwiki in requirements engineering course projects. In *2010 5th International Workshop on Requirements Engineering Education and Training*, pages 1–6, New York, NY, USA. IEEE, Institute of Electrical and Electronics Engineers.

Eckhard Limpert, Werner A Stahel, and Markus Abbt. 2001. Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question. *BioScience*, 51(5):341–352.

Ramon C Littell, PR Henry, and Clarence B Ammerman. 1998. Statistical analysis of repeated measures data using sas procedures. *Journal of animal science*, 76(4):1216–1231.

Center for Digital Music Queen Mary University of London. 2023. Sonic visualiser. `https://www.sonicvisualiser.org/`. (Accessed on 07/04/2023).

Michelle Lui. 2012. Designing immersive simulations for collective inquiry. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, page 943–946, New York, NY, USA. Association for Computing Machinery.

Alan Lundgard, Crystal Lee, and Arvind Satyanarayan. 2019. Sociotechnical considerations for accessible visualization design. In *2019 IEEE Visualization Conference (VIS)*, pages 16–20, New York, NY, USA. IEEE, Institute of Electrical and Electronics Engineers.

Alan Lundgard and Arvind Satyanarayan. 2021. Accessible visualization via natural language descriptions: A four-level model of semantic content. *IEEE transactions on visualization and computer graphics*.

Kelly Mack, Edward Cutrell, Bongshin Lee, and Meredith Ringel Morris. 2021a. Designing tools for high-quality alt text authoring. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–14.

Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. 2021b. What do we mean by "accessibility research"? a literature survey of accessibility papers in chi and assets from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA. Association for Computing Machinery.

Moira Maguire and Brid Delahunt. 2017. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education*, 9(3):3351–3364.

Poukong Colin Manikoth. 2016. *Web-Enabled Technologies and Individual User Agency*. University of Idaho.

Yotam Mann. 2020. Tone.js. `https://tonejs.github.io/`. (Accessed on 08/02/2021).

Kim Marriott, Bongshin Lee, Matthew Butler, Ed Cutrell, Kirsten Ellis, Cagatay Goncu, Marti Hearst, Kathleen McCoy, and Danielle Albers Szafir. 2021. Inclusive data visualization for people with disabilities: a call to action. *Interactions*, 28(3):47–51.

Randall B Martin. 1991. The assessment of involvement in role playing. *Journal of clinical psychology*, 47(4):587–596.

Peter McCullagh. 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127.

Tom McEwan and Ben Weerts. 2007. Alt text and basic accessibility.

David K McGookin and Stephen A Brewster. 2006. Soundbar: exploiting multiple views in multimodal graph browsing. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, pages 145–154.

Richard D McKelvey and William Zavoina. 1975. A statistical model for the analysis of ordinal level dependent variables. *Journal of mathematical sociology*, 4(1):103–120.

Verizon Media. 2020. Introducing accessible audio charts - an open source initiative for android apps. `https://developer.yahoo.com/blogs/612790529269366784/`. (Accessed on 07/04/2023).

Arthur G Miller. 1972. Role playing: An alternative to deception? a review of the evidence. *American Psychologist*, 27(7):623.

Silvia Mirri, Silvio Peroni, Paola Salomoni, Fabio Vitali, and Vincenzo Rubano. 2017. Towards accessible graphs in html-based scientific articles. In *2017 14th IEEE Annual Consumer Communications Networking Conference (CCNC)*, pages 1067–1072, Las Vegas, NV, USA. IEEE.

Martez E Mott, John Tang, and Edward Cutrell. 2023. Accessibility of profile pictures: Alt text and beyond to express identity online. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–7, New York, NY, USA. Association for Computing Machinery.

André Natal, Glen Shires, and Philip Jägenstedt. 2020. Web speech api. `https://wicg.github.io/speech-api/`. (Accessed on 08/07/2021).

National Federation of the Blind. 1940. Homepage | national federation of the blind. `https://nfb.org/`. (Accessed on 03/12/2022).

National Geographic Society. 2023. United states regions | national geographic society. `https://www.nationalgeographic.org/maps/united-states-regions/`. (Accessed on 03/29/2022).

United Nations. 2021. Universal declaration of human rights | united nations. `https://www.un.org/en/about-us/universal-declaration-of-human-rights`. (Accessed on 04/08/2024).

Sandra Ohly, Sabine Sonnentag, Cornelia Niessen, and Dieter Zapf. 2010. Diary studies in organizational research. *Journal of Personnel Psychology*.

Keita Ohshiro, Amy Hurst, and Luke DuBois. 2021. Making math graphs more accessible in remote learning: Using sonification to introduce discontinuity in calculus. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–4.

Abiodun Olalere and Jonathan Lazar. 2011. Accessibility of us federal government home pages: Section 508 compliance and site accessibility statements. *Government Information Quarterly*, 28(3):303–309.

Manuel M Oliveira. 2013. Towards more accessible visualizations for color-vision-deficient individuals. *Computing in Science & Engineering*, 15(5):80–87.

Dan R. Olsen. 2007. Evaluating user interface systems research. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, page 251–258, New York, NY, USA. Association for Computing Machinery.

Ricky Onsman. 2023. Making data visualizations accessible - tpgi. `https://www.tpgi.com/making-data-visualizations-accessible/`. (Accessed on 03/08/2024).

Mike Paciello. 2000. *Web accessibility for people with disabilities*. Crc Press, Boca Raton, FL, USA.

Rachael Rickta Patrick and Syahrul Nizam Junaini. 2021. Bibliometric visualisation of computer science and covid-19: A review and proposed method. In *2021 IEEE 19th Student Conference on Research and Development (SCOReD)*, pages 13–18, New York, NY, USA. IEEE, Institute of Electrical and Electronics Engineers.

Michael Quinn Patton. 1990. *Qualitative evaluation and research methods*. SAGE Publications, inc.

Helen Petrie, Fraser Hamilton, Neil King, and Pete Pavan. 2006. Remote usability evaluations with disabled people. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1133–1141, New York, NY, USA. Association for Computing Machinery.

Pablo Picazo-Sanchez, Juan Tapiador, and Gerardo Schneider. 2020. After you, please: browser extensions order attacks and countermeasures. *International Journal of Information Security*, 19(6):623–638.

Azzurra Pini, Jer Hayes, Connor Upton, and Medb Corcoran. 2019. Ai inspired recipes: Designing computationally creative food combos. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6.

Steven Pinker. 1990. A theory of graph comprehension. *Artificial intelligence and the future of testing*, 73:126.

Catherine Plaisant. 2004. The challenge of information visualization evaluation. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, page 109–116, New York, NY, USA. Association for Computing Machinery.

Jonathan Potter and Margaret Wetherell. 1987. *Discourse and social psychology: Beyond attitudes and behaviour*. Sage.

Sarbeswar Praharaj, Patricia Solis, and Elizabeth A Wentz. 2023. Deploying geospatial visualization dashboards to combat the socioeconomic impacts of covid-19. *Environment and Planning B: Urban Analytics and City Science*, 50(5):1262–1279.

Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca A. Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. 2021. Iterative design of sonification techniques to support people with visual impairments in obstacle avoidance. *ACM Trans. Access. Comput.*, 14(4).

Harry T Reis and Shelly L Gable. 2000. Event-sampling and other methods for studying everyday experience.

Benjamin Roussey. 2024. How to create accessible infographics and data visualizations. `https://www.accessibility.com/blog/how-to-create-accessible-infographics-and-data-visualizations`. (Accessed on 03/08/2024).

Tanumon Roy and Lakshmi Boppana. 2022. Interactive web-based image and graph analysis using sonification for the blind. In *2022 IEEE Region 10 Symposium (TENSYMP)*, pages 1–6.

Gery W Ryan, H Russell Bernard, et al. 2000. Data management and analysis methods. *Handbook of qualitative research*, 2(1):769–802.

Bahador Saket, Alex Endert, and Çağatay Demiralp. 2018. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics*, 25(7):2505–2512.

Jibonananda Sanyal, Song Zhang, Gargi Bhattacharya, Phil Amburn, and Robert Moorhead. 2009. A user study to compare four uncertainty visualization methods for 1d and 2d datasets. *IEEE transactions on visualization and computer graphics*, 15(6):1209–1218.

Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. 2016. Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics*, 23(1):341–350.

Nik Sawe, Chris Chafe, and Jeffrey Treviño. 2020. Using data sonification to overcome science literacy, numeracy, and visualization barriers in science communication. *Frontiers in Communication*, 5:46.

Anastasia Schaadhardt, Alexis Hiniker, and Jacob O. Wobbrock. 2021. Understanding blind screen-reader users' experiences of digital artboards. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA. Association for Computing Machinery.

Freedom Scientific. 1995. Jaws® – freedom scientific. `https://www.freedomscientific.com/products/software/jaws/`. (Accessed on 08/08/2021).

Philip Sedgwick and Nan Greenwood. 2015. Understanding the hawthorne effect. *Bmj*, 351.

Ather Sharif, Neha Aitharaju, Srihari N. Krishnaswamy, and Jacob O. Wobbrock. 2024a. Data sonification for screen-reader users: When and when not to use. Under Review.

Ather Sharif, Sanjana S. Chintalapati, Jacob O. Wobbrock, and Katharina Reinecke. 2021. Understanding screen-reader users' experiences with online data visualizations. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '21, New York, NY, USA. Association for Computing Machinery.

Ather Sharif and Babak Forouraghi. 2018. evographs — a jquery plugin to create web accessible graphs. In *2018 15th IEEE Annual Consumer Communications Networking Conference (CCNC)*, pages 1–4, Las Vegas, NV, USA. IEEE.

Ather Sharif, Joo Gyeong Kim, Jessie Z. Xu, and Jacob O. Wobbrock. 2024b. Understanding and reducing the challenges faced by creators of accessible online data visualizations. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '24, New York, NY, USA. Association for Computing Machinery.

Ather Sharif, Aedan Liam McCall, and Kianna Roces Bolante. 2022a. Should i say "disabled people" or "people with disabilities"? language preferences of disabled people between identity-and person-first language. In *Proceedings of the 24th international ACM SIGACCESS conference on computers and accessibility*, pages 1–18.

Ather Sharif, Venkatesh Potluri, Jazz R. X. Ang, Jacob O. Wobbrock, and Jennifer Mankoff. 2024c. Touchpad mapper: Exploring non-visual touchpad interactions for screen-reader users. In *Proceedings of the 21st International Web for All Conference*, page To Appear.

Ather Sharif, Ploypilin Pruekcharoen, Thrisha Ramesh, Ruoxi Shang, Spencer Williams, and Gary Hsieh. 2022b. "what's going on in accessibility research?" frequencies and trends of disability categories and research domains in publications at assets. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '22, New York, NY, USA. Association for Computing Machinery.

Ather Sharif, Aneesha Ramesh, Qianqian Yu, Trung-Anh H Nguyen, and Xuhai Xu. 2023a. Unlockedmaps: A web-based map for visualizing the real-time accessibility of urban rail transit stations. In *Proceedings of the 20th International Web for All Conference*, pages 5–17.

Ather Sharif, Olivia H. Wang, and Alida T. Muongchan. 2022c. "what makes sonification user-friendly?" exploring usability and user-friendliness of sonified responses. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '22, New York, NY, USA. Association for Computing Machinery.

Ather Sharif, Olivia H. Wang, and Alida T. Muongchan. 2022d. "what makes sonification user-friendly?" exploring usability and user-friendliness of sonified responses. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '22, New York, NY, USA. Association for Computing Machinery.

Ather Sharif, Olivia H. Wang, Alida T. Muongchan, Katharina Reinecke, and Jacob O. Wobbrock. 2022e. Sonifier: Javascript library that converts a two-dimensional data into a sonified response. `https://github.com/athersharif/sonifier`. (Accessed on 06/12/2022).

Ather Sharif, Olivia H. Wang, Alida T. Muongchan, Katharina Reinecke, and Jacob O. Wobbrock. 2022f. Voxlens: Making online data visualizations accessible with an interactive javascript plug-in. In *CHI*

Conference on Human Factors in Computing Systems, CHI '22, New York, NY, USA. Association for Computing Machinery.

Ather Sharif, Andrew M. Zhang, Katharina Reinecke, and Jacob O. Wobbrock. 2023b. Understanding and improving drilled-down information extraction from online data visualizations for screen-reader users. In *Proceedings of the 20th International Web for All Conference*, W4A '23, page 18–31, New York, NY, USA. Association for Computing Machinery.

Ather Sharif, Andrew M Zhang, and Jacob O. Wobbrock. 2024d. Increasing agency of screen-reader users in consuming information from online data visualizations. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '24, New York, NY, USA. Association for Computing Machinery.

Ather Sharif, Ruican Zhong, and Yadi Wang. 2023c. Conveying uncertainty in data visualizations to screen-reader users through non-visual means. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '23, New York, NY, USA. Association for Computing Machinery.

Lei Shi, Idan Zelzer, Catherine Feng, and Shiri Azenkot. 2016. Tickers and talker: An accessible labeling toolkit for 3d printed models. In *Proceedings of the 2016 chi conference on human factors in computing systems*, pages 4896–4907.

Kristen Shinohara and Josh Tenenberg. 2007. Observing sara: a case study of a blind person's interactions with technology. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, pages 171–178.

Ben Shneiderman, Catherine Plaisant, and Bradford W Hesse. 2013. Improving healthcare with interactive visualization. *Computer*, 46(5):58–66.

Katie A Siek, Gillian R Hayes, Mark W Newman, and John C Tang. 2014. Field deployments: Knowing from using in context. *Ways of Knowing in HCI*, pages 119–142.

Mandhatya Singh, Muhammad Suhaib Kanroo, Hadia Showkat Kawoosa, and Puneet Goyal. 2023. Towards accessible chart visualizations for the non-visuals: Research, applications and gaps. *Computer Science Review*, 48:100555.

Nikhil Singh, L. Lucy Wang, and Jonathan Bragg. 2024. Figura11y: Ai assistance for writing scientific alt text. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*, IUI '24, page To Appear, New York, NY, USA. Association for Computing Machinery.

Alexa Siu, Gene S-H Kim, Sile O'Modhrain, and Sean Follmer. 2022. Supporting accessible data visualization through audio data narratives. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery.

Alison Smith, Varun Kumar, Jordan Boyd-Graber, Kevin Seppi, and Leah Findlater. 2018. Closing the loop: User-centered design and evaluation of a human-in-the-loop topic modeling system. In *23rd International Conference on Intelligent User Interfaces*, pages 293–304.

Boris Smus. 2013. *Web Audio API: Advanced Sound for Games and Interactive Apps*. O'Reilly Media, California, USA.

Statistical Analysis Software. 2023. Sas graphics accelerator - chrome web store. `https://chrome.google.com/webstore/detail/sas-graphics-accelerator/ockmipfaiiahknplinepcaogdillgoko`. (Accessed on 07/04/2023).

C. Spearman. 1904. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.

Katta Spiel, Oliver L Haimson, and Danielle Lottridge. 2019. How to do better with gender on surveys: a guide for hci researchers. *Interactions*, 26(4):62–65.

Aaron Springer and Henriette Cramer. 2018. " play prblms" identifying and correcting less accessible content in voice interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, New York, NY, USA. Association for Computing Machinery.

Arjun Srinivasan, Tim Harshbarger, Darrell Hilliker, and Jennifer Mankoff. 2023. Azimuth: Designing accessible dashboards for screen reader users. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–16.

Arjun Srinivasan, Nikhila Nyapathy, Bongshin Lee, Steven M. Drucker, and John Stasko. 2021. Collecting and characterizing natural language utterances for specifying data visualizations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA. Association for Computing Machinery.

Constantine Stephanidis, Alex Paramythis, Michael Sfyrakis, A. Stergiou, N. Maou, A. Leventis, G. Paparoulis, and Charalampos Karagiannidis. 1998. Adaptable and adaptive user interfaces for disabled users in the avanti project. In *Proceedings of the 5th International Conference on Intelligence and Services in Networks: Technology for Ubiquitous Telecom Services*, IS&N '98, page 153–166, Berlin, Heidelberg. Springer-Verlag.

Adam Strantz. 2021. Using web standards to design accessible data visualizations in professional communication. *IEEE Transactions on Professional Communication*, 64(3):288–301.

Jacqueline Strecker. 2012. Data visualization in review: summary; evaluating idrc results-communicating research for influence.

Desmos Studios. 2011. Desmos | accessibility. `https://www.desmos.com/accessibility`. (Accessed on 07/04/2023).

Dag Svanaes and Gry Seland. 2004. Putting the users center stage: Role playing and low-fi prototyping enable end users to design mobile systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, page 479–486, New York, NY, USA. Association for Computing Machinery.

Hironobu Takagi, Chieko Asakawa, Kentarou Fukuda, and Junji Maeda. 2003. Accessibility designer: Visualizing usability for the blind. In *Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '04, page 177–184, New York, NY, USA. Association for Computing Machinery.

Jonathan E. Thiele, Michael S. Pratte, and Jeffrey N. Rouder. 2011. On perfect working-memory performance with large numbers of items. *Psychonomic Bulletin & Review*, 18(5):958–963.

John R Thompson, Jesse J Martinez, Alper Sarikaya, Edward Cutrell, and Bongshin Lee. 2023. Chart reader: Accessible visualization experiences designed with screen reader users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA. Association for Computing Machinery.

Ingo R. Titze and Daniel W. Martin. 1998. Principles of voice production.

Heidi Toivonen and Francesco Lelli. 2024. The varieties of agency in human–smart device relationships: The four agency profiles. *Future Internet*, 16(3):90.

Márcio Josué Ramos Torres and Regina Barwaldt. 2019. Approaches for diagrams accessibility for blind people: a systematic review. In *2019 IEEE Frontiers in Education Conference (FIE)*, pages 1–7, New York, NY, USA. Institute of Electrical and Electronics Engineers.

J Gregory Trafton, Susan S Kirschenbaum, Ted L Tsui, Robert T Miyamoto, James A Ballas, and Paula D Raymond. 2000. Turning pictures into numbers: extracting and generating information from complex visualizations. *International Journal of Human-Computer Studies*, 53(5):827–850.

Shari Trewin, Brian Cragun, Cal Swart, Jonathan Brezin, and John Richards. 2010. Accessibility challenges and tool features: An ibm web developer perspective. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, W4A '10, New York, NY, USA. Association for Computing Machinery.

Suppawong Tuarob, Poom Wettayakorn, Ponpat Phetchai, Siripong Traivijitkhun, Sunghoon Lim, Thanapon Noraset, and Tipajin Thaipisutikul. 2021. Davis: a unified solution for data collection, analyzation, and visualization in real-time stock market prediction. *Financial Innovation*, 7(1):1–32.

Michael B Twidale. 2005. Over the shoulder learning: supporting brief informal learning. *Computer supported cooperative work (CSCW)*, 14(6):505–547.

The Ohio State University. 2022. Alternative (alt) text guide | engineering technology services (ets). `https://ets.osu.edu/digital-accessibility/alternative-alt-text-guide`. (Accessed on 04/10/2024).

Frances Van Scoy, Don McLaughlin, and Angela Fullmer. 2005. Auditory augmentation of haptic graphs: Developing a graphic tool for teaching precalculus skill to blind students. In *Proceedings of the 11th Meeting of the International Conference on Auditory Display*, volume 5.

Tamás Vörös, Sean Paul Bergeron, and Konstantin Berlin. 2023. Web content filtering through knowledge distillation of large language models. In *2023 IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 357–361. IEEE.

W3C. 2006. Wai-aria overview | web accessibility initiative (wai) | w3c. `https://www.w3.org/WAI/standards-guidelines/aria/`. (Accessed on 04/11/2021).

W3C. 2021. Web content accessibility guidelines (wcag) 2.1. `https://www.w3.org/TR/WCAG21/`. (Accessed on 11/14/2022).

Jakub Wabiński, Albina Mościcka, and Marta Kuźma. 2021. The information value of tactile maps: A comparison of maps printed with the use of different techniques. *The Cartographic Journal*, 58(2):123–134.

Bruce N Walker and Michael A Nees. 2011. Theory of sonification. *The sonification handbook*, 1:9–39.

Jennifer Wall et al. 2012. Nasa-helping the blind to see science and math.

Ruobin Wang, Crescentia Jung, and Yea-Seul Kim. 2022a. Seeing through sounds: Mapping auditory dimensions to data and charts for people with visual impairments. In *The 24th EG/VGTC Conference on Visualization (EuroVis' 22), Rome, Italy, 13-17 June, 2022*. Eurographics-European Association for Computer Graphics.

Yanan Wang, Ruobin Wang, Crescentia Jung, and Yea-Seul Kim. 2022b. What makes web data tables accessible? insights and a tool for rendering accessible tables for people with visual impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–20.

Web Accessibility In Mind. 1999. Webaim: Contrast checker. `https://webaim.org/resources/contrastchecker/`. (Accessed on 07/09/2023).

WebAIM. 1999. Webaim: Css in action - invisible content just for screen reader users. `https://webaim.org/techniques/css/invisiblecontent/`. (Accessed on 09/01/2021).

WebAIM. 2016. Wave web accessibility evaluation tools. `https://wave.webaim.org/`. (Accessed on 08/21/2023).

Xiang Wei, Xingyu Cui, Ning Cheng, Xiaobin Wang, Xin Zhang, Shen Huang, Pengjun Xie, Jinan Xu, Yufeng Chen, Meishan Zhang, et al. 2023. Zero-shot information extraction via chatting with chatgpt. *arXiv preprint arXiv:2302.10205*.

Daniel Weiskopf. 2022. Uncertainty visualization: Concepts, methods, and applications in biological data visualization. *Frontiers in Bioinformatics*, 2:10.

Sue Widdicombe and Robin Wooffitt. 1995. *The language of youth subcultures: Social identity in action*. Harvester/Wheatsheaf.

Frank Wilcoxon. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.

Christopher Winship and Robert D Mare. 1984. Regression models with ordinal variables. *American sociological review*, pages 512–525.

Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011a. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 143–146, New York, NY, USA. Association for Computing Machinery.

Jacob O. Wobbrock, Krzysztof Z. Gajos, Shaun K. Kane, and Gregg C. Vanderheiden. 2018. Ability-based design. *Communications of the ACM*, 61(6):62–71.

Jacob O Wobbrock, Shaun K Kane, Krzysztof Z Gajos, Susumu Harada, and Jon Froehlich. 2011b. Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing (TACCESS)*, 3(3):1–27.

Jacob O Wobbrock and Julie A Kientz. 2016. Research contributions in human-computer interaction. *interactions*, 23(3):38–44.

Robin Wooffitt and Sue Widdicombe. 1995. *The language of youth subcultures: social identity in action*. Harvester Wheatsheaf, Birmingham, UK.

Keke Wu, Emma Petersen, Tahmina Ahmad, David Burlinson, Shea Tanis, and Danielle Albers Szafir. 2021. Understanding data accessibility for people with intellectual and developmental disabilities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA. Association for Computing Machinery.

Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In *proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 1180–1192.

Susan P Wyche and Rebecca E Grinter. 2009. Extraordinary computing: religion as a lens for reconsidering the home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 749–758.

Wai Yu and Stephen Brewster. 2002. Multimodal virtual reality versus printed medium in visualization for blind people. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies*, Assets '02, page 57–64, New York, NY, USA. Association for Computing Machinery.

Wai Yu, Ramesh Ramloll, and Stephen Brewster. 2000. Haptic graphs for blind computer users. In *International Workshop on Haptic Human-Computer Interaction*, pages 41–51. Springer.

Lotfi A Zadeh. 1988. Fuzzy logic. *Computer*, 21(4):83–93.

Limin Zeng, Gerhard Weber, et al. 2011. Accessible maps for the visually impaired. In *Proceedings of IFIP INTERACT 2011 Workshop on ADDW, CEUR*, volume 792, pages 54–60.

Dongsong Zhang, Lina Zhou, Judith O. Uchidiuno, and Isil Y. Kilic. 2017. Personalized assistive web for improving mobile web browsing and accessibility for visually impaired users. *ACM Trans. Access. Comput.*, 10(2).

Lidong Zhang, B Vinodhini, and T Maragatham. 2021. Interactive iot data visualization for decision making in business intelligence. *Arabian Journal for Science and Engineering*, Special Issue:1–11.

Mingrui Ray Zhang, Mingyuan Zhong, and Jacob O. Wobbrock. 2022. Ga11y: An automated gif annotation system for visually impaired users. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery.

Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. 2008. Data sonification for users with visual impairment: a case study with georeferenced data. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15(1):1–28.

Luming Zhao and WeiMing Ye. 2022. Visualization as infrastructure: China's data visualization politics during covid-19 and their implications for public health emergencies. *Convergence*, 28(1):13548565211069872.

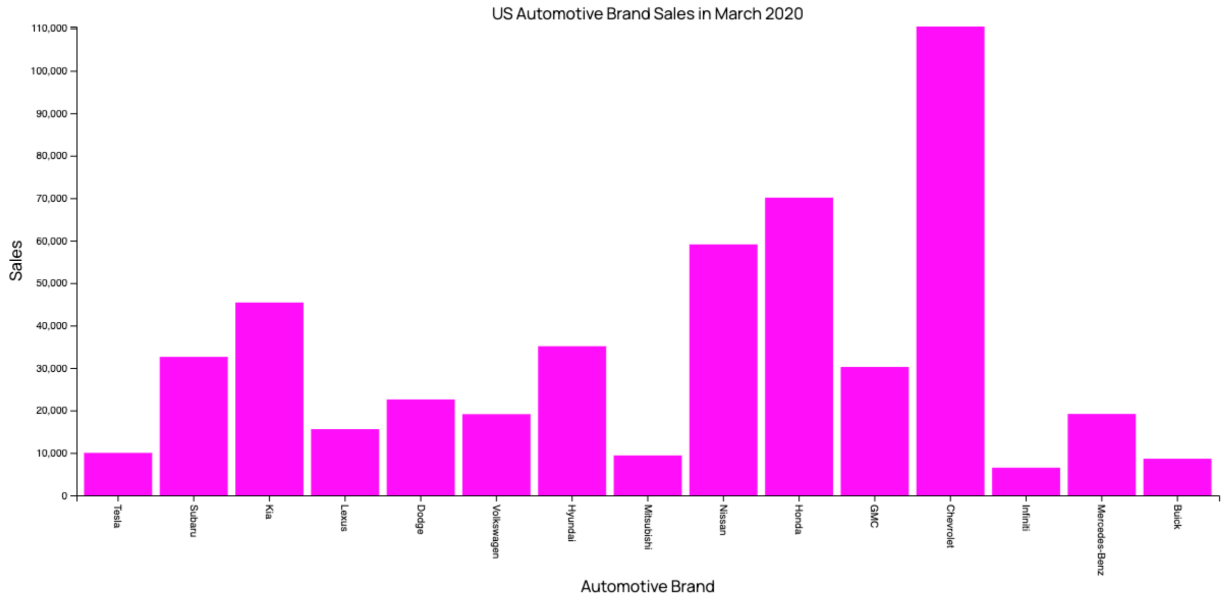
Shuai Zheng, Jonathan R Edwards, Margaret A Dudeck, Prachi R Patel, Lauren Wattenmaker, Muzna Mirza, Sheri Chernetsky Tejedor, Kent Lemoine, Andrea L Benin, Daniel A Pollock, et al. 2021. Building an interactive geospatial visualization application for national health care–associated infection surveillance: Development study. *JMIR public health and surveillance*, 7(7):e23528.

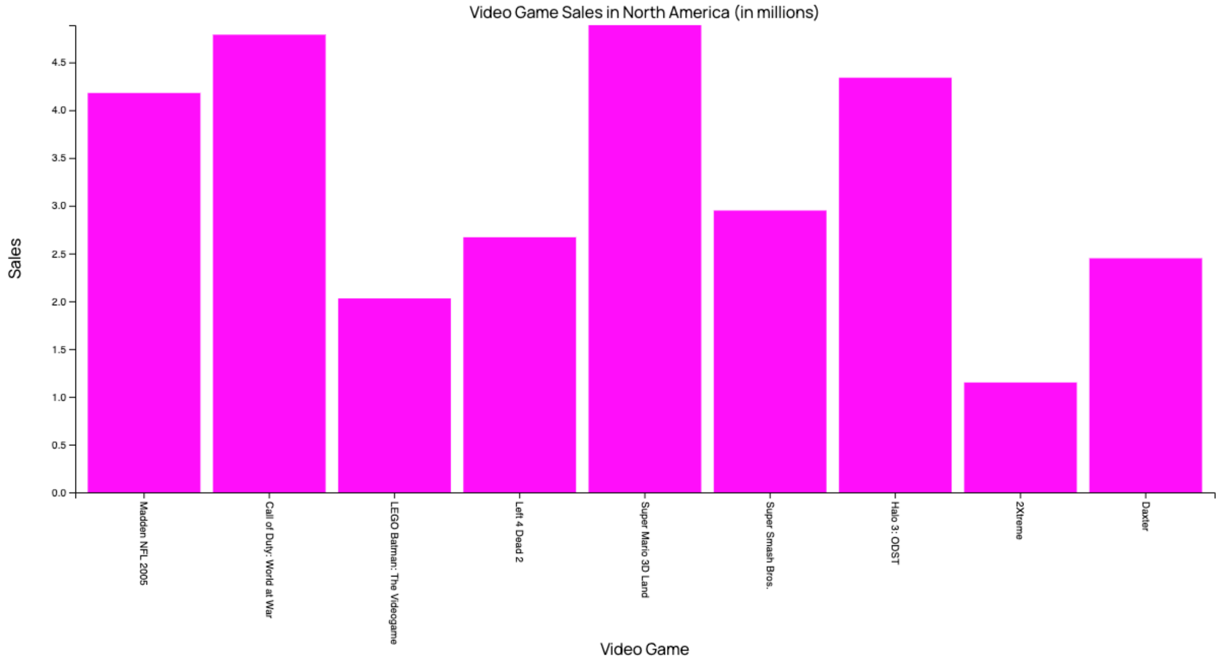
Jonathan Zong, Crystal Lee, Alan Lundgard, JiWoong Jang, Daniel Hajas, and Arvind Satyanarayan. 2022. Rich screen reader experiences for accessible data visualization. *Computer Graphics Forum*, 41(3):15–27.

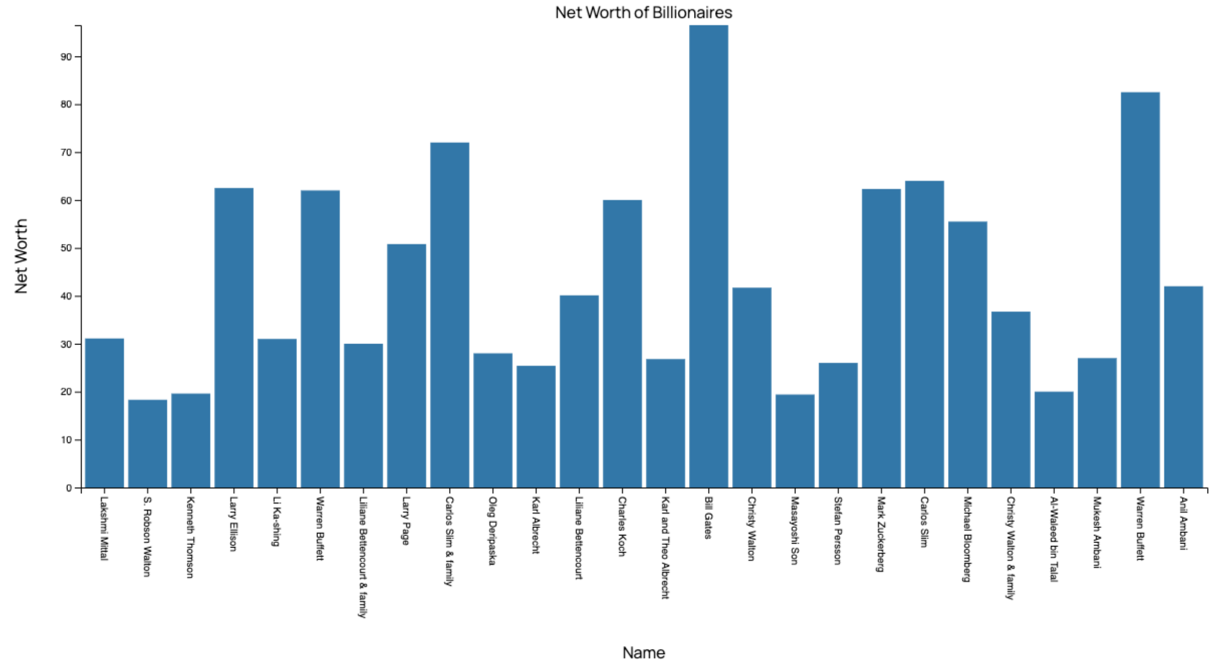
Appendix A

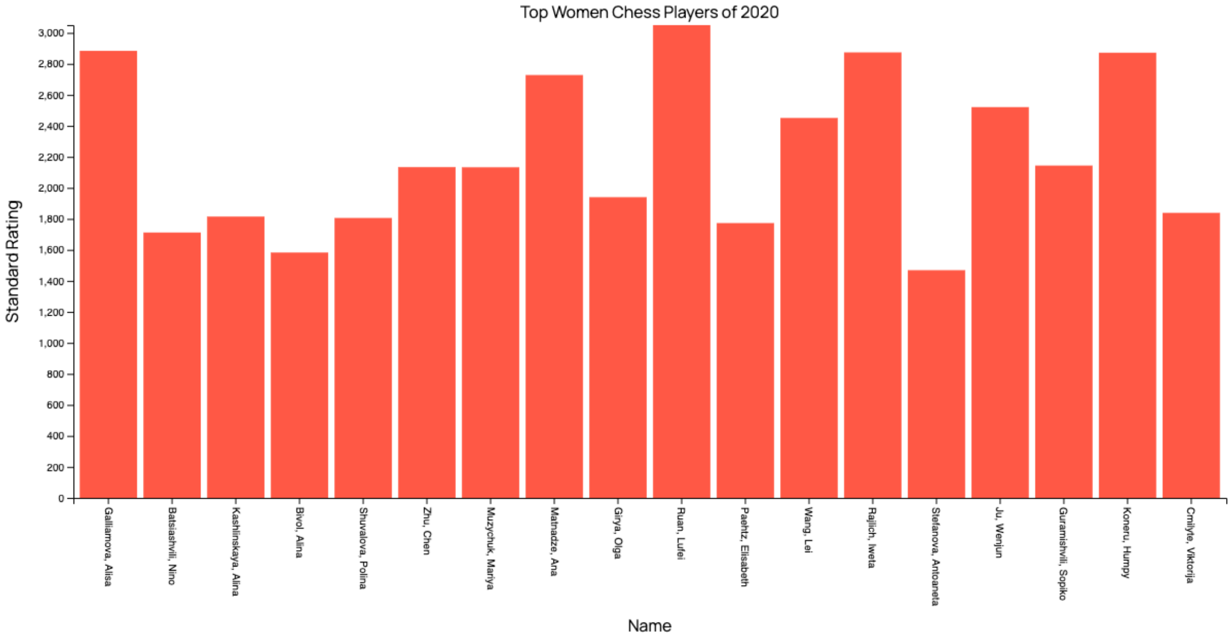
Bar Chart Stimuli

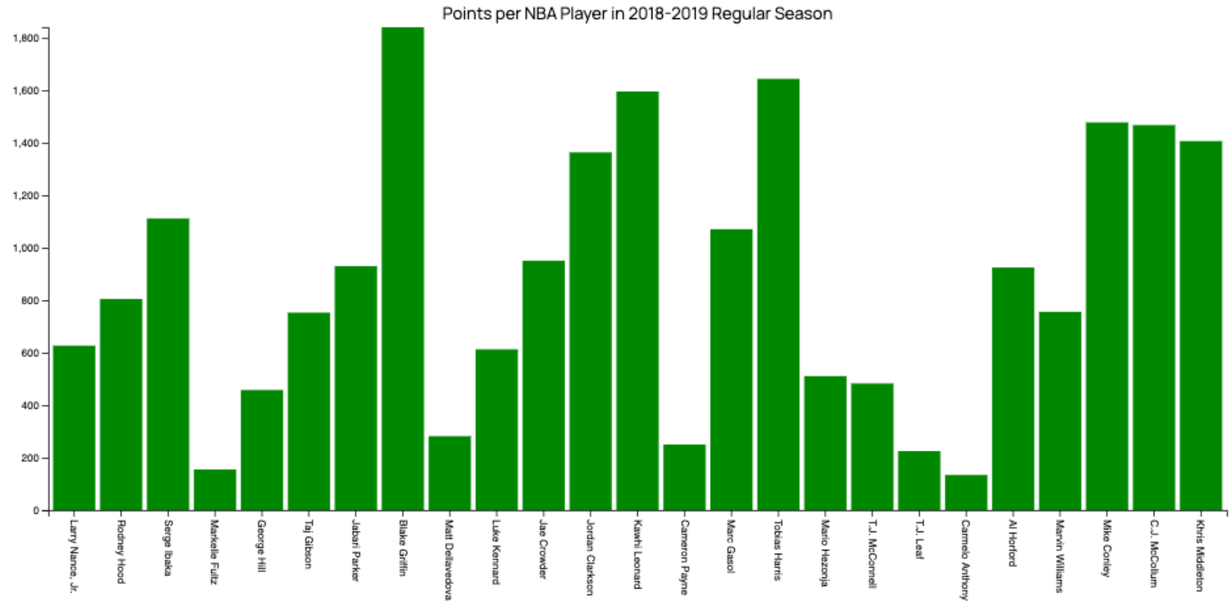
This appendix shows the bar chart stimuli used during the task-based experiment described in Chapter 3.

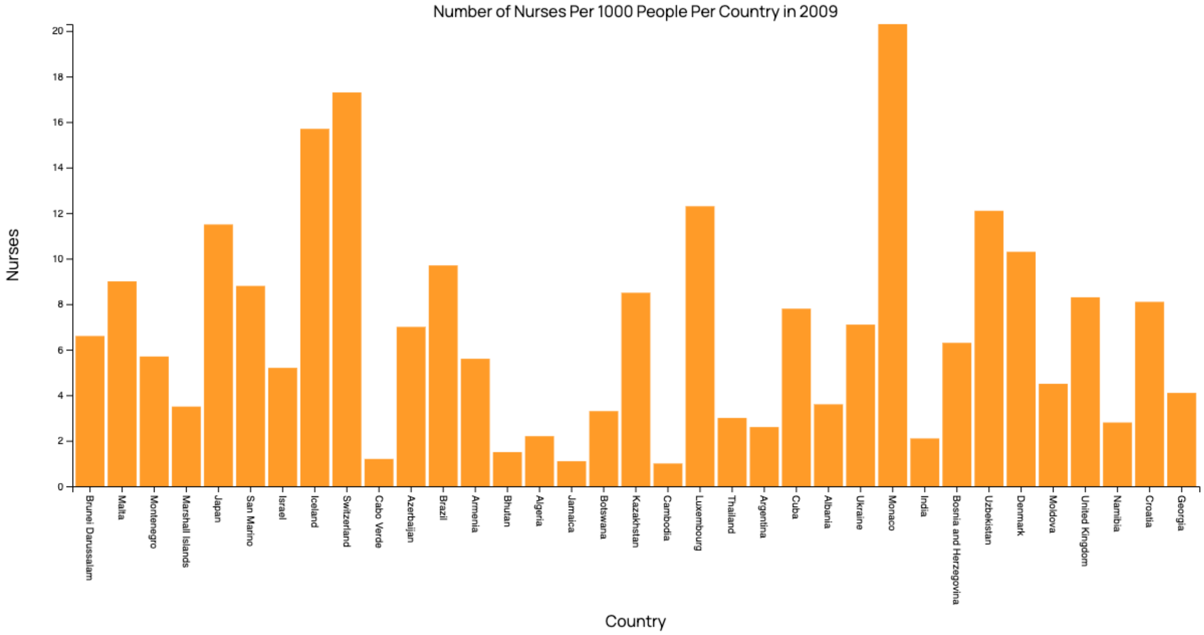

US Automotive Brand Sales in March 2020
Sales
110,000
100,000
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
Tesla
Subaru
Kia
Lexus
Dodge
Volkswagen
Hyundai
Mitsubishi
Nissan
Honda
GMC
Chevrolet
Infiniti
Mercedes-Benz
Buick
Automotive Brand

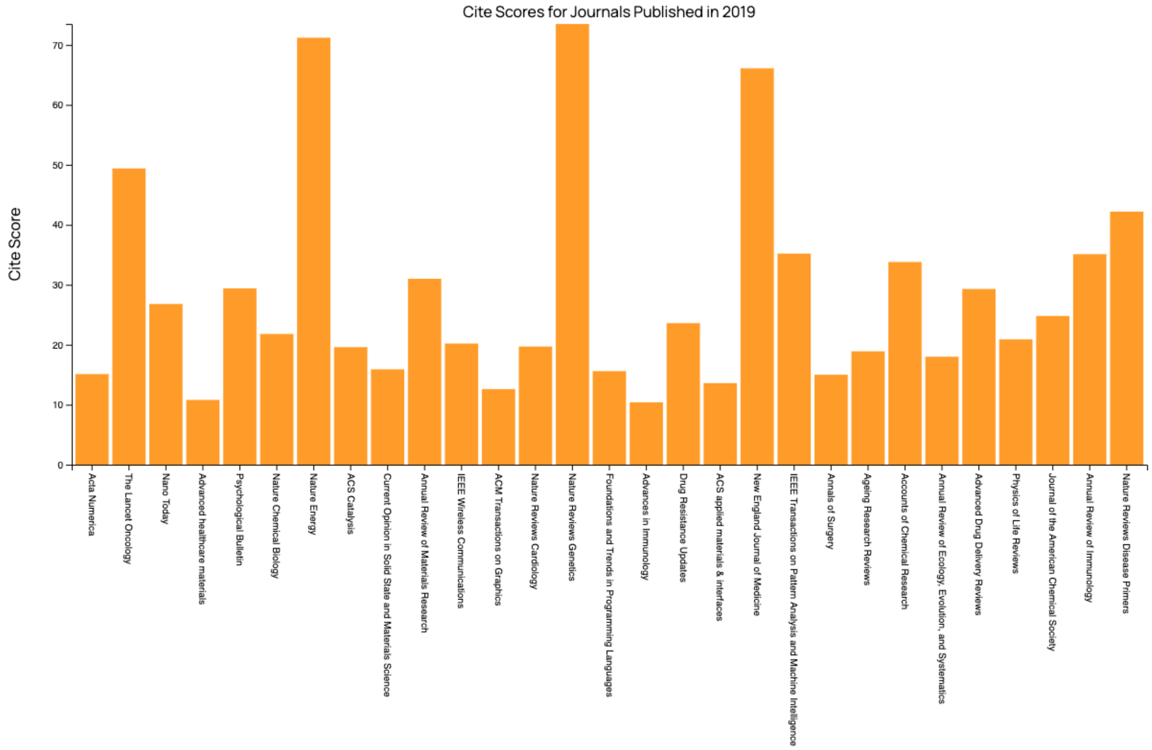

Video Game Sales in North America (in millions)
Sales
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Madden NFL 2005
Call of Duty: World at War
LEGO Batman: The Videogame
Left 4 Dead 2
Super Mario 3D Land
Super Smash Bros.
Halo 3: ODST
2Xtreme
Daxter
Video Game

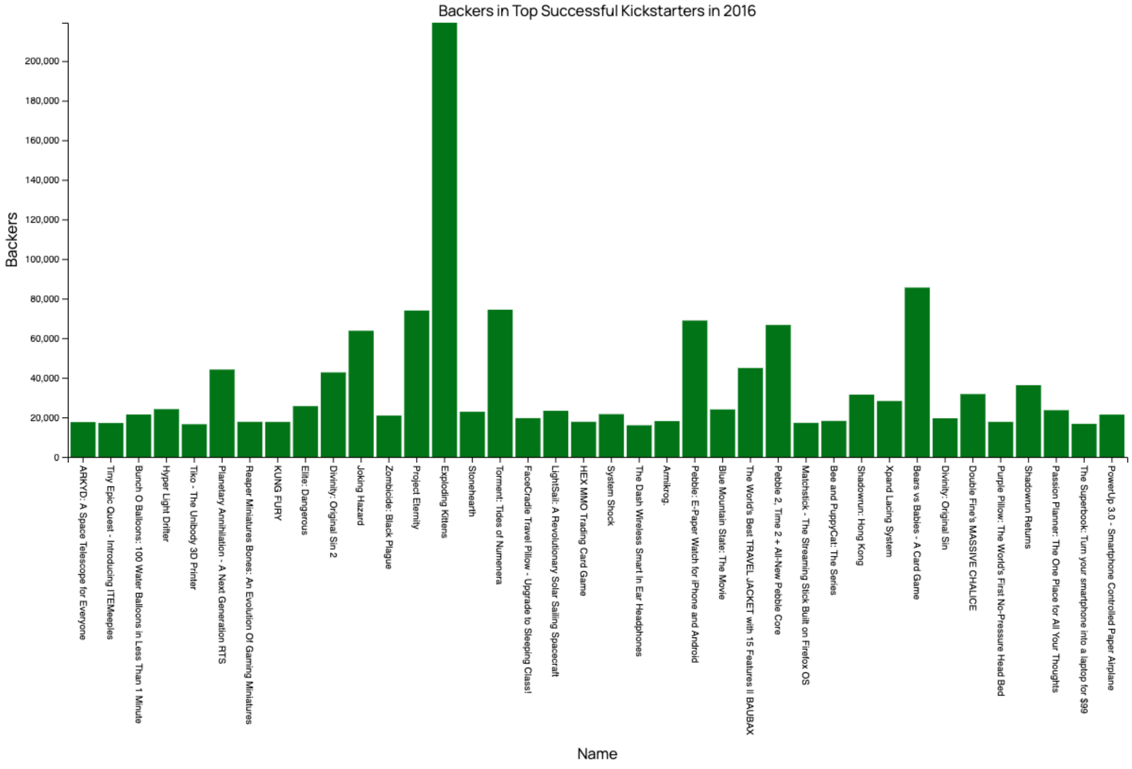

Net Worth of Billionaires
Net Worth
0
10
20
30
40
50
60
70
80
90
Lakshmi Mittal
S. Robson Walton
Kenneth Thomson
Larry Ellison
Li Ka-shing
Warren Buffett
Liliane Bettencourt & family
Larry Page
Carlos Slim & family
Oleg Deripaska
Karl Albrecht
Liliane Bettencourt
Charles Koch
Karl and Theo Albrecht
Bill Gates
Christy Walton
Masayoshi Son
Stefan Persson
Mark Zuckerberg
Carlos Slim
Michael Bloomberg
Christy Walton & family
Al-Waleed bin Talal
Mukesh Ambani
Warren Buffett
Anil Ambani
Name


Top Women Chess Players of 2020
Standard Rating
0
200
400
600
800
1,000
1,200
1,400
1,600
1,800
2,000
2,200
2,400
2,600
2,800
3,000
Galliamova, Alisa
Batsiashvili, Nino
Kashlinskaya, Alina
Bivol, Alina
Shuvalova, Polina
Zhu, Chen
Muzychuk, Mariya
Matnadze, Ana
Girya, Olga
Ruan, Lufei
Paehtz, Elisabeth
Wang, Lei
Rajlich, Iweta
Stefanova, Antoaneta
Ju, Wenjun
Guramishvili, Sopiko
Koneru, Humpy
Cmilyte, Viktorija
Name


Points per NBA Player in 2018-2019 Regular Season


Number of Nurses Per 1000 People Per Country in 2009


Cite Scores for Journals Published in 2019
Cite Score
0
10
20
30
40
50
60
70
Source Title
Acta Numerica
The Lancet Oncology
Nano Today
Advanced healthcare materials
Psychological Bulletin
Nature Chemical Biology
Nature Energy
ACS Catalysis
Current Opinion in Solid State and Materials Science
Annual Review of Materials Research
IEEE Wireless Communications
ACM Transactions on Graphics
Nature Reviews Cardiology
Nature Reviews Genetics
Foundations and Trends in Programming Languages
Advances in Immunology
Drug Resistance Updates
ACS applied materials & interfaces
New England Journal of Medicine
IEEE Transactions on Pattern Analysis and Machine Intelligence
Annals of Surgery
Ageing Research Reviews
Accounts of Chemical Research
Annual Review of Ecology, Evolution, and Systematics
Advanced Drug Delivery Reviews
Physics of Life Reviews
Journal of the American Chemical Society
Annual Review of Immunology
Nature Reviews Disease Primers


Backers in Top Successful Kickstarters in 2016
Backers
0
20,000
40,000
60,000
80,000
100,000
120,000
140,000
160,000
180,000
200,000
Name
ARKYD: A Space Telescope for Everyone
Tiny Epic Quest - Introducing ITEMeeples
Bunch O Balloons: 100 Water Balloons in Less Than 1 Minute
Hyper Light Drifter
Tiko - The Unibody 3D Printer
Planetary Annihilation - A Next Generation RTS
Reaper Miniatures Bones: An Evolution Of Gaming Miniatures
KUNG FURY
Elite: Dangerous
Divinity: Original Sin 2
Joking Hazard
Zombicide: Black Plague
Project Eternity
Exploding Kittens
Stonehearth
Torment: Tides of Numenera
FaceCradle Travel Pillow - Upgrade to Sleeping Class!
LightSail: A Revolutionary Solar Sailing Spacecraft
HEX MMO Trading Card Game
System Shock
The Dash Wireless Smart In Ear Headphones
Amiibog.
Pebble: E-Paper Watch for iPhone and Android
Blue Mountain State: The Movie
The World's Best TRAVEL JACKET with 15 Features ll BAUBAX
Pebble 2, Time 2 + All-New Pebble Core
Matchstick - The Streaming Stick Built on Firefox OS
Bee and PuppyCat: The Series
Shadowrun: Hong Kong
Xpand Lacing System
Bears vs Babies - A Card Game
Divinity: Original Sin
Double Fine's MASSIVE CHALICE
Purple Pillow: The World's First No-Pressure Head Bed
Shadowrun Returns
Passion Planner: The One Place for All Your Thoughts
The Superbook: Turn your smartphone into a laptop for $99
PowerUp 3.0 - Smartphone Controlled Paper Airplane

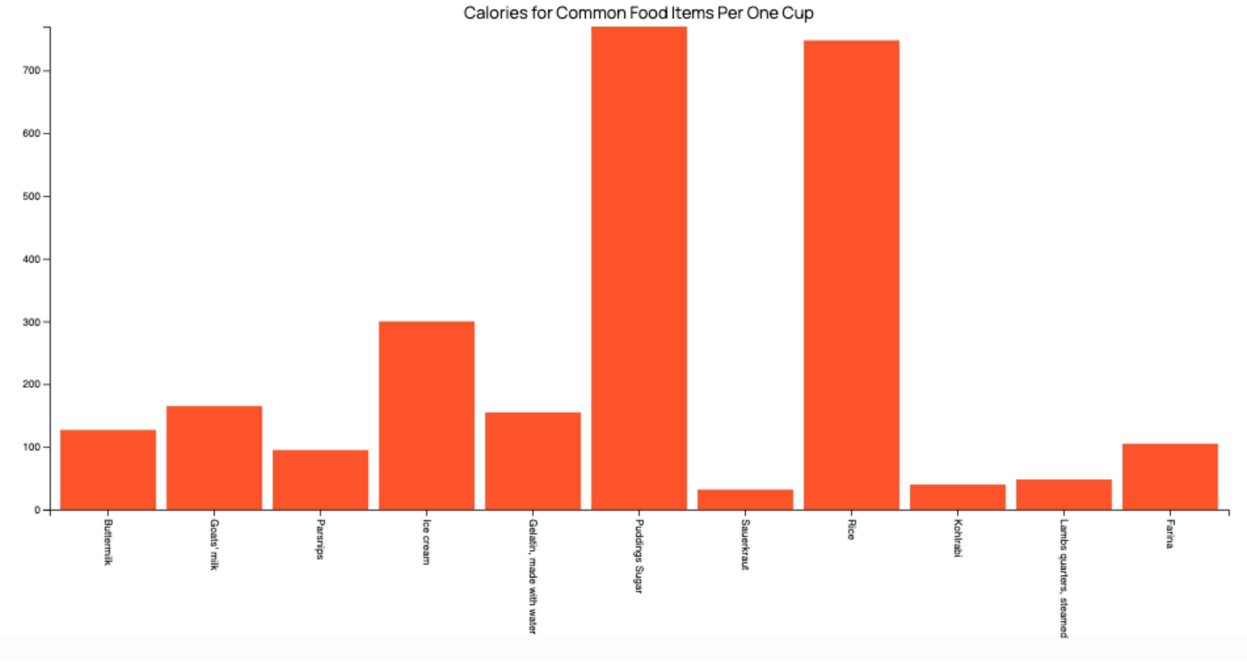
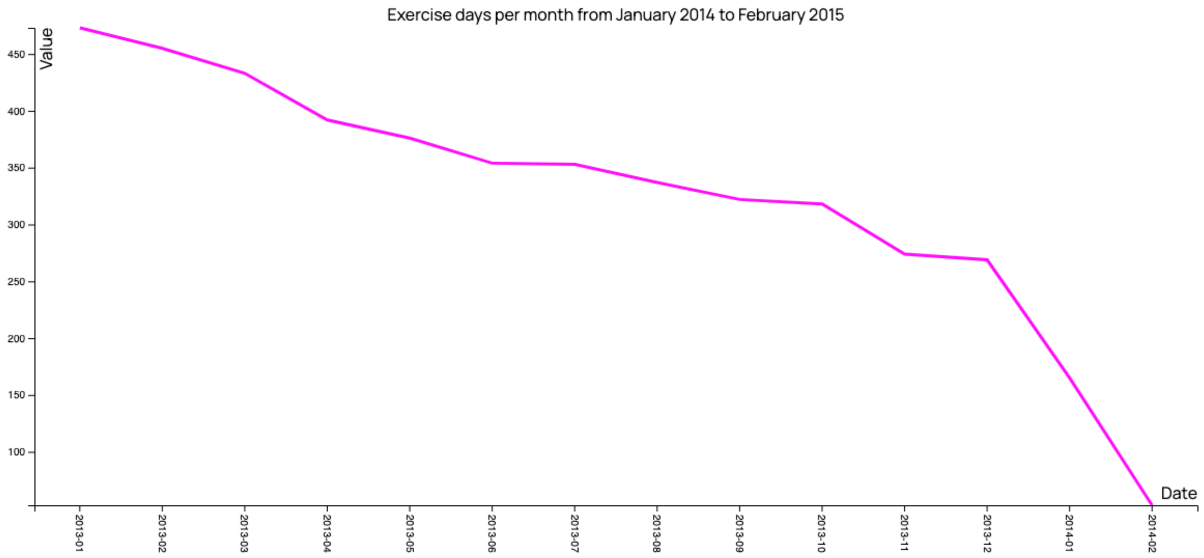


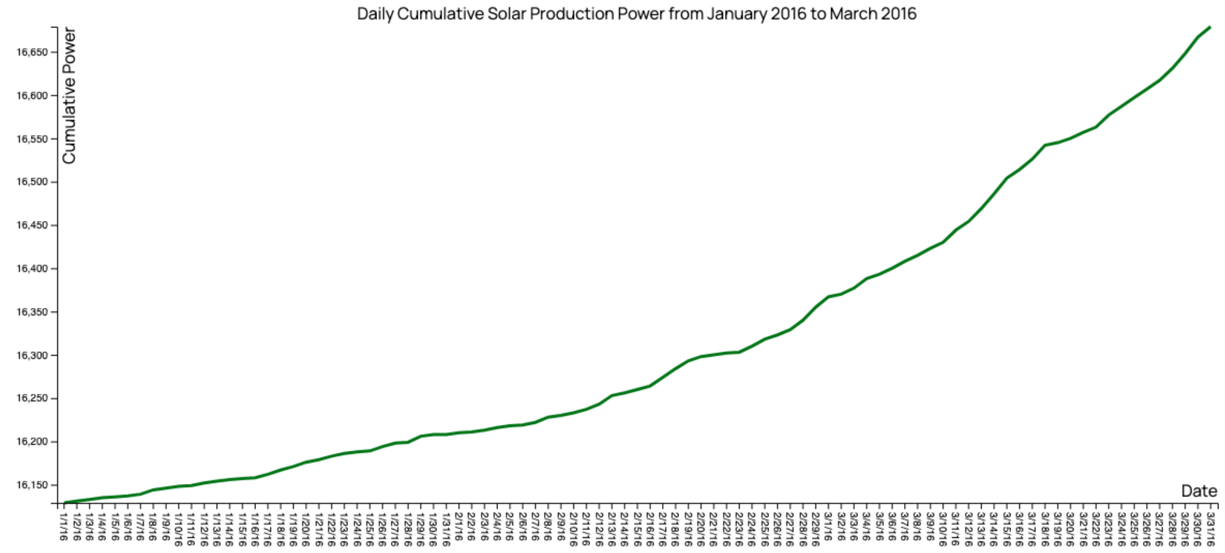
Figure A.1: Bar charts shown during the task-based experiment to measure accuracy of extracted information and interaction times. These visualizations are created using D3 (Bostock et al. [2011]).

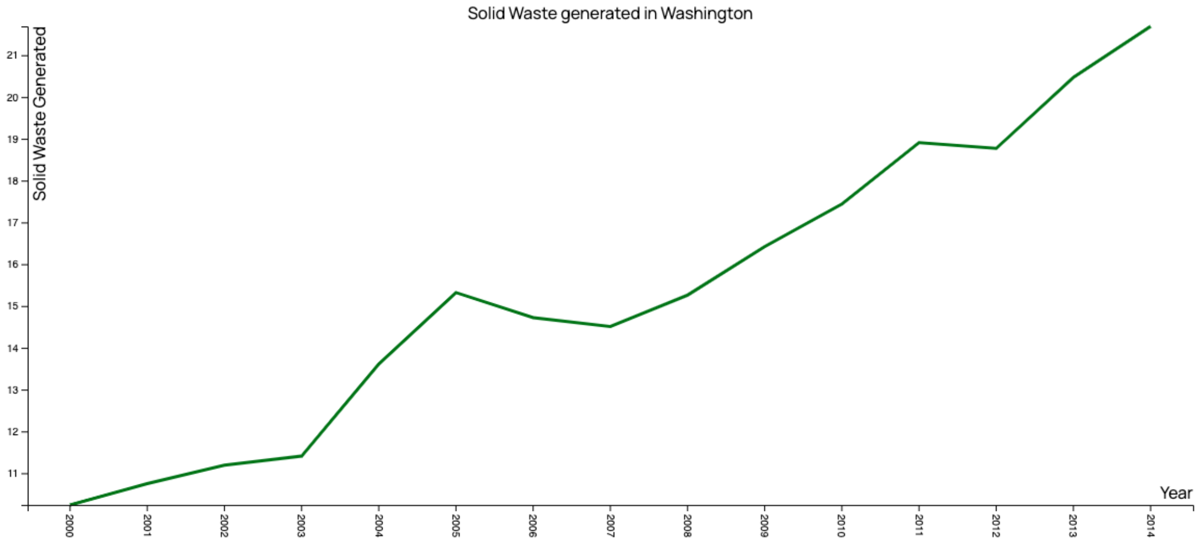
Appendix B

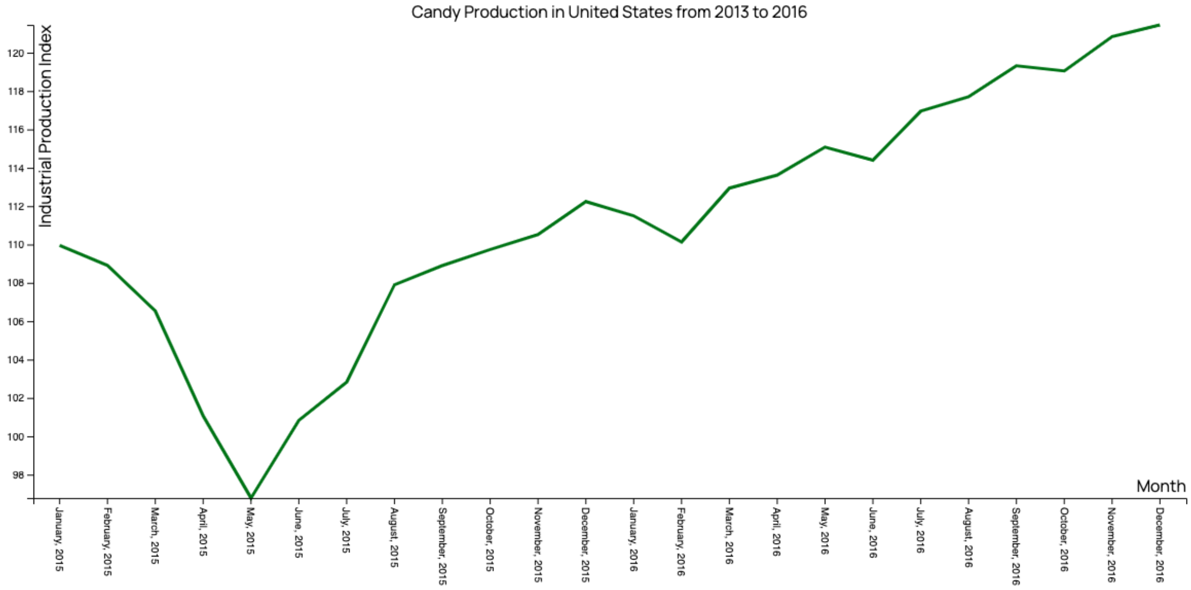
Line Chart Stimuli

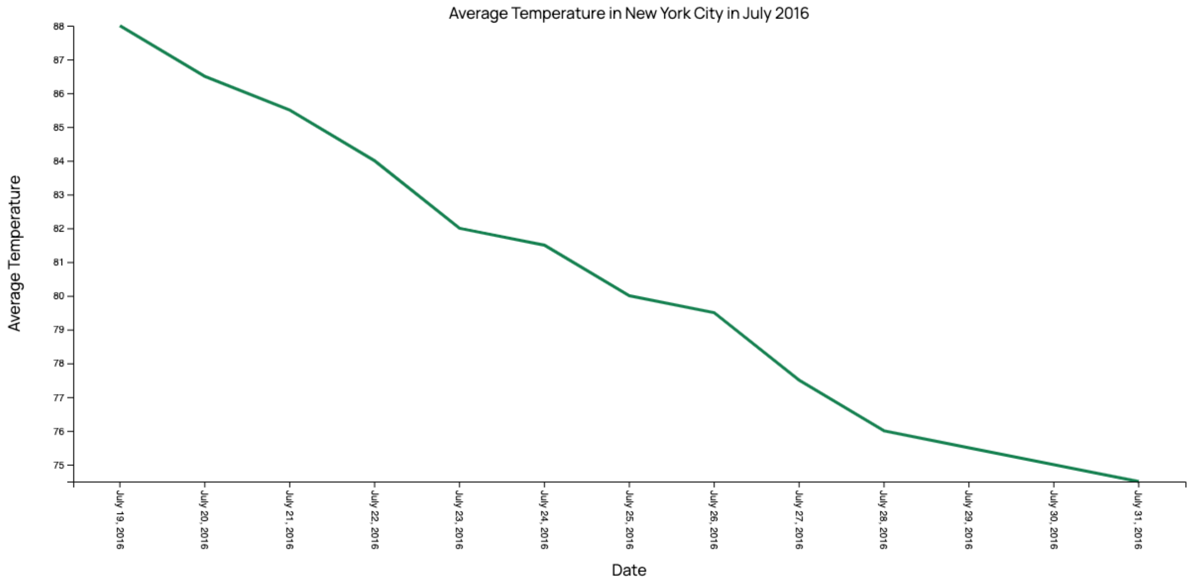
This appendix shows the line chart stimuli used during the task-based experiment described in Chapter 3.

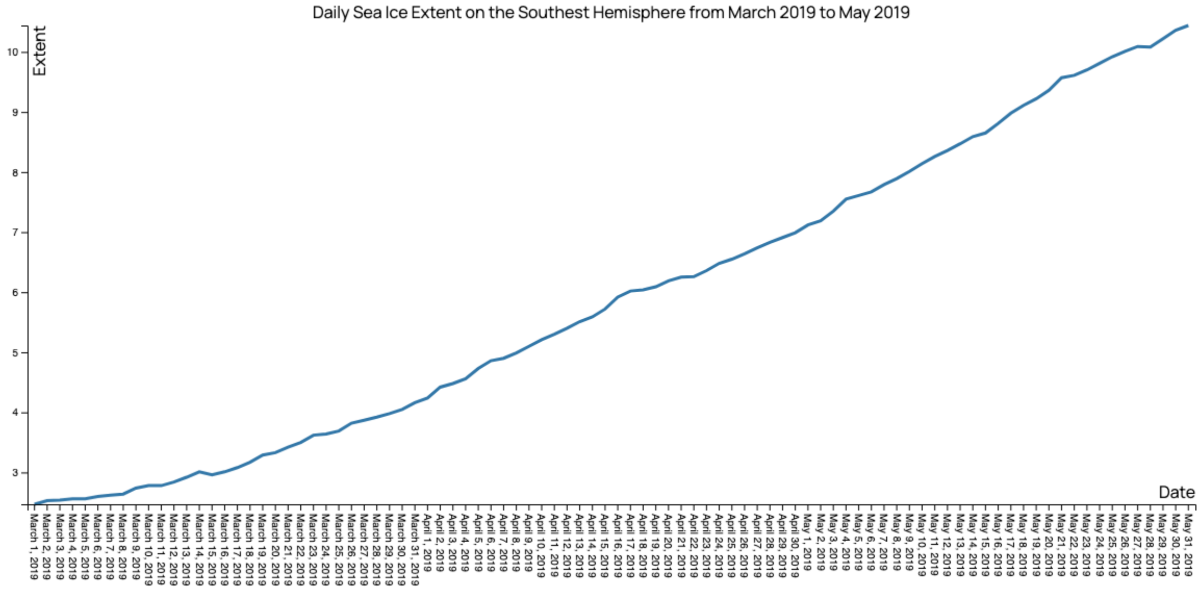

Exercise days per month from January 2014 to February 2015
Value
450
400
350
300
250
200
150
100
Date
2013-01
2013-02
2013-03
2013-04
2013-05
2013-06
2013-07
2013-08
2013-09
2013-10
2013-11
2013-12
2014-01
2014-02

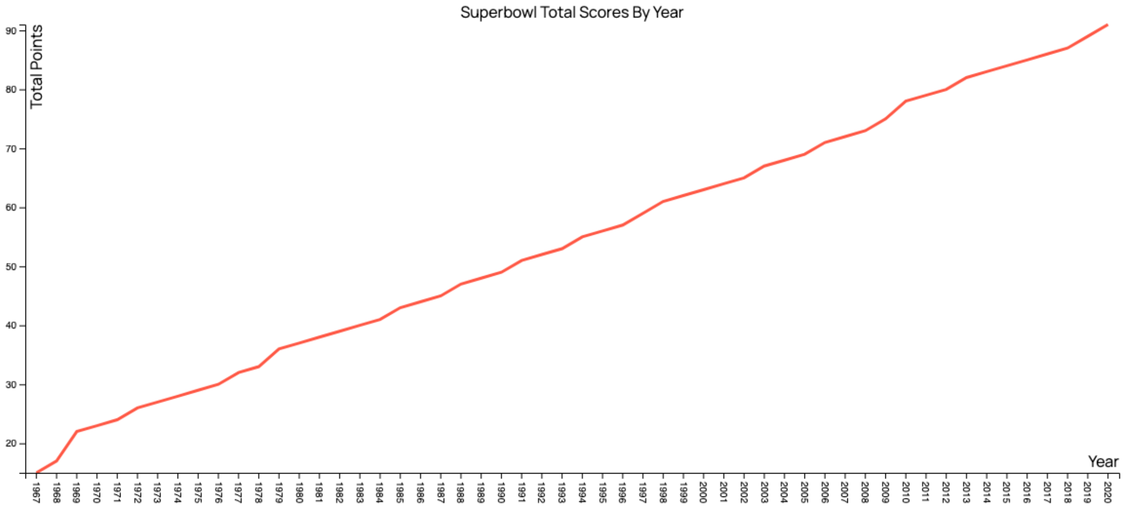

Daily Cumulative Solar Production Power from January 2016 to March 2016
Cumulative Power
16,650
16,600
16,550
16,500
16,450
16,400
16,350
16,300
16,250
16,200
16,150
Date
1/1/16
1/2/16
1/3/16
1/4/16
1/5/16
1/6/16
1/7/16
1/8/16
1/9/16
1/10/16
1/11/16
1/12/16
1/13/16
1/14/16
1/15/16
1/16/16
1/17/16
1/18/16
1/19/16
1/20/16
1/21/16
1/22/16
1/23/16
1/24/16
1/25/16
1/26/16
1/27/16
1/28/16
1/29/16
1/30/16
1/31/16
2/1/16
2/2/16
2/3/16
2/4/16
2/5/16
2/6/16
2/7/16
2/8/16
2/9/16
2/10/16
2/11/16
2/12/16
2/13/16
2/14/16
2/15/16
2/16/16
2/17/16
2/18/16
2/19/16
2/20/16
2/21/16
2/22/16
2/23/16
2/24/16
2/25/16
2/26/16
2/27/16
2/28/16
2/29/16
3/1/16
3/2/16
3/3/16
3/4/16
3/5/16
3/6/16
3/7/16
3/8/16
3/9/16
3/10/16
3/11/16
3/12/16
3/13/16
3/14/16
3/15/16
3/16/16
3/17/16
3/18/16
3/19/16
3/20/16
3/21/16
3/22/16
3/23/16
3/24/16
3/25/16
3/26/16
3/27/16
3/28/16
3/29/16
3/30/16
3/31/16

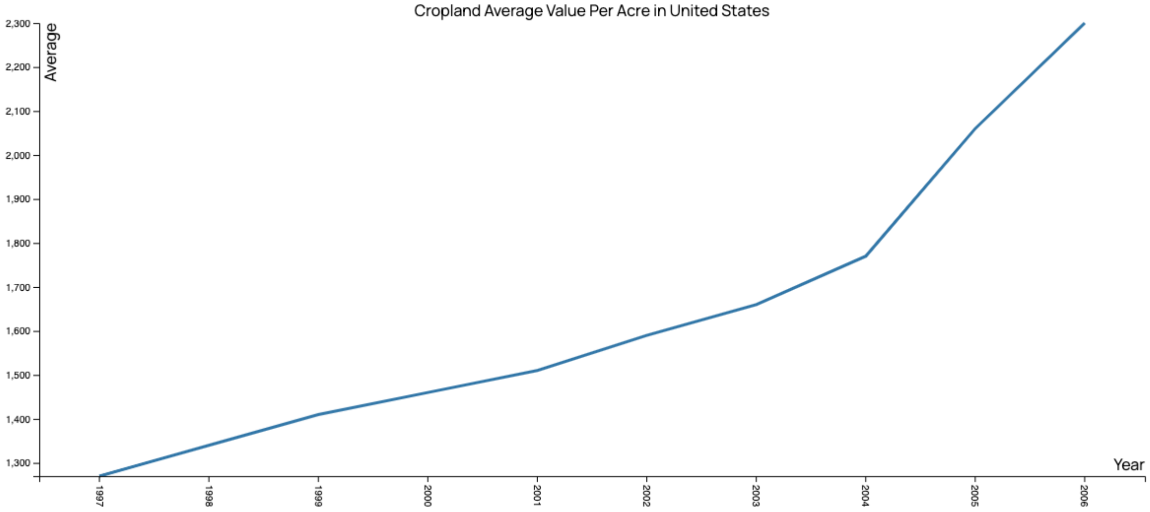

Solid Waste generated in Washington
Solid Waste Generated
21
20
19
18
17
16
15
14
13
12
11
Year
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014


Candy Production in United States from 2013 to 2016
Industrial Production Index
120
118
116
114
112
110
108
106
104
102
100
98
Month
January, 2015
February, 2015
March, 2015
April, 2015
May, 2015
June, 2015
July, 2015
August, 2015
September, 2015
October, 2015
November, 2015
December, 2015
January, 2016
February, 2016
March, 2016
April, 2016
May, 2016
June, 2016
July, 2016
August, 2016
September, 2016
October, 2016
November, 2016
December, 2016


Average Temperature in New York City in July 2016
Average Temperature
88
87
86
85
84
83
82
81
80
79
78
77
76
75
July 19, 2016
July 20, 2016
July 21, 2016
July 22, 2016
July 23, 2016
July 24, 2016
July 25, 2016
July 26, 2016
July 27, 2016
July 28, 2016
July 29, 2016
July 30, 2016
July 31, 2016
Date


Daily Sea Ice Extent on the Southest Hemisphere from March 2019 to May 2019
Extent
10
9
8
7
6
5
4
3
Date
March 1, 2019
March 2, 2019
March 3, 2019
March 4, 2019
March 5, 2019
March 6, 2019
March 7, 2019
March 8, 2019
March 9, 2019
March 10, 2019
March 11, 2019
March 12, 2019
March 13, 2019
March 14, 2019
March 15, 2019
March 16, 2019
March 17, 2019
March 18, 2019
March 19, 2019
March 20, 2019
March 21, 2019
March 22, 2019
March 23, 2019
March 24, 2019
March 25, 2019
March 26, 2019
March 27, 2019
March 28, 2019
March 29, 2019
March 30, 2019
March 31, 2019
April 1, 2019
April 2, 2019
April 3, 2019
April 4, 2019
April 5, 2019
April 6, 2019
April 7, 2019
April 8, 2019
April 9, 2019
April 10, 2019
April 11, 2019
April 12, 2019
April 13, 2019
April 14, 2019
April 15, 2019
April 16, 2019
April 17, 2019
April 18, 2019
April 19, 2019
April 20, 2019
April 21, 2019
April 22, 2019
April 23, 2019
April 24, 2019
April 25, 2019
April 26, 2019
April 27, 2019
April 28, 2019
April 29, 2019
April 30, 2019
May 1, 2019
May 2, 2019
May 3, 2019
May 4, 2019
May 5, 2019
May 6, 2019
May 7, 2019
May 8, 2019
May 9, 2019
May 10, 2019
May 11, 2019
May 12, 2019
May 13, 2019
May 14, 2019
May 15, 2019
May 16, 2019
May 17, 2019
May 18, 2019
May 19, 2019
May 20, 2019
May 21, 2019
May 22, 2019
May 23, 2019
May 24, 2019
May 25, 2019
May 26, 2019
May 27, 2019
May 28, 2019
May 29, 2019
May 30, 2019
May 31, 2019


Superbowl Total Scores By Year


Cropland Average Value Per Acre in United States

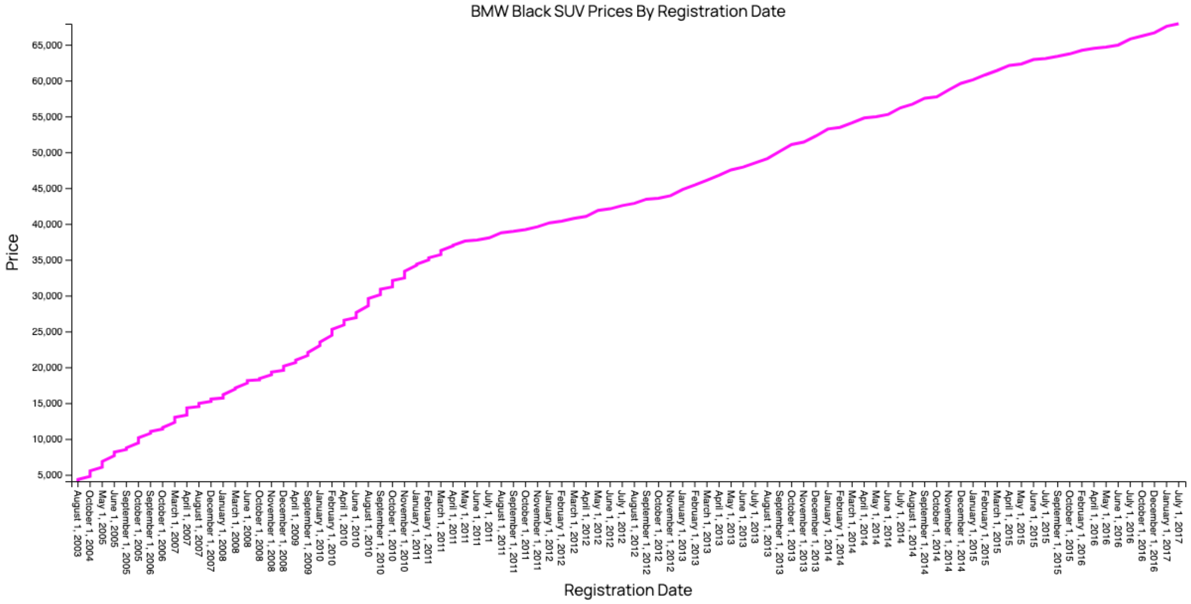
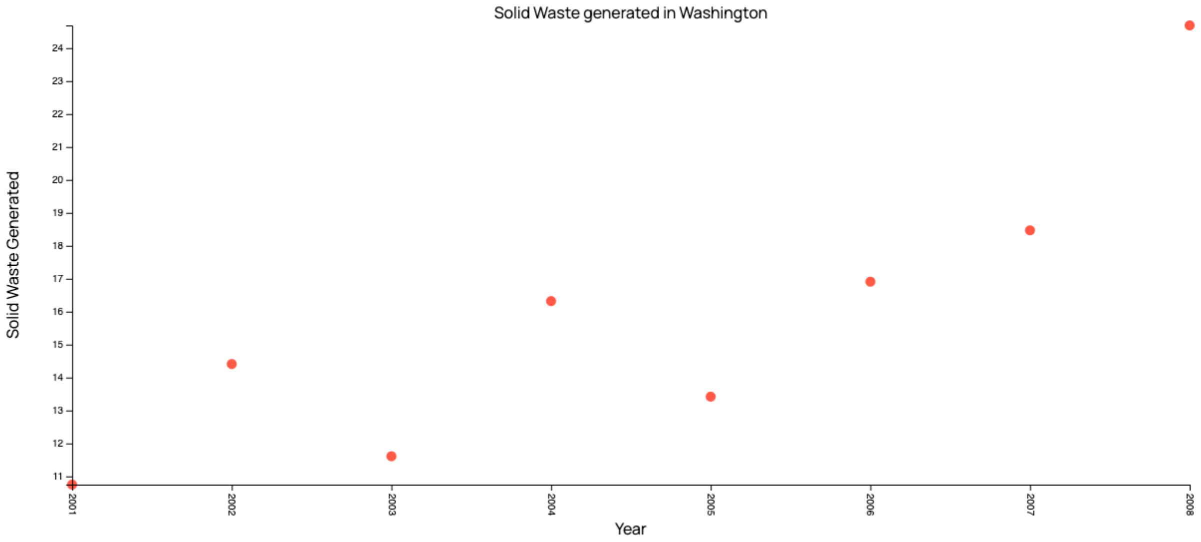


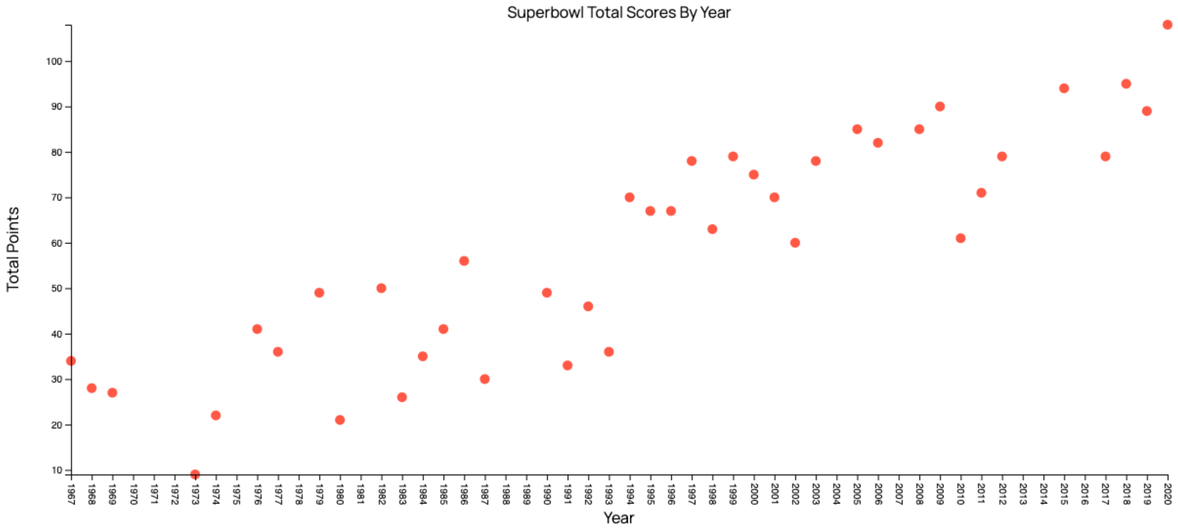
Figure B.1: Line charts shown during the task-based experiment to measure accuracy of extracted information and interaction times. These visualizations are created using D3 (Bostock et al. [2011]).

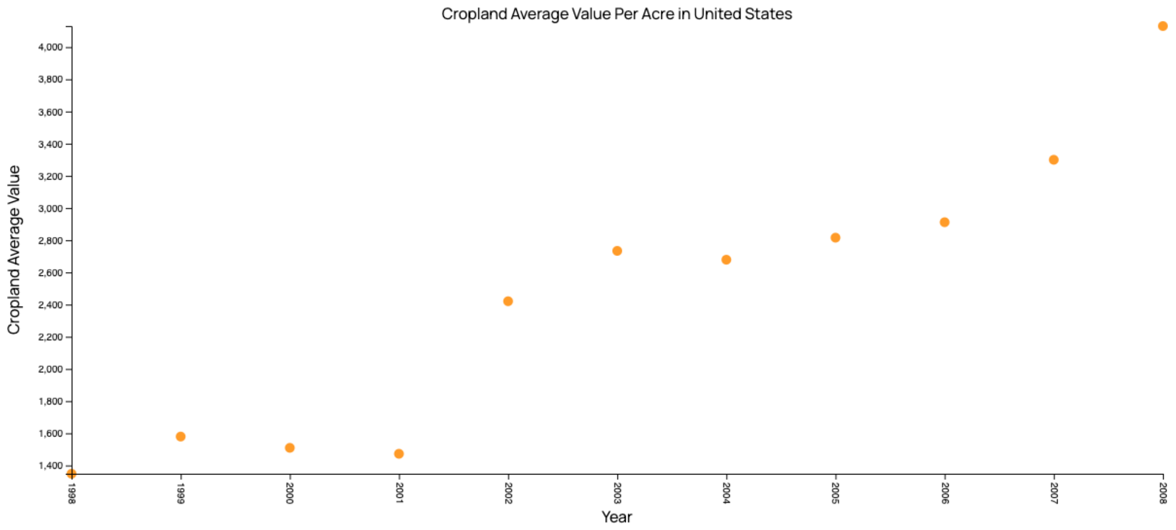
Appendix C

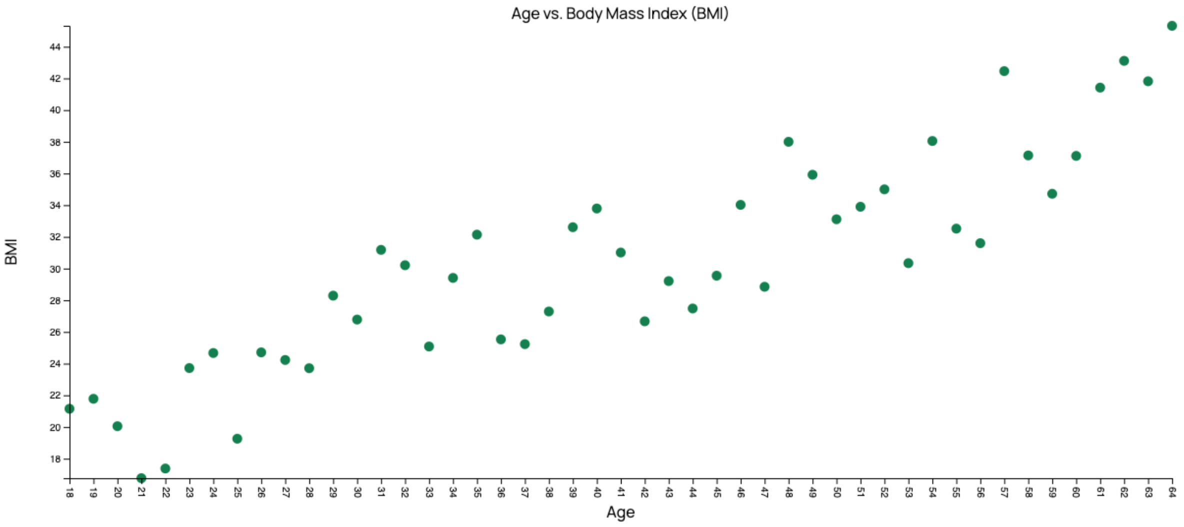
Scatter Chart Stimuli

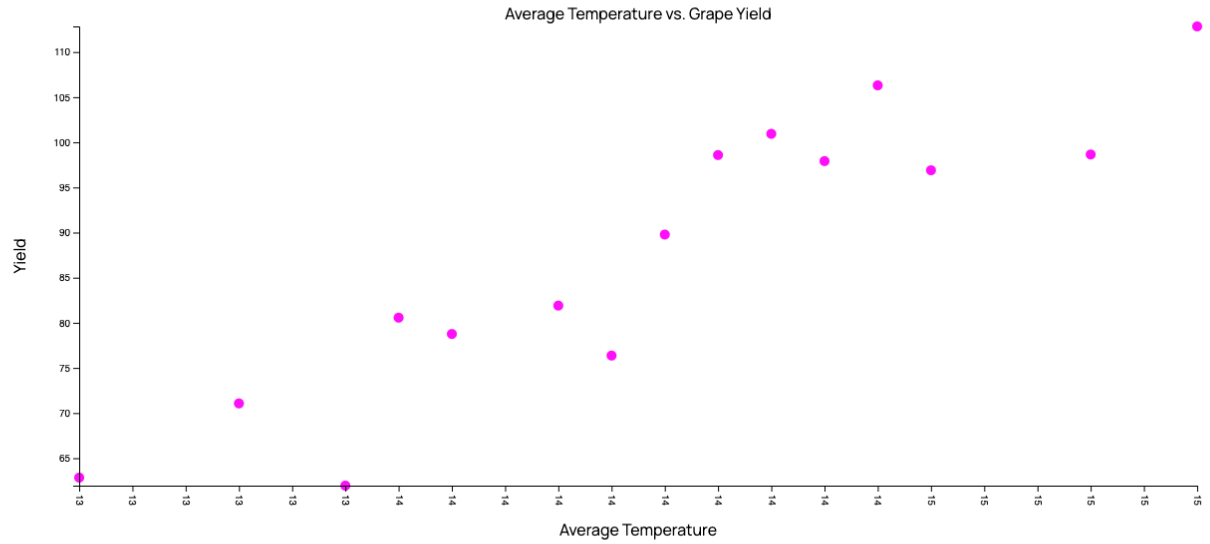
This appendix shows the scatter plot stimuli used during the task-based experiment described in Chapter 3.

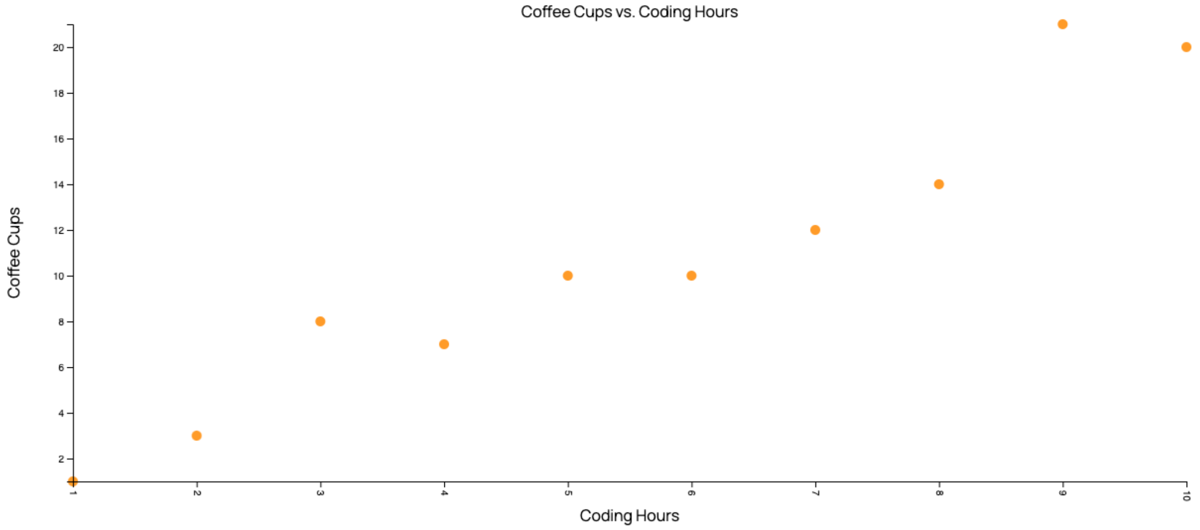

Solid Waste generated in Washington
Solid Waste Generated
24
23
22
21
20
19
18
17
16
15
14
13
12
11
2001
2002
2003
2004
2005
2006
2007
2008
Year

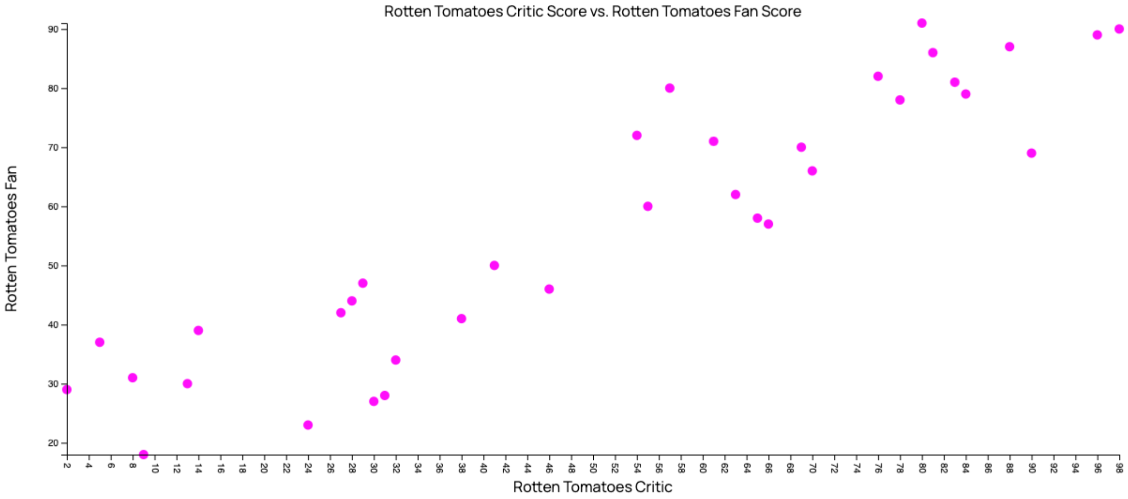

Superbowl Total Scores By Year
Total Points
100
90
80
70
60
50
40
30
20
10
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Year

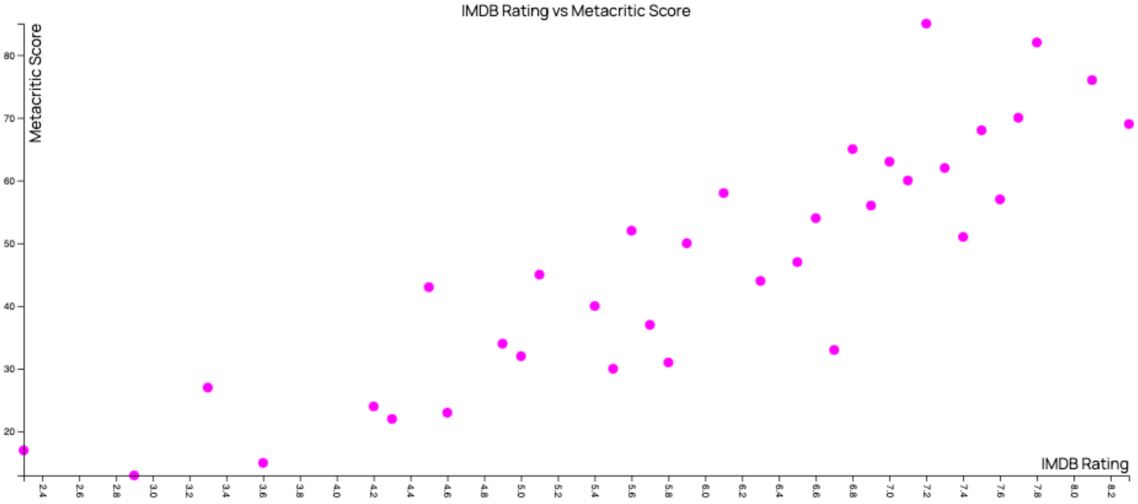

Cropland Average Value Per Acre in United States
Cropland Average Value
4,000
3,800
3,600
3,400
3,200
3,000
2,800
2,600
2,400
2,200
2,000
1,800
1,600
1,400
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
Year


Age vs. Body Mass Index (BMI)
BMI
44
42
40
38
36
34
32
30
28
26
24
22
20
18
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
Age


Average Temperature vs. Grape Yield
Yield
110
105
100
95
90
85
80
75
70
65
13
13
13
13
13
13
14
14
14
14
14
14
14
14
14
14
15
15
15
15
15
15
Average Temperature


Coffee Cups vs. Coding Hours
Coffee Cups
20
18
16
14
12
10
8
6
4
2
1
2
3
4
5
6
7
8
9
10
Coding Hours


Rotten Tomatoes Critic Score vs. Rotten Tomatoes Fan Score
Rotten Tomatoes Fan
90
80
70
60
50
40
30
20
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98
Rotten Tomatoes Critic


IMDB Rating vs Metacritic Score
Metacritic Score
80
70
60
50
40
30
20
IMDB Rating
2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 4.0 4.2 4.4 4.6 4.8 5.0 5.2 5.4 5.6 5.8 6.0 6.2 6.4 6.6 6.8 7.0 7.2 7.4 7.6 7.8 8.0 8.2

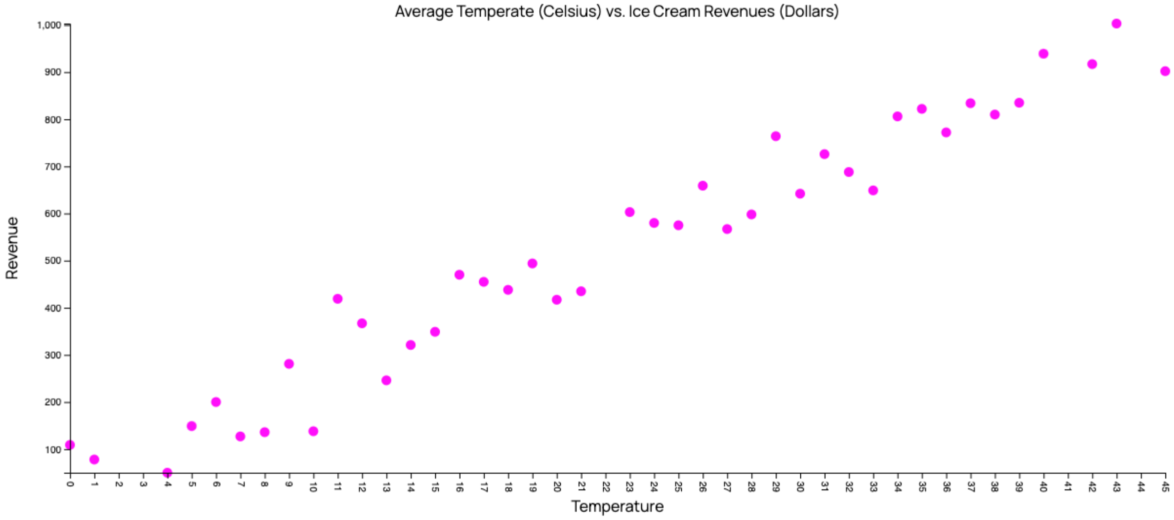


Figure C.1: Scatter charts shown during the task-based experiment to measure accuracy of extracted information and interaction times. These visualizations are created using D3 (Bostock et al. [2011]).

Appendix D

D3 Implementation With VOXLENS

This appendix shows the code for implementing a D3 visualization using VOXLENS.

```
import * as d3 from 'd3';
import max from 'lodash/max';
import startCase from 'lodash/startCase';
import voxlens from 'voxlens';

const createD3 = (options) => {
  let {
    chartType,
    data,
    title,
    xKey,
    yKey,
  } = options;

  const getDimensions = (maxXLabel) => {
    const margin = { top: 20, right: 40, bottom: maxXLabel * 5 + 10, left: 70 };

    return {
      margin,
      height: 500 - margin.top - 20,
      width: container.offsetWidth - margin.left - margin.right,
    };
  };

  const container = document.getElementById('chart');

  let maxXLabel = max(data.map((d) => d[xKey].toString().length));
  let { height, margin, width } = getDimensions(maxXLabel);
  let transform = margin.left + ',' + margin.top;

  const voxlensOptions = {
    x: xKey,
    y: yKey,
    title,
    chartType,
  };

  const svg = d3
    .select('#chart')
    .append('center')
    .append('svg')
    .attr('width', width + margin.left + margin.right)
    .attr('height', height + 40 + margin.top + margin.bottom)
    .append('g')
    .attr('transform', 'translate(' + transform + ')');

  const x = d3.scaleBand().range([0, width]).padding(0.1);
  const y = d3.scaleLinear().range([height, 0]);

  x.domain(data.map((d) => d[xKey]));
  y.domain([0, d3.max(data, (d) => parseFloat(d[yKey]))]);
```

```
  svg
    .selectAll('.bar')
    .data(data)
    .enter()
    .append('rect')
    .attr('class', 'bar')
    .style('fill', 'steelblue')
    .style('margin', '10px')
    .attr('x', (d) => x(d[xKey]))
    .attr('width', x.bandwidth())
    .attr('y', (d) => y(d[yKey]))
    .attr('height', (d) => height - y(d[yKey]))
    .call((d) => (voxlens('d3', d, data, voxlensOptions)));
  svg
    .append('text')
    .attr(
      'transform',
      'translate(' + width / 2 + ' ,' + (height + margin.bottom + 20) + ')'
    )
    .style('text-anchor', 'middle')
    .text(startCase(xKey));

  svg
    .append('text')
    .attr('transform', 'rotate(-90)')
    .attr('y', 0 - margin.left)
    .attr('x', 0 - height / 2)
    .attr('dy', '1em')
    .style('text-anchor', 'middle')
    .text(startCase(yKey));

  svg
    .append('g')
    .attr('transform', 'translate(0, ' + height + ')')
    .call(d3.axisBottom(x).ticks(data.length))
    .selectAll('text')
    .attr('y', 0)
    .attr('x', 9)
    .attr('dy', '.35em')
    .attr('transform', 'rotate(90)')
    .style('text-anchor', 'start')
    .style('opacity', 1);

  svg.append('g').call(d3.axisLeft(y));
};

export default createD3;
```

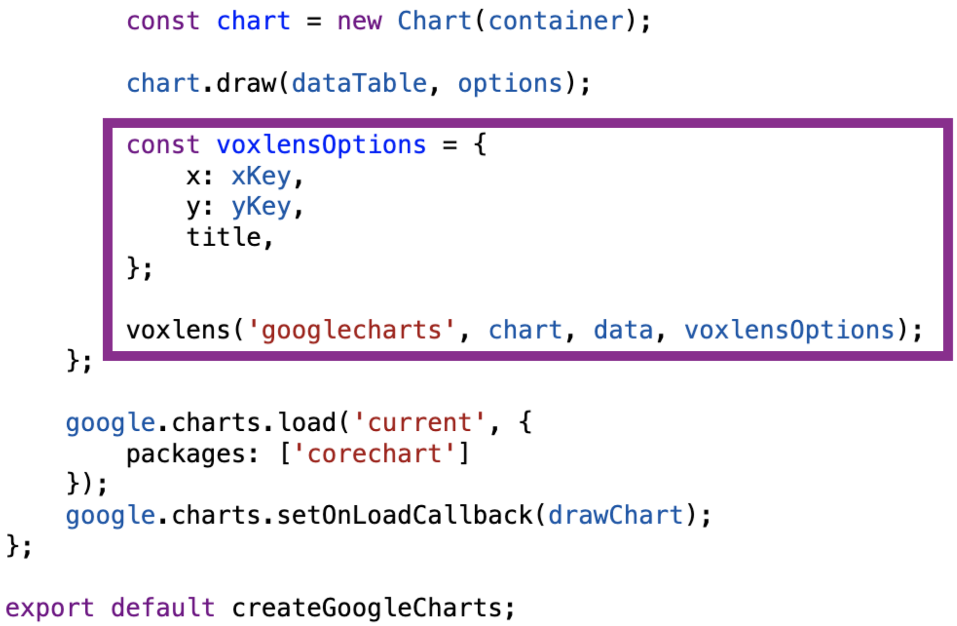
Figure D.1: D3 code for implementing a bar chart using VoxLens. The code is written in JavaScript. Code relevant to VoxLens is highlighted with a rectangular box.

Appendix E

Google Charts Implementation With VOXLENS

This appendix shows the code for implementing a Google Charts visualization using VOXLENS.

```
import startCase from 'lodash/startCase';
import voxlens from 'voxlens';

const createGoogleCharts = (options) => {
    let {
        data,
        fillColor,
        title,
        xKey,
        yKey
    } = options;

    const container = document.getElementById('chart');
    const margin = {
        top: 20,
        right: 40,
        bottom: 20,
        left: 70
    };
    const height = 700 - margin.top - margin.bottom;
    const width = container.offsetWidth - margin.left - margin.right;
    const google = window.google;

    const drawChart = () => {
        const dataTable = new google.visualization.DataTable();

        let Chart = google.visualization.ColumnChart;

        dataTable.addColumn('string', startCase(xKey));
        dataTable.addColumn('number', startCase(yKey));
        dataTable.addRows(
            data.map((d) => [d[xKey].toString(), parseFloat(d[yKey])])
        );

        const options = {
            legend: {
                position: 'none'
            },
            bars: 'horizontal',
            colors: [fillColor],
            title,
            width,
            height,
            hAxis: {
                title: startCase(xKey),
                slantedText: false,
            },
            vAxis: {
                title: startCase(yKey),
                baseline: 0,
                gridlines: {
                    color: 'black',
                },
            },
        };
```



```
        const chart = new Chart(container);

        chart.draw(dataTable, options);

        const voxlensOptions = {
            x: xKey,
            y: yKey,
            title,
        };

        voxlens('googlecharts', chart, data, voxlensOptions);
    };

    google.charts.load('current', {
        packages: ['corechart']
    });
    google.charts.setOnLoadCallback(drawChart);
};

export default createGoogleCharts;
```

Figure E.1: Google Charts code for implementing a bar chart using VoxLens. The code is written in JavaScript. Code relevant to VoxLens is highlighted with a rectangular box.

Appendix F

ChartJS Implementation With VOXLENS

This appendix shows the code for implementing a ChartJS visualization using VOXLENS.

```
import Chart from 'chart.js';
import startCase from 'lodash/startCase';
import voxlens from 'voxlens';

const createChartJS = (options) => {
    const {
        data,
        fillColor,
        title,
        xKey,
        yKey
    } = options;

    const container = document.getElementById('chart');
    const margin = {
        top: 20,
        right: 40,
        bottom: 40,
        left: 70
    };
    const height = container.height - margin.top - margin.bottom;
    const width = container.width - margin.left - margin.right;

    const ctx = container.getContext('2d');
    ctx.canvas.width = width;
    ctx.canvas.height = height;

    const mappedData = data.map((d) => d[yKey]);

    new Chart(ctx, {
        type: 'bar',
        data: {
            labels: data.map((d) => d[xKey]),
            datasets: [{
                label: startCase(yKey),
                data: mappedData,
                fill: true,
                backgroundColor: fillColor,
                borderColor: fillColor,
            }, ],
        },
        options: {
            maintainAspectRatio: true,
            title: {
                display: true,
                text: title,
            },
            legend: false,
```

```
            scales: {
                xAxes: [{
                    scaleLabel: {
                        display: true,
                        labelString: startCase(xKey),
                    },
                    gridLines: {
                        display: true,
                    },
                }, ],
                yAxes: [{
                    scaleLabel: {
                        display: true,
                        labelString: startCase(yKey),
                    },
                    gridLines: {
                        display: true,
                    },
                    ticks: {
                        beginAtZero: true,
                    },
                }, ],
            },
        },
    });

    const voxlensOptions = {
        x: xKey,
        y: yKey,
        title,
    };

    voxlens('chartjs', container, data, voxlensOptions);
};

export default createChartJS;
```

Figure F.1: ChartJS code for implementing a bar chart using VOXLENS. The code is written in JavaScript. Code relevant to VOXLENS is highlighted with a rectangular box.

Vita

Ather Sharif was born in Karachi, Pakistan, in 1989. Ather earned a Bachelor of Science in Software Engineering from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Pakistan, in 2011. He moved to the United States in 2010 to pursue a Master's degree at the University of North Dakota. Due to a car accident that rendered him paralyzed from the neck down, he relocated to Philadelphia in 2013, which he calls home, and where he earned a Master of Science degree in Computer Science from Saint Joseph's University in 2016. He started his doctorate in Computer Science at the Paul G. Allen School of Computer Science and Engineering at the University of Washington in 2018. As a doctoral student, he pioneered two first-of-its-kind systems, VoxLens and UnlockedMaps, that aim to reduce the digital divide for disabled people. His research has been published at top-tier academic venues and received several accolades, including a Best Technical Paper award, a Best Paper Nomination, and two Best Artifact awards. In addition, he received several fellowships and awards for his research work, including the Allen School Fellowship, Google CMD-IT LEAP Dissertation Fellowship, Open Scholarship Commons Fellowship, and Husky 100 Award. Several media outlets, including the UW News and Philadelphia Inquirer, have showcased his work.