Within Context: Exploring Visualization and Qualitative Research

Fida AlSughayer

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Computer Science & Software Engineering

University of Washington

2017

Committee:

David Socha, Chair

Kelvin Sung

Joseph McCarthy

Program Authorized to Offer Degree:

Computing and Software Systems

University of Washington

Abstract

Within Context: Exploring Visualization and Qualitative Research

Fida AlSughayer

Chair of the Supervisory Committee:
Dr. David Socha
Computing and Software Systems

The literature of visualization exhibits more focus on quantitative forms of data and analysis and the literature of qualitative research has few works that study and structure the relationship between visualizations and qualitative data. This thesis inquires: how can we reason about what is qualitative in visualizations? We explore this question through three approaches. First, this thesis examines and refutes the positioning of qualitative-ness as opposite of quantitative-ness in visualization. Second, this thesis studies and organizes the "multiplicity of engagements" (Cope and Elwood, 2009) between visualizations and the literature of qualitative research. Third, the thesis presents the findings from a series of qualitative interviews which explored how researchers use visualizations in the process of qualitative data analysis. Our findings in this thesis suggest that, when approached qualitatively, a visualization is strongly embedded in the context that is represents. A visualization provides an abstract view of some context and it is read, understood, and synthesized within that context.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The work reported in this thesis is my work, however, I use the words “we” and “our” throughout it because this thesis would not have been possible without the assistance of many others. My utmost thanks and sincere gratitude go to my advisor and mentor, Professor David Socha, for his insights, guidance, continuous support, and, above all, his patience throughout the process of this thesis. Many thanks for Professor Joe McCarthy and Professor Kelvin Sung for their support and constructive feedback, to Nancy Kool for helping me choose the right words, and to Megan Jewell for her assistance in navigating the graduate school process.

DEDICATION

This work, with gratitude, is dedicated to my mother and father and to Bushra, Omar, Salman and Jana – apart together.

Chapter 1. INTRODUCTION

This thesis asks: How can we reason about what is qualitative in visualization? Let us consider this question in the context of the visualization in Figure 1. Socha and Sutanto [1] present this visualization in their paper "The Pair as a Problematic Unit of Analysis for Pair Programming," in which they study interactions among software developers at their place of work, BeamCoffer, a software company in Seattle, WA. Specifically, the authors investigate interactions that occur between a pair of software developers (sitting at a pairing station) and anyone *outside* of their pairing station. Figure 1 represents data collected from one pairing station at the company, BeamCoffer. The figure, a time-series, depicts extra-pair interactions throughout a single day, where the blue-shaded rectangle represents the pairing station, and the arrows represent extra-pair interactions.

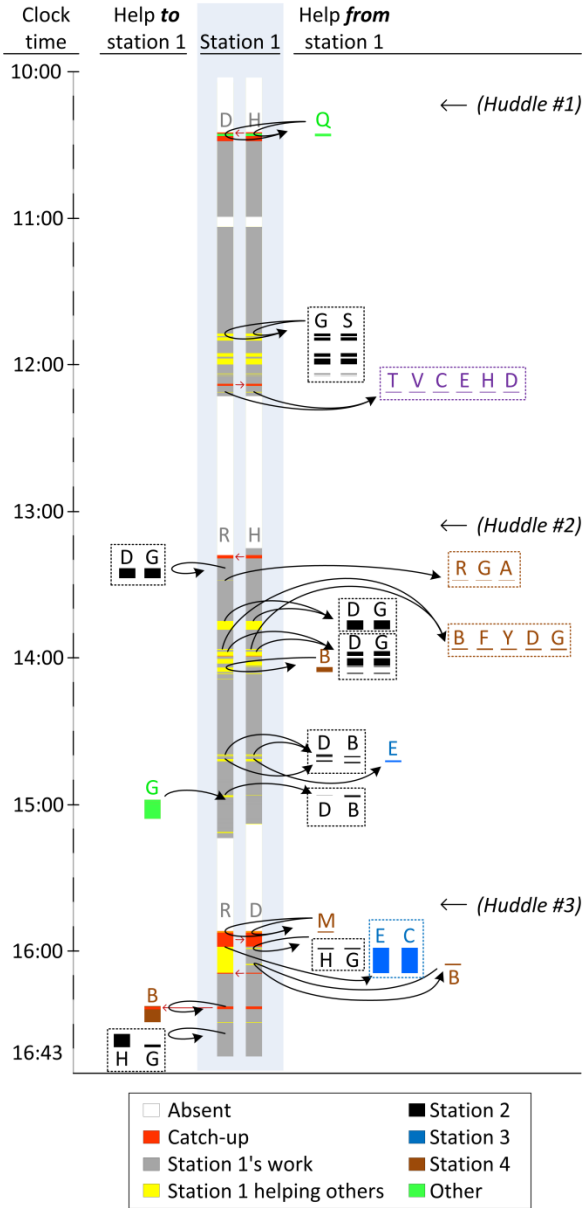

Figure 1 – Interaction Score (created by Socha and Sutanto [1])

This visualization features several qualitative aspects. For one, the data represented in this visualization were generated using qualitative fieldwork methods. Socha [2] spent days in

the field at BeamCoffer, taking notes and recording hours of video and audio. The resultant dataset, from which the visualization was derived, was large (more than 8 terabytes), multi-modal (including various data types), and rich with context (physical, social, and digital). This detailed and contextual data supports the qualitative construction of thick descriptions [3] and rounded understanding [4] of many phenomena. In addition to the qualitative-ness of the data within the visualization, Socha and Sutanto created this visualization as part of their qualitative analysis activities. Specifically, the authors used Interaction Analysis [5], an ethnographically-informed method of analysis in which researchers repeatedly playback video segments and inspect interactions, while taking into account the distributed nature of cognition in social contexts. As the researchers analyzed extra-pair interactions, they constructed a version of this visualization on a whiteboard. This enabled them to see, observe, and communicate their analysis insights. Finally, the authors used this visualization to report and evidence the results of their qualitative and empirical work in their research paper [1].

Figure 1 elucidates qualitative features in data, analysis, and research results. Our inspection of these features motivated a range of inquiries: Is the qualitative-ness of a visualization characteristic of the data, the analysis method, or, perhaps, the visualization itself? Are visualizations of qualitative data characteristically different from other types of visualizations? Does the creation of a visualization itself influence qualitative approaches? Do qualitative approaches influence visualizations? Again, *how can we reason about what is qualitative in visualization?*

These questions have not been previously explored in the literature of visualization, despite the wide-spread application of visualizations across many fields of knowledge, scientific, artistic, and technological [6] [7] [8]. In fact, the literature of visualization exhibits an emphasis

on *quantitative* data sources, processes, and analyses [9] [10] [11]. The works of leading contributors to the study of visualizations (e.g., Tufte [12], Tukey [13], and Few [14]) exhibit this imbalance towards statistical models and frameworks for quantitative data. Similarly, the discourse in leading conferences and journals on visualizations (e.g., IEEE Conference on Information Visualization and IEEE Transaction in Visualizations and Graphics) highlights innovations in technique (i.e., algorithmic improvements), design (i.e., novel visual representation), systems (i.e., architecture), and modeling (e.g., taxonomies of visual encodings) [15].

The literature of qualitative research, on the other hand, depicts many examples of using visualizations to generate, analyze and report qualitative data (Section 3.2). Many qualitative researchers advocate the use of visualizations to support qualitative processes. For example, Corbin and Strauss [16] encourage the use of diagrams for constructing grounded theories, Russel Bernard and Ryan [17] describe an array of visualizations to support systematic approaches to qualitative analysis, and Miles and Huberman [18] advocate visual display as a cornerstone of iterative, qualitative knowledge production. Additionally, in an attempt to fill the gap between visualization and qualitative research, Henderson and Segal [10] and Onwuegbuzie et al. [19] have developed frameworks for displaying some types of qualitative data (Section 3.2.2). Even so, beyond these references, the use of visualizations within the realm of *qualitative* research remains an area that is less explored (as I evidence through literature review in Chapter 3).

In this thesis, we ask:

How can we reason about qualitative-ness in visualization?

We explore this question through three approaches:

- In the first approach, we ask: Can a visualization be qualitative? Is *qualitative-ness the opposite of quantitative-ness*?
- In the second approach, we ask: How do visualizations engage with qualitative research? We pursue these questions in the literature of qualitative research and we structure, categorize, and report the *engagements between visualization and the discourse of qualitative research*.
- In the third approach, we ask: How do qualitative researchers use visualizations? We report our findings from 7 interviews with researchers who use visualizations to support their qualitative analysis activities. Specifically, we share insights regarding *qualitative approaches to visualizations*, and we explain how these approaches describe the role of visualizations as abstract *views of context* embedded *within context*.

Chapter 2. THE PROBLEM WITH FRAMING THE QUALITATIVE AS AN OPPOSITE TO THE QUANTITATIVE

Earlier, when we began to ask questions about the qualitative-ness of visualizations, we found ourselves drawn to categorize visualizations as either quantitative or qualitative. In doing so, we were, perhaps unintentionally, delineating the qualitative nature we investigate in visualizations as separate, or even contradictory, to the quantitative. Indeed, we found echoes of this classification in discussions with colleagues, responses from interview participants (Chapter 4), and examples from literature that we present throughout this chapter.

> We ask:
>
> *Is the qualitative-ness of visualizations an opposite of quantitative-ness?*

The perception of dichotomy between what is qualitative and what is quantitative, which is here applied to visualizations, could be considered an extension of the varying degrees of distinction and contention that exist between the qualitative and quantitative methods of research. Given its multi-faceted definition, "it is often argued that the binding feature of qualitative research is its opposition to positivism, the philosophical basis for quantitative research" [20, p. 1]. Further, this perception of dichotomy can be attributed to the discourse of information visualization, scientific or technical, which predominantly uses quantitative descriptors when addressing the *process* or *artifacts* of visualization. To exemplify the quantitative description of *process*, we reference the theoretical influence of Tukey [13] who wrote the book on *Exploratory Data Analysis* where he advocates for statistical exploration of phenomena in quantitative data, in part through the use of visualizations. To exemplify quantitative description of visual *artifact*, we reference the ground-setting works of Playfair [14],

inventor of the bar chart and the pie chart, which were primarily statistical; and the volumes of Tufte [12], particularly *The Visual Display of Quantitative Information,* in which he defines the practices for crafting "statistical graphics".

The visualization in Figure 2 challenges this dichotomy of perception. Figure 2 is prepared by the Office of Planning and Development at the City of Seattle [21] based on census data; it features a horizontal bar chart that represents the rates of poverty among a selection of social groups, including people of color, people with disabilities and foreign-born people. At first inspection, the visualization might be considered quantitative because (a) it uses a bar chart, which is traditional artifact in the discourse of quantitative visualizations; and (b) its percentages are calculated based on statistical processing of census data. However, the visualization shows many traces of qualitative characteristics as well: for example (a) the data might be collected in part using fieldwork, which involves qualitative methods; (b) the categorizations reflect a complex history of social development; (c) the categorizations are based on social understandings that are by nature subject to qualitative interpretation; and (d) the categorizations represent topics that are subject to rich theoretical discussions in qualitative research, such as race, gender, immigration and disability. Additionally, this visualization can create *opportunities*

for qualitative work; for example, it might prompt questions about the experiences of those undergoing poverty among these social groups or the relation and causality between them.

Figure 2 – A horizontal bar chart representing the rates of poverty in Seattle in between 2011 and 2013 among some selected social groups (prepared by Office of Planning and Community Development - City of Seattle [21]).

The qualitative characteristics and opportunities reflected in Figure 2, particularly with its social context, might not be as prominent in all examples of visualizations. However, all quantitative visualizations, with the exception of scatterplots, include categories in addition to quantitative scales [22] [23]. The categories define *what* the visual encoding represents (e.g., the bars in Figure 2 represent social categories), while the quantitative scales define *how many* of the members of a category are represented (e.g., the width of the bar in Figure 2 represents the percentage of poverty among a social category). Categories have been defined to comprise a variety of data types in in the science of statistics [24] and visualizations [22] [25]. Figure 3 shows examples of three categorical data types as defined and exemplified by Stephen Few [23]: (a) nominal types, which denote discrete labels that have no intrinsic order, such as department, gender, or car type; (b) ordinal types, which denote discrete labels that have an intrinsic order, such as rank, dates, or age; and (c) interval types, which denote quantities that have an intrinsic

order; which includes numbers of various representations, such as integers, ratios, or percentages.

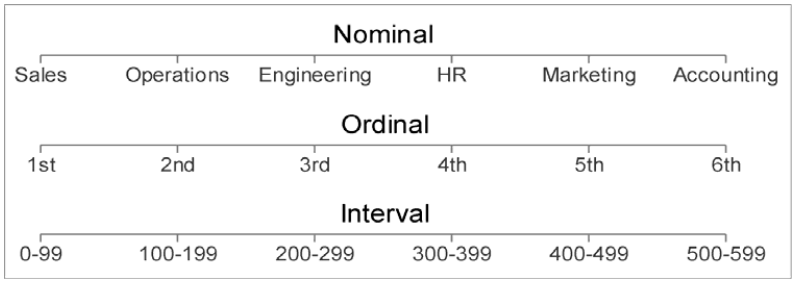

Figure 3 – Examples of Categorical data types as presented by Stephen Few [23]

With their inclusion of non-numerical data types, categories allow for qualitative characteristics and opportunities in varying degrees. Even in visualizations which are derived through purely statistical analysis, categories remain constructed concepts that are subject to human design and interpretation. For example, the nominal values in Figure 3 depict department categories that are ubiquitous in the structure of modern companies. Those departments represent social organizations of work that developed over a history of economic, social, and spatial factors; these factors are complex, rich, and qualitative. The constructed nature of categories in visualizations explains their prevalence in supporting the statistical analysis of human phenomena in social sciences, political sciences, and user research [26]. Categories also enable the visualization of data that is collected and analyzed through qualitative methods. For example, in content analysis of text (traditionally considered a qualitative data type [10] [18]), human coders can classify content into categories that can be represented in a variety of visualizations [27, p. 149]. We provide more examples on visualizing qualitative data in Section 3.2.2.

We continue our challenge of the perceived dichotomy between quantitative and qualitative descriptions of visualizations by presenting examples that are more difficult to

classify. Figure 4 depicts an interactive visualization, developed by a group of students at Stanford University, which represents three categories: (1) the departments at Stanford University (represented as color-coded circles); (2) the number of dissertations written by PhD students in 2008 (represented by the size of the circle); and (3) the dissertation topic similarity between departments (represented as the distance between circles). This visualization bears visual similarities to quantitative node-link diagrams [28] [29]. It is not, however, a direct application of any well-known quantitative visualization. The visual encodings in Figure 4 are computed using statistical models (quantitative) that process collections of abstracts (text, traditionally perceived as qualitative [10] [18]). Similarly, Figure 5 depicts a Word Cloud used as a supplementary tool to explore the discussions of focus-group participants [30]. Word Clouds are widely-used visualizations for content exploration [31] [32] despite being subject to usability critique [33] [34]. In this example, the size of the word corresponds to the frequency of the word (quantitative) in the transcript of the focus group discussion (qualitative). Both visualizations depict a mixture of quantitative and qualitative characteristics and are thus more difficult to classify than the bar chart and categorical scales in Figure 2 and Figure 3.

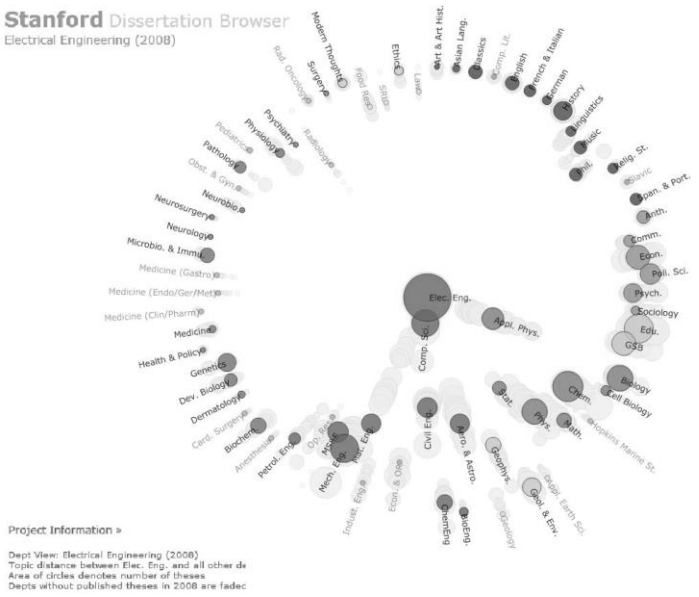

Figure 4 – Stanford Dissertation Viewer (created by Jason Chuang and Daniel Ramage [35])

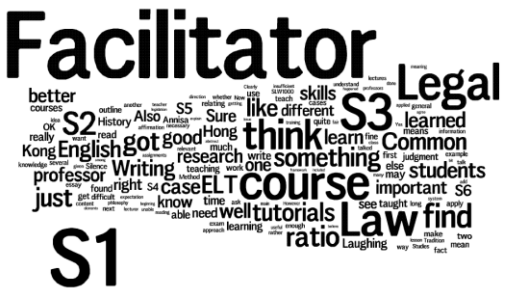

Figure 5 – A Word Cloud of common words in a discussion of a group of law students (created by Carmel McNaught and Paul Lam [30])

Visualizations are powerful because they utilize the human mind's capacity for visual thinking and processing [36] [37], regardless of whether the processing is qualitative or quantitative. The many viewpoints of visualization as a discipline (scientific, technological, and artistic [6]) intersect at the human mind. So, despite the nature of data or research method, these remain persistent: the constraints of sight, the graphical perception of the mind, and the psychophysics of interaction between the artifact and the sensory system [38]. Furthermore, in

both quantitative and qualitative fields of research, visualizations pursue the same values of integrity, efficiency, effectiveness, and aesthetics [12] [11].

The examples in this section depict the various ways in which the qualitative and quantitative coexist in visualizations: (a) visualizations traditionally described as quantitative include qualitative characteristics and opportunities (Figure 2 and Figure 3); (b) qualitative methods are often used to collect quantitative data (Figure 2); (c) qualitative data can be statistically processed (Figure 4 and Figure 5); and (d) visualizations can be difficult to classify simply as either qualitative or quantitative (Figure 4 and Figure 5). These statements do not refute the fact that there is a distinction between qualitative and quantitative methods in general and that analysis can have more of either. In the context of visualizations, however, we believe that this distinction becomes less important. It is not valuable to say that a visualization is *more* qualitative because it has *less* intrinsically quantitative characteristics.

On the contrary, one of the strongest features of visualization is that it is a space where qualitative and quantitative data and methods *meet and work*. This enables/makes possible mixed-method approaches. There are two examples in the literature that reference this theory about the role of visualization in mixed-method approaches: (1) Cope and Elwood [39] position their description of *Qualitative GIS* as a mixed-method approach and they provide the rationale for this position in their book with the same title (see Section 3.1). (2) Onwueghuzie et al. [11] echo the sentiment in their article "Mixed-methods Analysis and Information Visualization: Graphical Display for Effective Communication of Research Results" where they provide a taxonomy for visualizations that integrates results from mixed-methods research in a shared framework (see Section 3.2.2). These two works indicate an opportunity for more discourse on the link between visualization and mixed-method approaches. Our survey of literature did not

uncover any more references to this link. Still, the aforementioned works solidified our position that defining the qualitative as an opposite to the quantitative in visualizations is not important and could even be counterproductive.

Furthermore, we do not wish to add to the divide between qualitative and quantitative analysis in general. There are movements in both fields advocating for triangulation using mixed-methods and asserting that a conjunction of qualitative and quantitative methods could support stronger findings [40] [41] [42]. For these reasons, we refrain from framing our questions about visualizations and qualitative research in a manner that posits a dichotomy with quantitative processes or visualizations.

We conclude:

Because qualitative-ness and quantitative-ness co-exist in visualizations, we do not define the qualitative-ness of a visualization to be an opposite of the quantitative-ness. A visualization is not entirely quantitative nor entirely qualitative; rather, visualizations include qualitative and qualitative characteristics.

Chapter 3. THE "MULTIPLICITY OF ENGAGEMENTS" BETWEEN VISUALIZATION AND QUALITATIVE RESEARCH

The literature of visualization is rooted in quantitative discourse [10] [9] [11] [43]. By contrast, the literature of Geographic Information Systems (GIS) has benefited from rich methodical and philosophical discussions that disrupted the field and shifted it towards qualitative approaches since the 1990s [44] [45] [46]. GIS is a relevant field to visualization as it produces artifacts that are visual representations of spatial data; though it had evolved as a separate field from visualization [47].

> We ask:
>
> *What can we learn from the qualitative movements in GIS? How can we use the insights gained from Qualitative GIS to reason about qualitative-ness in visualization?*

3.1 Learning from the Qualitative Movements In GIS

Geographic Information Systems are "digital technologies for storing, managing, analyzing and representing geographic information" [39, p. 9]. The cartographic and non-cartographic artifacts produced by GIS are a subset of visualizations representing spatial data, termed as geo-visualizations [47]. While there is a strong link between GIS and visualization, the two have evolved as separate fields of research and knowledge creation.

The progress of GIS towards qualitative discourse dates from a 1993 meeting between human geographers and GIS scholars at a workshop that was organized by the University of

Washington at Friday Harbor [48]. That meeting is widely considered to mark the birth of the dialogue on critical GIS and the catalyst for the emergence and progress of feminist GIS [44] and participatory GIS [46] [49]. These discussions have shifted and shaped new directions in GIS research that advocate for its integration with qualitative methods of knowledge creation, analysis and representation [39]. In what follows, we provide examples and explanations for a variety of movements and integration methods in GIS.

The qualitative movements in GIS advocate for the inclusion of multiple forms of data from multiple sources in a visualization [39] [48] [50]. They encourage innovation beyond the quantitative techniques that rely on conventional, location-based, or Cartesian data. Many examples of participatory GIS and Feminist GIS embed images, audio, video, or textual narratives into cartographic or non-cartographic artifacts [51] [52]. The argument is that qualitative data can preserve the voices and experiences behind the GIS artifact, and express factors such as social contexts, subjective positions in knowledge creation, and power relations. Figure 6 presents an example created by Kwan [50], a leading contributor to the field of Feminist GIS. The visualization in Figure 6 represents the daily life paths of African-American women plotted in a 3D model, where the horizontal planes represent spatial travel and the vertical planes represent time. The data used in this visualization were collected through qualitative methods, namely travel diaries kept by participants. The visualization maps women's lives into space-time and constructs a cartographic narrative that is enriched with personal and social context. The visualization also supports Kwan's findings that the movement of women is constricted to a small metropolitan area [53]; a "closeted-spatiality" [50, p. 654] that she links to gender and ethnic factors.

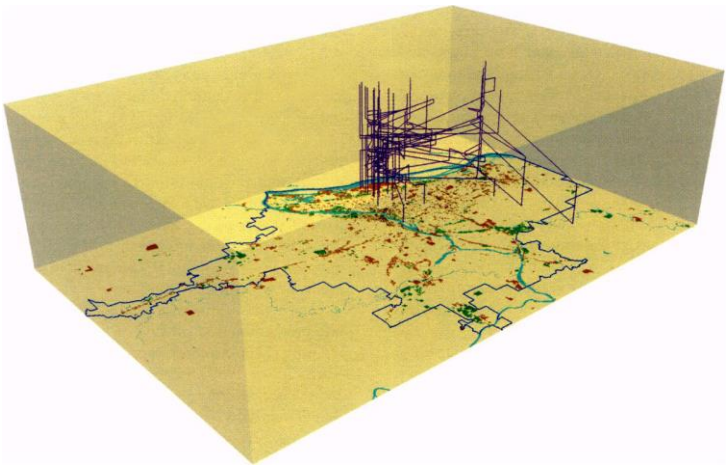

Figure 6 – A 3D geo-visualization that represents the paths of a sample of African-American women in Portland, Oregon (created by Kwan [50])

In addition to representing multi-form data, qualitative movements in GIS advocate for the use of visualization at the analysis level of research. Geo-visualizations can support the process of coding, triangulation of multi-form data, iterative investigation, and a variety of analysis techniques (e.g., content analysis and discourse analysis) [39]. The theory is that both visualizations and qualitative analysis support exploratory approaches. For example, Knigge and Cope [54] argue that visualizations and Grounded Theory [16] both "enable, and, in fact, demand, researchers to query the data from multiple angles, ponder emerging consistencies or dis-junctures, make new or revised connections, and entertain rival explanations" [54, p. 2033]. The authors conceptualize Grounded Visualization as an approach to integrating qualitative and quantitative analysis with support from GIS. They demonstrate it through a case study about community gardens in Buffalo, New York where they iteratively (a) collected evidence through ethnographic fieldwork; (b) coded and analyzed their evidence; and (c) explored geo-visualizations of census data surrounding public gardens they study. These iterations were reflexive and contextual, providing multiple views of reality. The authors report that Grounded Visualization facilitated their exploration and interpretation of a complex social world.

The advances in the fields of Critical GIS, Participatory GIS and Feminist GIS forged the development of Qualitative GIS as a major research approach [44]. Qualitative GIS theorizes that the development of GIS processes and artifacts involves the negotiation of qualitative meanings (e.g., social or political) and geographical phenomena at different stages of production. As a result, qualitative methods are not considered simply as add-ons to GIS (like linking images to map locations), but as opportunities for engagement in knowledge generation, representation, and analytical process [39] [55].

3.2 The Multiplicity of Engagements Between Qualitative Research and Visualization

The qualitative movements in GIS spanned over two decades and culminated in the rich discourse of Qualitative GIS; but the literature of visualizations in general did not benefit from the same philosophical and methodical disruption. Rather, it remains rooted in quantitative epistemologies [11].

Nevertheless, qualitative research literature is rich with examples where visualizations are applied throughout all phases of research. The visualizations used are of different types (e.g., static or interactive; formal or researcher-defined) and are applied towards different research goals. We ask: Is it possible to learn what constitutes the qualitative-ness of a visualization by studying the way visualizations are used in qualitative research? We conjecture that the qualitative approaches in research influences the manner in which visualizations are used as processes and as artifacts. Cope and Elwood state that Qualitative GIS can be understood as the "multiplicity of engagements" between GIS and qualitative forms of data and analysis [39, p. 4]. In this section, we investigate the multiplicity of engagements between visualizations and qualitative research.

3.2.1 *Using Visualizations to Generate Qualitative knowledge*

Among the engagements we study in this chapter, the use of visual methods to generate qualitative knowledge is the most prominent and most studied topic in literature. *Visual research* is a field of research that studies the methodology and practice of using visual media either as sources or as supporting methods for data collection and representation in qualitative research. A more general term than *visualization*, a *visual* refers to photographic images, videos, paintings, drawings, and digital images, in addition to visualizations [56]. In visual research, visuals are used for a variety of goals: a researcher might take videos or photographs of an environment as part of gathering rich, contextual evidence in fieldwork [57]; a researcher might collect existing visuals about a topic of interest from newspapers or archives [58]; a researcher might ask participants to carry a camera and collect photographs that are relevant to the subject of the study [59]; and a researcher might use visual media as supplementary artifacts for data elicitation during the interview process. Visual methods of research are used in many qualitative disciplines, including anthropology [60], ethnography [59], psychology [61], and health sciences [62]. The visual captures more context than what is accessible through verbal accounts or textual notes [56]. Additionally, the visual can work as a stimuli for memories and emotions, generating more or deeper insights from participants [63].

As we are interested in visualizations this thesis, we inspect the use of visualizations specifically as a subset of visual research methods. Qualitative researchers use visualizations to elicit feedback from participants in a variety of ways. Comi and Eppler [63] group visual methods into two techniques: (1) projection, where visuals (including visualizations) are used as stimuli to prompt reactions from participants; and (2) facilitation, where the researcher and the participant collaboratively engage in the creation of the visual.

Using visualization as a *projection* technique in qualitative research aims to provoke the memory, establish rapport, and evoke reactions. For example, Figure 7 depicts a poster that was displayed at MS Life Conference 2012 by Yauner and Tennant [64]. The researchers were interested in provoking a debate and collecting the insights of attendees on the subject of "daily living with MS." In lieu of using traditional data collection methods such as questionnaires or surveys (that had garnered little feedback for them in the past), the researchers used this poster as form of "Art/Design Intervention". As a result, the researchers reported a few findings: (a) participants were excited and needed little prompting; (b) they displayed less of the hesitation which often occurs when sharing personal health information; and (c) the intervention yielded 48 written and drawn contributions.

Figure 7 - The World's Tallest Signpost, a printed visualization used to ignite debate and elicit feedback from conference goers (created by Freddie Yauner and Andy Tennant [64])

Using visualization as a *facilitation* technique in qualitative research aims to collaboratively produce a visual account of the participant's experience, while also stimulating their verbal account. In facilitation, the visualization is "drawn, commented and interpreted" during the conversation between the researcher and the participant [63]. This technique is gaining increased interest in the literature of qualitative interviewing, including the use of diagrams [65], concept maps [66] [67], or personal timelines [68] [69]. For example, in Figure 8, Umoquit et al. define the term "diagrammatic elicitation," which they use to describe the application of various forms of charts, maps, and drawings in the data collection process [65]. In diagrammatic elicitation, the researcher prompts the participant to construct a concept map as they respond to the interview question (participant-led diagramming); alternatively, the researcher herself can build the diagram as she listens to the participant's response (researcher-led diagramming). The collaborative construction of diagrams is described as a graphical representation of the participant's experience [67] [65]. As the participant responds to questions, they recall their experiences, plot memories or ideas, and draw links between them. In addition, they are able to supplement their verbal accounts with graphic abstractions, which may aid the presentation of complex ideas [65]. Further, facilitation can provide a visual, shared understanding between the interviewer and the participant of the topics covered and therefore guide the interview toward more relevant directions.

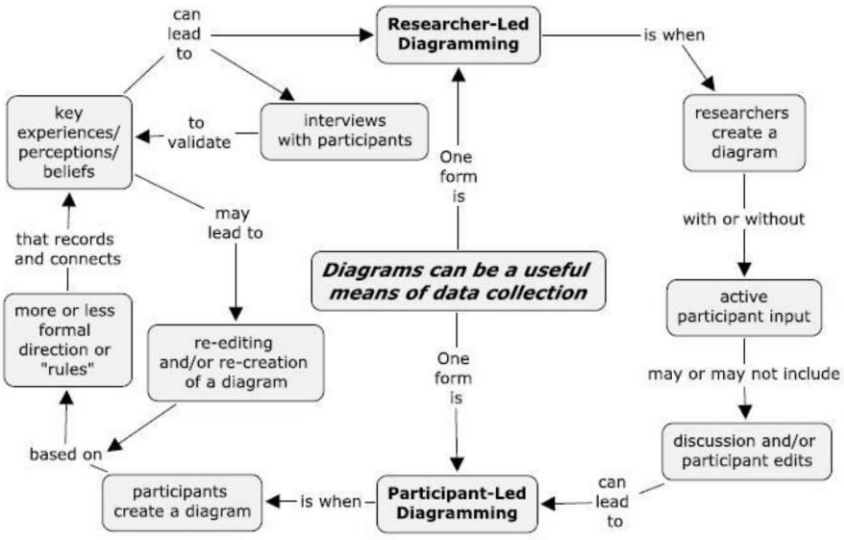

Figure 8 - A concept map describing the use of diagrammatic elicitation in qualitative interviews (created by Umoquit et al. [65])

We note that as visualizations (projective or facilitative) support the data elicitation process, visualizations also become part of the data. They become "repositories of emergent knowledge" [63] collected using a qualitative approach. As such, visualization as an elicitation technique also has an analytical advantage in that it can offer an organizing framework for analysis [63] or a record of possible codes and themes for Grounded Theory analysis [66].

3.2.2 *Visualizing Qualitative Data*

Unlike the visual methods for qualitative knowledge generation, the literature on visualizing qualitative data remains less voluminous [11] [9]. There are two sides to this statement: (1) the literature in qualitative disciplines underutilizes visualizations in research publications [9] [11]; and (2) the literature on visualization places an emphasis on quantitative data, with few exceptions that we reference in this section.

In an effort to survey the use of visual displays in qualitative research, Verdinelli and Scagnoli reviewed 784 articles from three refereed qualitative research journals published by SAGE Publications. In particular, they reviewed articles published between 2007 and 2009 in the Qualitative Health Research journal, the Qualitative Inquiry journal, and the Qualitative Research journal. Their review focused on visual displays as defined by Miles and Huberman: "an organized, compressed assembly of information that permits conclusion drawing and action" [18, p. 11]. This definition is less strict than that of quantitative visualizations, which aim to provide a mapping between data and visual encoding [11]. Nonetheless, Verdinelli and Scagnoli found that only 27% of the articles they reviewed contained any form of visual display. Furthermore, 60% of the visual displays they found were in the form of tables. This low level of representation is intriguing, particularly because some qualitative methods, such as Grounded Theory, encourage the use of visual displays in the process of analysis [70]. This finding, in part, motivated our study in Chapter 4.

Onwuegbuzie and Dickenson published a series of articles to address the imbalance of visualizations towards statistical data [11] [19]. Onwuegbuzie and Dickenson assert that graphical methods can be used for both quantitative and qualitative data. Given this assertion, they present a taxonomy that aims to help researchers in selecting a visualization that is most effective in communicating their data. The taxonomy is based on Tufte's framework of visual representation, which contends that data can be represented on three levels of abstraction: sentences, tables, and graphics [12]. Onwuegbuzie and Dickenson map qualitative data to Tufte's quantitative framework by proposing to *quantify qualitative data*. Quantification is done by taking codes or themes of qualitative analysis, counting them, and mapping them to categories in quantitative visualizations. This approach is appropriate but limited to quantifiable data.

3.2.3 *Using Visualizations for Qualitative Analysis of Data*

In the previous sections, we presented two types of engagements between visualizations and qualitative research: using visualizations to collect qualitative data and using qualitative data to construct visualizations. In this section, we show how visualizations engage with qualitative research in a bidirectional manner to support the activities of data analysis. In the literature of qualitative research, the term *analysis* is a hypernym that groups together a variety of activities such as coding, theme derivation, interpretation, sense-making, theory construction, and conclusion drawing [70] [16] [71] [17]. The process of creating a visualization, when designed or generated manually, drives the researcher to think about, organize, and draw connections between elements from the context of their data collection [17] [70] [18] [16]. The resulting visualization then enables the researcher to see their data, observe patterns, detect salient themes, and visually compare data points across different metrics.

For example, Figure 9 shows a custom visualization designed and implemented by Hiller [72], who generated the visualization using biographical interview data. Hiller improves on the concept of timeline, a chronological visualization which is often used as a data collection and analysis device in biographical interviews [68], by including multiple layers of categories depicting various aspects of the interviewee's personal life (e.g., education and occupation), in addition to major historical events that are relevant to the context of the interview (e.g., in this case, political activism). Hiller notes that, although the contents of this chronological chart are objective (i.e., free from interpretation beyond the interviewee's own statements), it elucidated his process of interpretation. The parallel representation of personal and historical events along a common timeline enabled Hiller to observe and make sense of temporal connections between different events globally (across categories) and locally (within one category). In addition to

drawing connections, the construction of the visualizations and subsequent analysis allowed themes to emerge on Hiller's topic of interest (i.e., the intersection between the personal and the social history) based on the juxtaposition of different data points. Hiller's visualization illustrates the bi-directionality and ongoing iteration of feedback between visualization and qualitative research: (a) qualitative interview data was used to generate the visualization; (b) the visualization assisted in generating themes and insights for his qualitative research questions; and (c) the resulting insights could be added to the same visualization or reported elsewhere.

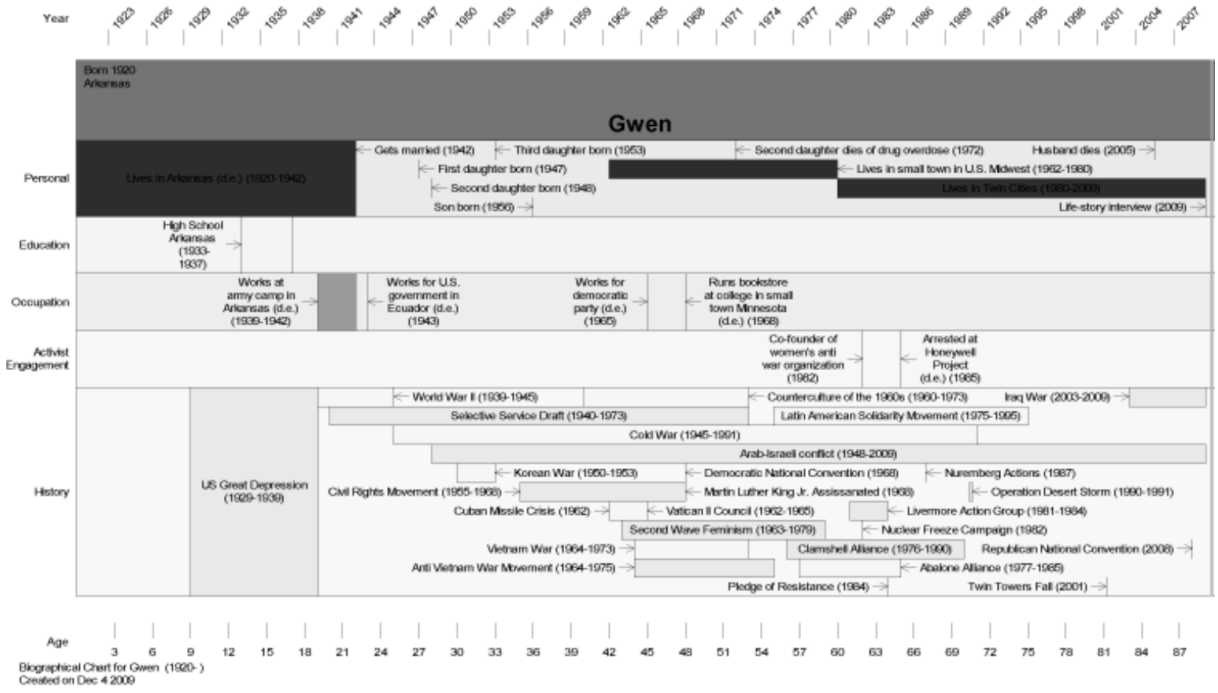

Figure 9 – Multi-layered chronological charts (created by Patrick Hiller [72])

Similarly, examples from literature illustrate how visualizations can aid the activities of analysis in a variety of ways: diagrams and concept maps are used to *conceptualize and construct themes* from qualitative data in grounded analysis [16] [70] and in historical research [73]; maps and floor plans are used to *navigate and make sense* of qualitative data within spatial context [74]; word clouds are used to *explore* textual content [75]; tree structures and scatter

plots are used to *detect salient patterns* among data points [17]; and network maps are used to *represent or discern relationships* among clusters of categorical nodes [76].

Furthermore, Qualitative Data Analysis (QDA) software tools provide many visualizations that are generated using primary data (e.g., interview transcripts or audio recordings) or analysis data (e.g., analytic codes or memos) [77]. Some tools offer visualizations that are designed for specific use case scenarios; such as ChronoVis which provides parallel time-series that organize custom analysis annotations with media content [78] [79] optimized for observational research using video analysis [80]. Most well-known QDA tools (such as Atlas.ti, Dedoose, and NVivo) feature common visualizations. Consider the following three examples. Figure 10 shows the Code Co-occurrence table in Atlas.ti (available in other QDA tools as well). Each cell in this visualization plots and color-codes how frequently the two corresponding analytic codes appeared in close proximity (proximity is calculated differently in different tools) within a selected collection of data. This visualization enables the researcher to discern connections between analytic codes and scan for possible patterns of emergent themes. Figure 11 shows a Packed Code Cloud in Dedoose, which is an implementation of a word cloud that aggregates content based on the analytic codes researchers assign to text, rather than the text itself. This visualization enables the researches to explore their codes and discern their frequency. Figure 12 shows a project map in NVivo [81] (a variation of it is termed a Network View in Atlas.ti). The map is manually created by the researcher to depict various elements from their analysis context (e.g., field notes, memos, and codes) and to plot and annotate the connection between them. This enables the researcher to construct a visualization that corresponds to their mental map of the research project.

	Education Ref	Enthusiasm fc	Goals	Math	MOTIVATION	Motivation: Ec	Motivation: M	Prospects	RESERVATION	Reservation: C	Skills needed	Teach for Am	Teacher equal	why join TFA
Education Reform		n/a	1 - 0.50	n/a	1 - 0.25	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Enthusiasm for Teaching	n/a		n/a	2 - 0.40	1 - 0.17	n/a	n/a	n/a	n/a	n/a	n/a	n/a	1 - 0.20	1 - 0.25
Goals	1 - 0.50	n/a		1 - 0.20	1 - 0.20	n/a	n/a	1 - 0.33	n/a	n/a	n/a	n/a	n/a	n/a
Math	n/a	2 - 0.40	1 - 0.20		n/a	n/a	n/a	1 - 0.20	n/a	n/a	n/a	n/a	1 - 0.17	1 - 0.20
MOTIVATION	1 - 0.25	1 - 0.17	1 - 0.20	n/a		1 - 0.25	1 - 0.20	n/a	n/a	n/a	n/a	1 - 0.25	1 - 0.17	1 - 0.20
Motivation: Educational D	n/a	n/a	n/a	n/a	1 - 0.25		n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Motivation: Moving Away	n/a	n/a	n/a	n/a	1 - 0.20	n/a		n/a	1 - 0.25	1 - 0.50	n/a	1 - 0.50	1 - 0.25	n/a
Prospects	n/a	n/a	1 - 0.33	1 - 0.20	n/a	n/a	n/a		1 - 0.25	n/a	n/a	n/a	n/a	n/a
RESERVATION	n/a	n/a	n/a	n/a	n/a	n/a	1 - 0.25	1 - 0.25		1 - 0.33	n/a	n/a	n/a	n/a
Reservation: Changing Pla	n/a	n/a	n/a	n/a	n/a	n/a	1 - 0.50	n/a	1 - 0.33		n/a	n/a	n/a	n/a
Skills needed	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a		n/a	n/a	n/a
Teach for America	n/a	n/a	n/a	n/a	1 - 0.25	n/a	1 - 0.50	n/a	n/a	n/a	n/a		1 - 0.33	n/a
Teacher equality	n/a	1 - 0.20	n/a	1 - 0.17	1 - 0.17	n/a	1 - 0.25	n/a	n/a	n/a	n/a	1 - 0.33		n/a
why join TFA	n/a	1 - 0.25	n/a	1 - 0.20	1 - 0.20	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	

Figure 10 – A Code Co-Occurrence table in Atlas.ti

Figure 11 – A Packed Code Cloud in Dedoose [82]

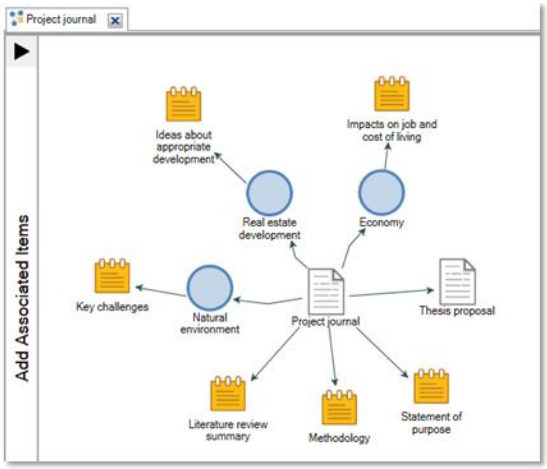

Figure 12 – A Project Map in NVivo [83]

In our survey of visualizations and qualitative data analysis, we observed that the visualizations used to support analysis in literature *vary in type*: some of the visualizations were defined by the researcher for their specific use case (e.g., the chronological chart in Figure 9), while others were commonly used visualization types that are generated using qualitative data

(e.g. the visualizations offered by QDA tools). Additionally, the variation includes visualizations that are traditionally considered quantitative (e.g., bar charts, time series, and scatter plots). We do not observe common characteristics among those types that we can attribute to the qualitative-ness of the analysis task.

Our survey yielded fewer examples of visualizations than we expected, especially given that many qualitative analysis books encourage the use of visualizations in the process of analysis [16] [70] [18]. This low quantity could partially be due to the focus of final publications (e.g., articles or research papers) on reporting results rather than elaborating on the process of analysis, yet visualizations also can be useful for reporting methods or results, especially if there are interesting non-linear relations. Therefore, most of the examples we list in this section were found in method papers, or in books that promote specific visualizations and explain their benefit to the process of qualitative analysis.

Given the observations we noted in this section, we wonder: Why do qualitative studies, including this thesis, not seem to use many visualizations? Is it possible that, if we had access to a larger variety of visualizations used in qualitative research, we would be able to discern some common qualitative characteristics *within* these visualizations? These questions partially motivate our inquiry in Chapter 4, where we interview qualitative researchers to learn about how they use visualizations.

Based on our survey of literature, we organize the engagements between visualization and qualitative research into the model in Figure 13. We note that (a) the process of visualization can be used as a projection or facilitation technique to generate data for qualitative research; (b) visualizations can be used to display data resulting from qualitative research; and (c) in the activities of qualitative data analysis, knowledge creation is a bidirectional process where

qualitative data are used to generate visualizations, which are, in turn, used to generate qualitative insights that support the process of analysis.

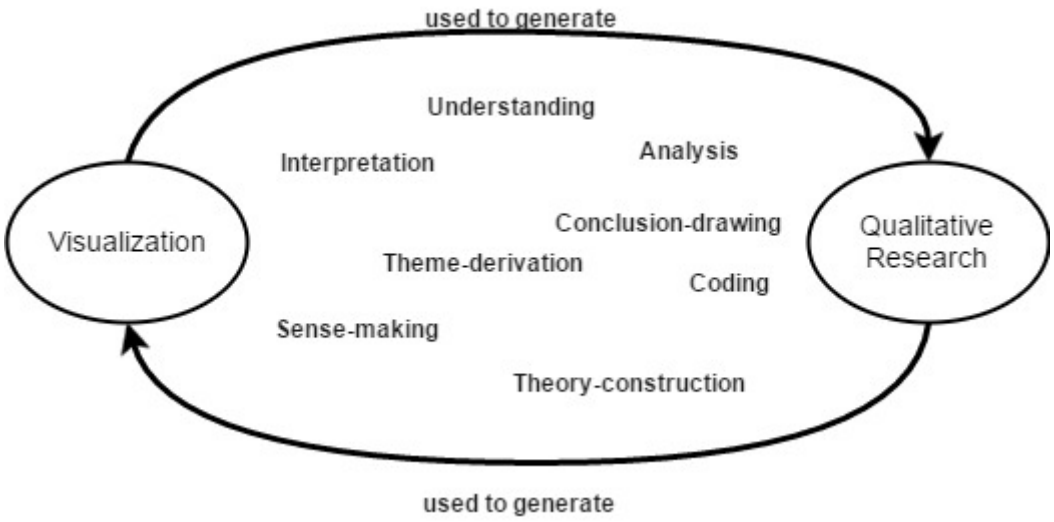

Figure 13 – A model representing the multiplicity of engagements between visualizations and qualitative research.

We conclude:

The conceptualization of Qualitative GIS as defined by Cope and Elwood [39]*, provides a theoretical framework for our exploration of the literature. Given this framework, qualitative-ness in visualization can be understood as a feature of the multiplicity of engagements between visualization and qualitative research. We structure those engagements into the model presented in* Figure 13*.*

Chapter 4. VISUALIZATIONS IN QUALITATIVE RESEARCH ANALYSIS

4.1 BACKGROUND

This thesis asks: How might we characterize qualitative approaches to visualizations? In the previous chapter, we explored this question within literature. We learned from the example of GIS scholars, particularly the movement of Qualitative GIS, and we used their definition to characterize and organize the "multiplicity of engagements" between the literature of visualizations and the literature of qualitative research.

To describe qualitative approaches to visualizations in the context of engagements between visualizations and qualitative research, however, is to limit these approaches within the discourse of qualitative research. We do not wish to constrain our characterization to qualitative research alone. Rather, we foresee the application of qualitative approaches to visualizations to be ecumenical and independent of any research method.

As we explore qualitative-ness outside of the literature of qualitative research, we conjecture that we might gain insights from researchers who use qualitative methods.

We ask:

How do qualitative researchers use visualizations? What can we learn from their patterns of use? How might we use the insights we gain to characterize qualitative approaches to visualizations?

4.2 METHODS

In addressing the intellectual puzzles, our target population were *researchers* who meet the following criteria: (1) using qualitative methods of any type (e.g., interviews or participant observation) and within any discipline (e.g., psychology, sociology or nursing); (2) employing visualizations of any form or medium; and (3) using visualizations to inform the analysis activities of their work. We constrained our sample to the *analysis* phase because we wanted to limit responses from researchers who use visual methods of data collection (Section 3.2.1), a topic which has been sufficiently covered in literature.

Our primary method for collecting evidence was through conducting *qualitative interviews* [84] [85] with our target population. The interviews were (a) semi-structured, thematic, and relatively informal; (b) lasted 60 minutes or less, depending on the availability of the subject, and (c) conducted on-site, taking place at the researcher's place of work whenever possible (otherwise, interviews were conducted using video-conferencing).

In our recruitment process, we used the following *channels*: (a) circulating a call-for-participation advertisement using mailing lists that serve academics in disciplines that use qualitative research methods (e.g., the QUAL concentration mailing list at the University of Washington); (b) personally contacting researchers who use visualizations in their published works and whose organizations are located within Seattle area, and (c) speaking with representatives (e.g., program coordinators) from different programs and schools at UW to inquire about possible study subjects and ask for referrals within their department.

We used interviews to generate evidence from three *data sources*: (1) the researchers' own accounts of their experiences of using visualizations, (2) the environment where they

conduct their analysis, both physical (offices or labs) and digital (e.g., QDA software), and (3) documents and references provided by the researchers either before or after an interview.

The first source of evidence is the researchers' own accounts. Throughout the interview process, we attempted to gather an account of the researchers' experience in using visualizations as part of their analysis activities. In doing so, we first worked to situate our conversation within the context of their research, inquiring about the topic they study and the qualitative methods they use. We then steered the conversation into the topic of visualizations, focusing initially on generating details about *what* they use and *when* and *why* they use it. This includes a survey of the varieties of visualizations used (colloquial or researcher-defined), the medium of analysis and display (computer-assisted or paper-based) and the types of data abstracted (e.g., textual or multimedia). It also includes an inquiry into the specific goals of using a visualization and the research activity within which it was commissioned.

The most delicate and complex part of the interview process was generating an account of *how* visualizations are used. The complexity stems from its focus on the *experience of an interaction*, which often is not immediately memorable, unless it was in some way outstanding. The challenge in this part was to revive the memory of participants. We attempted to do so in a number of ways: (a) by conducting the interview at the participant's place of work, which could provide cues that refreshed their memories; (b) by gently persisting and asking the same question in a variety of manners at different points in the interview; and (c) if the participant is using a QDA tool, we asked questions about experiences specific to the visualizations provided by this tool.

The second source of evidence is the environment in which researchers conduct their analysis. In the call-for-participation, we asked researchers to meet at their place of work (if this

was possible and convenient). In addition to prompting memories, it allowed the participant to reach for and present pieces of evidence in their environment, which might not have been accessible otherwise. For example, as we were holding our first interview, the participant frequently left their desk to show us a quotation from a book on their shelf or to discuss a visualization that they hung on their wall. Of interest as well was the digital environment contained in the researchers' work stations. In five out of seven interviews that we conducted in participants' offices, the participants opened their personal computers, switched the monitors in our direction and clarified many of our interview conversation with examples and references.

The final source of evidence in our inquiry comprised references provided by researchers before or after an interview. In the call-for-participation, we had asked the researchers to provide us with references that they deemed important for us to read, in order to become familiar with their topic of research. Our goal in making this request was to be more efficient with the finite time (60 minutes) we had with the participant by acquiring advance knowledge of what they had previously written or published. Preparation allowed us to customize interview questions to be relevant to the specific case of the participant, referring to examples from their own work. For example, in one case, we learned that the participant has designed their own interactive visualization tool. This allowed us to ask specific questions about why they made specific design choices. Additionally, in three cases, references mentioned the QDA tool that the researcher uses. In the cases when we were familiar with the tool (e.g., ATLAS.ti), we were able to ask specific questions about the visualizations it provides (e.g., Network Views and Co-occurrence tables). In the cases when we were not familiar with the tool (e.g., Dedoose.com), we were able to research it before the interview.

4.3 ANALYSIS

Our sample consisted of 7 interview participants. They varied in their areas of research, their choice of qualitative methods, their choice of visualizations to use, and their choice of QDA software. Table 1 presents the variety of our sample across a number of attributes of our participants and their research. For every column representing a participant, the (●) symbol indicates that the corresponding attribute in that row applies to the participant.

Table 1 – Attributes of participants and their research

Categories	Attributes	P1	P2	P3	P4	P5	P6	P7	T
Participants' Field of Research	Geography	●							1
	Information Management		●						1
	Nursing			●					1
	Education				●		●		2
	Political Science					●			1
	Software Engineering							●	1
Academic Position	Professor	●					●	●	3
	Senior Research Analyst				●				1
	PhD Student		●	●		●			3
Research Approaches	Narrative research						●		1
	Mixed-methods	●			●	●			3
	Grounded Theory		●						1
	Visual methods	●	●				●		3
Qualitative Methods	Interviews	●	●	●	●	●	●	●	7
	Participant observation				●	●		●	3
	Archive Analysis					●			1
QDA Software	Atlas.ti	●	●		●	●			4
	Dedoose		●			●	●		3
	Excel								1
Visualization Medium	Digital	●	●	●	●	●	●	●	7
	Paper-based	●	●		●	●	●	●	6
	Whiteboard				●			●	2
Gender	Female (4)		●	●	●	●			4
	Male (3)	●					●	●	3

We proceeded in our analysis guided by grounded approaches to theory development [70] [16]. In using grounded approaches, we attempted to construct meaning in an inductive manner:

we were not attempting to confirm or prove some presupposed theory; rather, we were open to themes that emerge through our analysis of the interviews and the range of evidence that they produce.

Our analysis involved the following activities: (1) we initially transcribed all interviews as an exercise of active and close listening to the evidence; (2) we openly coded all interviews denoting concepts that stand for meaning [16] [86]; (3) we wrote memos where we analyzed in writing; and, finally, (4) we performed a second round of focused coding as guided by initial findings. We provide a record of our major themes and some of their associated excerpts in Appendix A.

4.4 FINDINGS

The questions of our interviews (as we described them in Section 4.2), were broad in nature, aiming to explore how qualitative researchers use visualizations by inquiring about *what* visualizations they use and *when* and *why* they use them. Table 2 lists the visualizations our participants shared or talked about during our interviews. We note these visualizations vary in type: (a) they include common visualization types (e.g., bar charts) in addition to researcher defined visualizations (e.g., sketches and collages); (b) they include spatial (e.g., geo-visualizations or heat maps), temporal (e.g., timelines), and social (e.g. network view or interaction diagram) representations; and (c) they depict data produced during analysis (e.g. analytic codes) in addition to primary data (e.g. text from interview transcripts).

Nonetheless, the themes that emerged from our analysis did not pertain to characteristics *within* visualizations; rather, we observed several themes regarding our participants' *approach* to visualizations (i.e., *how* the participants used visualizations). Table 2 lists three major themes that we observed (knowing the data, synthesizing and explaining, and reaching for more) which

are illustrated in the following sections with examples and excerpts from our interviews, in which the letter P, followed by a number, represents participants in our interviews. The (●) symbol indicates the presence of a theme in relation to the participant's approach to the corresponding visualization in the table. The table shows the distribution of themes across visualizations and participants. Note that those themes were present in 6 out the 7 interviews we conducted.

Table 2 – Distribution of visualizations and themes

P	Visualization	Knowing the Data	Synthesizing and Explaining	Reaching for More
1	Network view		●	
	Geo-visualization of census Data			
	Multimedia geo-visualization	●	●	●
	Geo-visualization of racial distribution			
	3D heat map	●	●	
	2D heat map		●	
	Geo-visualization of transit connections	●	●	●
	Cartogram			
	Collage of map, visual, and visualization			
	Geo-visualization of crime rates	●	●	
	Code cloud			
	Word cloud of geo-tweets			●
2	Timeline sketches	●	●	●
	Affinity diagram		●	
	Code co-occurrence		●	●
	Word cloud			
4	Word cloud			
	Code co-occurrence	●	●	●
	Network view			
	Tableau (various visualizations)			
	Dynamic map of research results			
	Flow chart			
5	Network view			●
	Code distribution		●	
	Code co-occurrence		●	
	Code primary document table	●		
	Concept diagram			●
6	Code co-occurrence		●	●
	Code presence table	●		●
	Code cloud		●	●
	Concept diagram	●		
	Line graph		●	
	Bar chart		●	
7	Interaction diagram	●	●	●

4.4.1 *Knowing the data*

Participants used visualizations to represent a wide *variety* of data during the process of qualitative analysis. The data was collected by participants using different methods, including interviews, participant observation and archival analysis. Participants stored and analyzed in multiple formats, which included collections of textual transcripts, audio recordings and images from the field. The collections ranged from records as small as 13 interviews to "hundreds of interviews and thousands of data points" (P4). Participants used a variety of descriptors in reference to their data collections, pertaining to size, complexity and privacy.

Despite this variety, participants consistently (in 6 out of 7 interviews) professed knowledge of many of the data points in visualizations. Notably, participants knew and remembered the sites where they collected the data, the subjects that they studied, the contexts within which words were uttered and actions took place, the questions that framed their inquiries, and the data points that emerged later in the course of analysis. As the participants showed me different visualizations of their data, they drew from these memories and connected elements in the visualization to recollections of the data.

Participants were close to the data as they *collected* the data. Using qualitative methods, the researchers collaborated with their subjects to generate data about their topic of interest. This is illustrated in P2's recollections about her interview data. P2 studied patients' experiences as they went through the timeline of their illness and recovery. She asked her interview subjects to draw a sketch of their journey as they answered her questions. P2 was describing the visual artifacts her interview subjects produced when she remarked:

> The other thing about these visual artifacts; it gives me very vivid memories of the interview. So, this one was with an artist and she was very bohemian. … She is very

> funky and laid back. We sat on the floor while she did this because she was more comfortable sitting on the floor. So she sat on the floor, she had the markers spread around. I think the vividness of it because I am so, I feel so close to my participants and protective of them a little bit. (P2)

P2 had conducted around 18 interviews, from which she retained 15 visual artifacts. Yet she remembered the actions and reactions of her interview subjects, their personal stories, and the visual encodings they chose to represent elements from their journey. When she showed me the artifacts, she narrated them with explanations of what they represent. For example, she explained how one subject used dark colors to represent negative emotions, another used circles to represent people, and another drew a rising line to represent an uplifting moment in their life. P2 *knew* and *remembered* the data that she collected.

Participants knew the data as they were *immersed* in the field where data were collected. In general, researchers using qualitative methods devote considerable time to studying the context surrounding their scope of inquiry, in order to be able to draw descriptions and interpretations. P5, for example, used a mixture of methods to study the democratic practices of community organizations. Her data collection included a hundred interviews and extensive notes taken during her observations in the field. Further, following her qualitative fieldwork, P5 circulated a survey on the same topic. She notes:

> Sometimes people filling in the surveys were actually people who I'd interviewed. I was able to infer, because it's clear from the survey who they are and what organization they're with and I know that I interviewed someone from that organization. (P5)

P5's extensive work in the field allowed her to draw connections between two different data sources. She was able to infer the identity of some of the participants she interviewed based on their survey responses. Despite the relatively large number of her interviews, she maintained knowledge of her interview subjects and their unique characteristics. She *inferred* and *remembered* her data because of her immersion and extensive work in the field.

Participants knew the data because they *analyzed* the data for prolonged periods of time. They conduct a variety of analysis activities iteratively, including transcribing audio recordings, coding data, writing analytic memos, and discussing themes with collaborators. These activities drive researchers to pay close attention to both the contextual richness and the individual distinctions within their collections. Knowledge of data also became evident when participants showed me visualizations that were based on their analytic codes. For example, P5 showed me the Code Primary Document table, a table that maps analytic codes as rows and interviews (where these codes were applied) as columns. P5 noted:

> I even just remember what the interviews were because I've spent so long with them that I can look at that Code Primary Document table and I know what every one of those columns is. I can remember who that column is, and so to some extent when I see that they flagged a public-government learning I can often remember what the thing they were describing was. Because that's one of the things that comes from translating your own interviews. You just get to know them really well. (P5)

P5 spent a long time with the data: she transcribed all hundred interviews, analyzed the transcripts (primary documents), assigned analytic codes to the contents of those transcripts, and iteratively refined those codes. With the thoroughness of analysis, when P5 looks at the Code

Primary Document table, she is able to discern the relevant content corresponding to a row or a column. P5 *recognized* and *remembered* the data because she analyzed them rigorously.

In the previous examples, we observe that participants demonstrated knowledge of the data that lies within or surrounds the visualizations used in the analysis process. This is exhibited in (a) their vivid recollections of the context and circumstances of data generation; (b) their synthesis of knowledge from different data sources; and (c) their ability to remember data sources given visual clues.

These exhibits, consistent in six out of seven interviews, lead us to describe *knowing the data* as a characteristic of qualitatively approaching visualizations. We understand knowing data to be the state of recognition, awareness, and familiarity with data, in part or as a whole, that enables the beholder to remember, comprehend, and make connections among data points or data sources. We intentionally adopt a broad understanding of data, encoded within a visualization or in connection to it, which encompasses any type or medium.

The condition of knowing the data can be understood both as a requisite and a byproduct of qualitative analysis. Qualitative research "aims to produce rounded understandings on the basis of rich, contextual, and holistic data" rather than "charting surface patterns, trends, and correlations" [4, p. 4]. To achieve rounded understanding and sound explanations, researchers must be thorough in their analysis, dissection, synthesis, inference, construction, and evidencing. All these activities require, and produce, intimate interactions and exchanges with the data. Additionally, qualitative research sampling features a balance between breadth and adequacy, seeking sample sizes that are small enough to be comprehensible and large enough to support validity within their scope [70] [87]. The richness of the data, the rigor of the analysis, and the accessibility of the sample size foster recognition, awareness, and familiarity.

Qualitative researchers refer to a condition of "closeness to data" [88, p. 73] that they desire to accomplish and maintain [89]. This statement is described to have many possible meanings [88] [89], including (a) the evocative nature of data which produces memories of context (e.g., reading field-notes could remind the researcher of the smells in the field); and (b) the tactile experience of constant interaction with physical or digital analysis artifacts (e.g., writing codes on the margins of a transcript). Both of those exemplary meanings indicate the recognition, awareness, and familiarity that we characterize as knowing the data.

When participants interacted with visualizations, they expressed knowledge and remembrance. To the participants, the visual encodings in a visualization were not mysterious data points; rather, they were familiar components of a familiar context. We stress that this perceived knowledge is not necessarily encompassing of *all* elements of a visualization but is indicative of *some* degree of knowledge within and surrounding a visualization.

4.4.2 *Synthesizing and explaining*

As researchers shared their visualizations, they read and interpreted these visualizations in a manner that connected their components to other elements within the context of analysis. P6, for example, studied the personal narratives of high-school dropouts. P6 and his collaborators used Dedoose as a QDA software and utilized the many visualizations it provides. P6 noticed that Dedoose had recently added a type of visualization that he had not previously used: a Packed Code Cloud (see example in Figure 11).

When asked about what he saw in the code cloud, P6 reviewed the visualization and said: "It kind of reflects what we heard [in the interviews]." P6's statement reports his general reading of the graph. Having analyzed the interviews thoroughly (and authored a paper and a

forthcoming book based on them), P6 concurs that the visualization is not in disagreement with what he knows about the interviews ("what we heard").

P6's reading of the visualization did not terminate at this general statement. He inspected the visualization further and interpreted the implication of the visual encodings. "Elementary" and "Secondary" were the two largest words in the code cloud, indicating that they were the most common analytic codes applied to the interviews. P6, however, appeared to be careful not to attach an analytic value to this high frequency by merit of size alone: "Of course, elementary and secondary are big because our questions and their answers had to do with those time frames." P6 explained that the interview questions asked specifically about subjects' memories in elementary and secondary school; consequently, they were more frequently coded into the transcripts. Rather than glossing over what the size indicated, P6 used evidence from the context of analysis (the interview questions and how they were coded) to explain why and how the major elements were derived. Similarly, noting that the code "Pregnancy" was relatively small, P6 explained: "Obviously, that'd have to be small, because at least half of our participants couldn't have gotten pregnant. Just there's that." P6 synthesized the low frequency of "Pregnancy" as a code with the context of the research sample, which was split evenly between males and females. This synthesis of visual encoding with context can explain why P6 did not dismiss pregnancy as a possible theme. Rather, through different means of analysis, pregnancy emerged to be one of 6 main narratives to be featured in P6's book.

Further, P6 connected and synthesized what he saw in the visualization to what he uncovered in the literature review and to the findings of his own research. P6 noted:

> What I like is that the things that tend to get emphasized by popular media anyway are as small as they should be. … Mental health, drugs and alcohol, attendance, behavior

> difficulties; those are not huge. But truancy is [large]. Because certainly students skipped and we have lots of, there were lots of patterns of skipping that we saw. (P6)

P6's analysis of the subjects' personal accounts had showed that truancy was a recurring pattern, which led to dropping out of high school more so than other issues popularized by the media. P6 viewed the visualization within the context of his findings. In this case, he agreed with the analytic significance of truancy as encoded with a larger size, because he could provide contextual interpretation for this significance through his previous literature review and analysis.

We observe that, as P6 read the visualization, (a) he understood the visual encodings (relative size of the words in this case) and used them to provide a general reading of the visualization; (b) he then interpreted the visual encodings and evaluated their significance in light of his previous analysis; and (c) he connected an synthesized his reading of the visualization with other data from the context of analysis. P6 interpreted the visualization as he read it, providing explanations that combine evidence from multiple data sources.

This pattern of reading, connecting, synthesizing, and explaining is also consistent among participants (in 6 out of 7 interviews). Rather than focusing on *what* a visualization represents, participants interpreted visual encodings in conjunction with other sources of meaning in order to provide explanations for *how* patterns relate to context or *why* they are significant [10]. This pattern is also consistent with the characteristics of qualitative analysis in general, which involves series of interpretation and sense-making activities [17]. In other words, the data deduced from visual encodings in a visualization are, in turn, qualitatively analyzed. Participants critically read the visualizations and figuratively asked "How does this visualization fit with other sources of knowledge?"

4.4.3 *Reaching for more*

The most frequently recurring theme observed among participants was that, as they shared their visualizations, they talked about how they usually *follow up on clues* provided by a visualization by *studying the data underlying* the visualization or connected to it. We observe that when participants interact with visualizations for different analysis goals, they reach for more data.

P2, for example, uses co-occurrence tables to *explore* interview data with her collaborating team. Co-occurrence tables are visualizations, afforded by many QDA software, which plot the correlation among analytic codes by representing (in color-coded cells) how often they occur in close proximity in primary documents (see Figure 10). In the initial stages of analysis, P2 and her collaborators code primary documents individually. They then meet and use Dedoose to generate the co-occurrence table of their most common codes, and they review the table to search for interesting patterns. P2 describes this process:

> I'll often start with the co-occurrence table when I meet with the coding team. And then you can double-click where you see no co-occurrence or you see lots of co-occurrence. … I'll put them up and I'll say 'what strikes you?' Like 'what strikes a chord with you? One; and two, what surprises you about this visualization?' (P2)

In the case of P2 and her coding team, when they see codes having a high co-occurrence rate or no co-occurrence, they consider them to be patterns of interest. When the team determines patterns of interest, they "track [them] down" to explore them further. P2 remarks that the table in Dedoose is "pretty; but also, it gives you things to zoom in on really fast. … We would, sort of, dive into it. We would double-click on things."

When the team clicks on the cell they find interesting, Dedoose produces a list of all the excerpts where the two codes corresponding to the cell co-occur. The team studies the excerpts: they review the context and determine whether an excerpt is "coded right"; they discuss the "overlap" between codes and provide potential interpretations for why it occurred; and they debate whether the pattern creates a theme. While the table initially provides the team with co-occurrence as a clue for possible patterns, the team has to reach for more data to explore further.

Participants also reported reaching for more data as part of systematic *analysis* activities beyond the exploration phase. P6, for example, described using the Code Presence Chart in Dedoose to analyze themes both broadly, across different subjects' accounts, and specifically, within one subject's account. The Code Presence Chart maps analytic codes as columns and primary documents as rows, such that a colored cell indicates the presence of a code within a document, and an empty cell indicates its absence. P6 describes looking at the chart and analyzing the theme of student retention:

> Those are the students who talked about retention. Who were those students and what did they say? So then we look at those excerpts. Then it's either, so you have two ways: One is to pull all of those excerpts with that code to see across the group. … The other is to look at one of those transcripts and to read through it to see where the retention occurred and what it was like for that individual. (P6)

By looking at the chart, P6 is initially able to determine the commonality and distribution of retention among his interview sample. But in order to analyze retention as a theme, he has to investigate it holistically and contextually. Accordingly, P6 uses the chart to reach for more data. He either extracts all the excerpts coded with retention to analyze "the group and the trend" or he

extracts a single transcript to analyze "the individual story". The visualization functions as a starting point, but P6 has to reach for more data to analyze further.

Participants also reported using visualizations to error-check or *debug* their process of analysis. P4 and her team, for example, found that one of their codes is associated with 1,500 excerpts of text. This large number raised their concern, because it could indicate that the code itself is underdeveloped or that the assignment of it was not carefully applied within the primary documents. P4 describes how they used the co-occurrence tables in Atlas.ti to debug this high occurrence rate:

> As it turns out we just have 1,500 elements that don't go with anything else. Make note of those places that I think require deeper dive. … So generally I will talk with [my collaborator] first and we will come up with what the data-mining strategy is, and then yeah, it often involves going back to the primary documents to figure out. (P4)

When P4 looks at the co-occurrence table, she searches for possible correlations between the code and other codes in the repository. When she identifies such correlations, she and her collaborator work on a "data-mining strategy" through which they will investigate those correlations. P4 reports that this "deeper dive" often involves extracting the relevant primary documents (interview transcripts) and "spot-checking" to examine why a code is behaving in this way: "It's co-occurring with X, Y, and Z in the following ways. So let's think about what that code is truly telling us. Is it completely redundant with these three things? Of the places where it's not co-occurring, is that actually telling us something unique?" The visualization provided P4 with possible correlations, but she had to reach for more data in order to debug her analysis.

4.5 Discussion

We set out on our investigation in this chapter, through qualitative interviewing, with the goal of exploring this question: How do qualitative researchers approach visualizations? In previous sections, we presented three consistent themes that we observed when our participants interacted with visualizations. We perceive these themes to be descriptive characteristics of qualitative approaches to visualizations.

Figure 14 presents a model of these characteristics. As researchers read, interacted with, and used visualizations, (a) they expressed closeness and familiarity with the data represented within the visualization and to the data in the context surrounding the visualization (*knowing the data*); (b) they combined their reading of the visualization with other data sources in the context of analysis and they interpreted the visualization in light of this holistic evaluation (*synthesizing and explaining*); and (c) they followed up on points of interest they noted in a visualization, dived deeper, and extracted relevant data from the context of analysis (*reaching for more*).

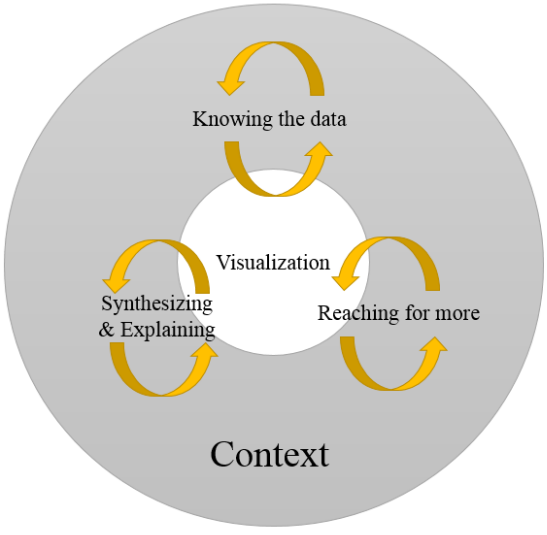

Figure 14 – A model of qualitative approaches to visualizations

4.5.1 *On Context*

In order to study the model in Figure 14, it is necessary to explore what we mean by "context". Context is a term that is widely used, particularly in the literature of qualitative research [90] [91]. Contextual understanding and explanation is often cited as a goal and a distinctive feature of qualitative methods [4] [3]. However, in his article "What we Talk about When we Talk about Context," Paul Dourish remarks that "context is a slippery notion. Perhaps appropriately, it is a concept that keeps to the periphery, and slips away when one attempts to define it" [92, p. 35]. Part of the trouble of defining context is that "by nature, it's about everything" [93, p. 3]. For example, it's a term that can be used to describe social situations as well as books or digital documents.

For our purposes in this model, we learn from Dourish's approach, i.e., we do not attempt to restrictively define context; rather, we "ask what work the term is doing as it is used" [92, p. 35]. In this sense, our understanding of context has four "works" (i.e., functions):

(1) Context indicates relationships [93]. It implies that a thing does not exist in isolation, but is connected to many elements outside its scope. It also implies that the meaning of a thing is interpreted and realized in connection to outside elements. So, as we witnessed in our interviews, a visualization is interpreted in juxtaposition with the larger scope of analysis.

(2) Context includes many sources of data. In our interviews, participants synthesized the visualization with data from literature, primary documents (e.g., transcripts), a research sample, or recollections from fieldwork.

(3) "Context implies boundedness" [88, p. 75]. While context as a term has broad applications, it intrinsically implies an understanding of locality. A word might mean

something in one context, but have a completely different meaning in another. In our interviews, participants interpreted one visualization in different contexts (e.g., understanding retention in the context of the group trend vs. in the context of an individual narrative).

(4) Context is generally larger and richer with data than any single element that it contains. That is, the context of a visualization is richer with data than the visualization itself.

4.5.2 *A Visualization as an Abstract View of Context*

Visualizations are visual representations of data [37]. This definition may include one or more activities of abstraction: (a) data points are selected and removed from a context (e.g., analytic codes and primary document titles are chosen and extracted from the analysis environment); (b) data undergoes some form of statistical modeling (e.g., the frequency of occurrence of an analytic code in a document is computed); and (c) the resulting model is mapped to an image (e.g., analytic codes, primary documents, and frequency are visually encoded as a matrix).

A visualization, by this definition, is an abstraction of a richer context. The context of analysis is typically larger, richer with data, and more complex than what a human brain can readily process [18]. So, the abstract nature of visualizations complements the processing power of the mind [37], enabling it to structure, draw connections, and discern patterns. We witnessed in our interviews the benefit of visualizations to data exploration, systematic analysis, and debugging (Section 4.4.3).

4.5.3 *A Visualization as embedded in Context*

With this understanding of context and of the visualization as an abstract view of context, we revisit the model presented in Figure 14. The model shows the visualization to be embedded in

the context that it represents. The visualization is an *abstract view of some context*, and it is *viewed within this context*. This statement can be understood both literally and figuratively: (a) literally, most participants reported using visualizations offered through QDA software, thus viewing the visualization within the context of the QDA environment; and (b) figuratively, the participants engaged with the visualization in a manner that places its connotations within the explanations provided by its context.

When participants approached visualizations, they did not review them in isolation from their context. This understanding of the connectedness of visualizations to context is not different from how other sources of data are treated in a qualitative environment. Interpretation, a goal of qualitative research, is a highly contextual act [94]. Therefore, in the course of interpretation and sense-making, a data point (regardless of its source, a visualization might be one) is embedded and studied within context.

In *the Visual Display of Quantitative Information*, Tufte writes: "data graphics are paragraphs about data and should be treated as such" [12, p. 181]. This view of a visualization as a data source frames it to be semantically equivalent to a collection of words. It is an anecdotal reading for the strong association of visualizations, like words, to their context. As Miles explains: "Words are meaningless unless you look backward or forward to other words" [18, p. 54].

4.5.4 *Negotiating the Role of the Abstract within Context*

In addition to interpreting a visualization *within context*, participants *negotiated the interaction* between the visualization and its context in their research. The visualization and its context both have strengths and weaknesses, affordances and challenges. The visualization, as an abstract representation of context, affords a holistic view of context, enabling the researcher to process

and comprehend a larger expanse of knowledge and making visible patterns that are difficult to detect in nuanced (i.e., complex and detailed) analysis. To be sure, a visualization can be a part of nuanced analysis. But the visualization has a limitation in that it can display only *some* amount of data, therefore *constraining* the ability to do nuanced analysis. The context of analysis, on the other hand, contains a large assembly of data (e.g., primary documents, multimedia files, analytic codes) that makes possible many forms of thorough examination, interpretation and conclusion drawing. But this breadth and richness, at once, escalate the challenge of rigorous processing or pattern detection.

These affordances and challenges drive researchers to use visualizations *in concert* with elements from their larger context, beyond what is explicitly included within the visualization. The visualization complements the interpretation and analysis of context, and context complements the interpretation and analysis of the visualization. As we observed in our interviews, researchers *negotiated the interaction* between the visualization and its context: they displayed familiarity with visual data points as informed by their contextual analysis, they synthesized patterns that are evident in the visualization with understandings from the environment of analysis, and they investigated these patterns by studying the underlying data from which they were induced. Researchers turned from context to visualization when they wanted to explore patterns or questions (e.g., asking how a number of analytic codes correlate, or how many times they are grounded), then the researchers turned from visualization to context to confirm, synthesize or explore further.

The literature also indicates that visualizations and the process of analysis are used in active interaction. Many theoretical frameworks for qualitative analysis in particular feature this bidirectional interaction between analysis and visual displays. Grounded theorists, as an

example, “treat creating visual images of their emerging theories as an intrinsic part of Grounded Theory methods” [70, p. 117]. Grounded theorists frequently use many visual metaphors to describe the process of theory development, referring to activities such as *constructing* theories, *mapping* concepts, and *drawing* connections between ideas or entities [16]. Furthermore, they complement the use of these visual metaphors by advocating the use of visual displays to represent elements of analysis and to reason about relationships between them. Charmaz [70] recommends the use of diagrams to “tease out relationships while constructing analysis.” In her view, diagramming enables the researcher to think about and to see categories and structural properties in their data (such as entities or social groups) and it allows them to reason about and to map their interactions. Charmaz’s diagramming is in active exchange with the context in different ways: (a) diagrams are used at different phases that serve different purposes of analysis; and (b) diagrams are repeatedly revised to reason about and to reflect changes in contextual interpretation. In this way, researchers create diagrams that are abstractions of their understanding of the context and it elements, *while* using this abstraction, as a process and an artifact, to see and interpret “the relative power, scope, and direction of the categories in [their] analysis” [70, p. 118].

Adding to the voice of grounded theorists, Miles and Huberman [18] situate visual displays to be even more integral to analysis. While visual displays are, for grounded theorists, one of many analysis activities, to Miles and Huberman, they are one of three core activities. Miles and Huberman define qualitative analysis to consist of (a) data reduction, which includes the process of selecting data sources, focusing on elements within those sources, coding, and summarizing; (b) data display, which includes organizing information into visible units; and (c) conclusion drawing/verification, which includes construction of meaning from data. Qualitative

analysis, in their view, is a continuous and iterative process among those activities. Within those, they advocate for the use of visual displays in particular (such as matrices, charts, and graphs). Visual displays interact with context because they compact information in a manner that enables the researcher to "see what is happening and either draw justified conclusions or move on to the next-step analysis the display suggests may be useful" [18, p. 22]. To summarize, the visual displays are reduced forms of context, they show a view of context, and they link back to context for verification or further investigation.

4.6 Design Implications

Previously in this chapter, we studied how qualitative researchers approach visualizations, we presented a set of patterns that we found to be characteristic of their approach, and we formulated those patterns into a discussion about the role of visualizations as abstractions within context. Our findings relate qualitative-ness as an *approach* rather than as an *intrinsic feature* of a visualization itself. In this section, we ask: how can we design visualizations in a manner that enables these qualitative approaches? What are the implications of our findings?

4.6.1 *Keep Context Close*

Our participants reported that, as they engaged with visualizations, they interacted with context through mental process (synthesizing visual information with their memories and previous knowledge) and action (reaching out for more data in the analysis environment). We discussed how, through these patterns, the participants' view of visualizations was embedded in the context they study. Therefore, when designing a visualization, we suggest that context be kept close to the visualization, in order to enable this qualitative approach.

This recommendation immediately raises two questions: What context? And how does one measure and reason about “close”?

To illustrate the difficulty of these two questions, we consider the useful design technique of *Focus and Context* [96] from the literature of information visualization. This technique is generally applied to interactive visualizations. It studies how to balance two important views simultaneously in one visualization: the first is a high-level view of a data-set (context) and the second is the detailed information that is focal to the user’s interest at the point of interaction (focus) [96]. Using this technique, the user can zoom into one aspect of a visualization to read its details while maintaining an understanding of its context on the periphery. Figure 15 shows an application: Lamping and Rao [95] designed a hyperbolic browser for viewing hierarchical structures. When the user selects a node, it is brought to the focal point of the visualization (becoming the focus of the visualization) and the remaining nodes rearrange around it in a circular view (maintaining the context of the focal node).

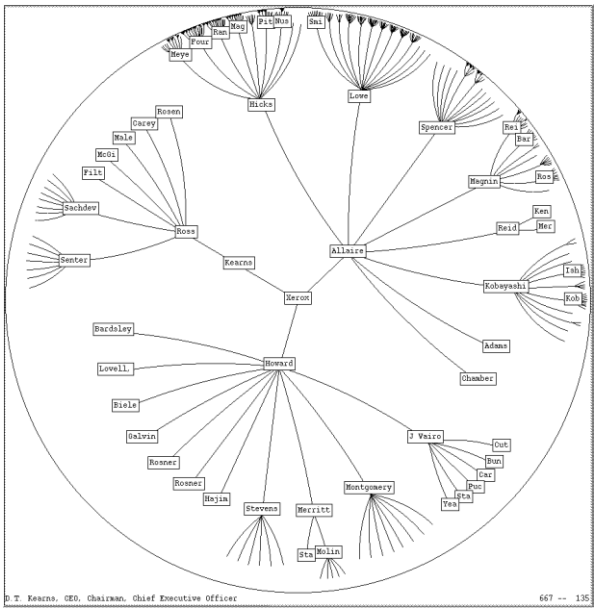

Figure 15 – Hyperbolic Browser (created by Lamping and Rao [95])

While this technique can be helpful in enabling qualitative approaches by making part of the context visible, it presents with two limitations. First, this technique enables navigation among foci and context that belong to the same data-set. In Figure 15, for example, both the focal point and the surrounding context are nodes representing employees and their relationships in an organizational chart. However, as we have seen in our interviews, participants synthesized or shifted from the visualization to context available through different data sources (e.g., P6 synthesized analytic codes with readings from literature). Second, the focus and context technique displays both focus and context *within the visualization*. On the other hand, participants often reached *outside the visualization* to inspect context (e.g., P6 used Code Presence table to extract interview transcripts). Indeed, the variable and contingent nature of context, which changes depending on the mental processing of the beholder, makes it likely for the beholder to reach outside of the visualization into whatever context they choose to investigate.

These limitations reinforce the questions we asked previously in this section: What context should we choose to keep close? How can we predict what context the user will synthesize with or reach for? If we do not display this context within the visualization, then where, and how close, should we keep it?

We believe that the answers to these questions are highly dependent of the design problem and the target users. For example, the visualizations presented in QDA software (e.g., co-occurrence tables) are optimized for the problem of qualitative analysis and for qualitative researchers as users. In terms of what context to keep close, the QDA software contains analysis artifacts (e.g., transcripts or interview recordings). In terms of how close it is kept, the

visualization is offered within the QDA environment, enabling the researchers to switch swiftly from the visualization to the analysis artifacts they deem relevant.

When we suggest *keeping context close* to enable the qualitative approaches we discussed in this chapter, we do not offer specific definitions for *which context* or *how close* it is to be kept. Rather, we believe that these two question should be part of the design research that precedes the construction of a visualization. These two questions and their implications are, in fact, opportunities for future research.

4.6.2 *Provide Access to Underlying data*

Easier than *keeping context* close is to *make accessible the underlying data* that make the visualization. As we expanded at different points in this chapter, the visualization, as an abstract, made visible to the participants certain features about the data. The participants often dived deeper and reached for more data. Therefore, when designing a visualization, we suggest providing as much access to underlying data as is reasonably possible in order to enable this qualitative approach.

The literature of information visualization presents many techniques for exposing more about the data underlying the visualization. These techniques range from providing a reading of a single data point through a tooltip (a small box containing tips about the data point which appears when a user hovers over it) to enabling access to more details about a portion of the visualization. Shneiderman [97] emphasizes the importance of data access as part of his visual information-seeking mantra: “overview first, zoom and filter, then details-on-demand”. For the purpose of our discussion, we regard qualitative approaches as information-seeking. So, we repeat after Shneiderman the call for details-on-demand.

When we ask for access to underlying data, we do not necessarily recommend a database dump. Rather, we want to make use of the power of visualizations to enable a diversity of queries in a simple and interactive manner [98]. For example, when P6 worked with the Code Presence table, he was able to query the underlying data in two ways: (1) vertically, by clicking on a column and extracting quotes from a group of participants; or (2) horizontally, by clicking on a row and extracting the transcript of a single participant.

We conclude:

Through interviews with qualitative researchers, we observed three themes that we consider to be characteristic of qualitative approaches to visualizations: researchers showed knowledge of many data points within the visualization, they synthesized the insights gained from the visualization with elements from its context, and they reached for more data from the context outside the visualization.

These themes elucidate the role of the visualization as an abstract view of context: a visualization is strongly embedded in the context that it studies, it is used in conjunction with its context, and it is interpreted within that context.

Chapter 5. CONCLUSION

5.1 FUTURE WORK

Our exploration stirred many questions that we were not able to explore within the scope of our work. In our pursuit of qualitative-ness in visualizations, we considered some alternative approaches that ultimately were not part of this thesis. For example, at the preliminary stage of our research, we considered whether the "level of context/detail" contained within a visualization is an indication of its "level of qualitative-ness." Henderson and Segal [10] use the concept of "level of display" to reason about the extent to which context is represented in visualizations. They organize "qualitative visualizations" along the spectrum of "level of display"; for example, in their framework, a narrative excerpt displays more context than a word cloud. However, the authors' definition of qualitative data is constrained to textual data alone. This allows them to organize the level of display into three categories in order of increasing representation of context (a word, a sentence, and a paragraph) such that a visualization representing words has less context than a visualization representing sentences (which shows the order and link between words).

Expanding beyond the constraint of Henderson and Segal's work (which is limited to text data alone) illustrates how difficult it is to *measure context*. As an example, consider Tufte's Data-Ink ratio [12], which measures the amount of ink used to represent data in a visualization, as an ordering criterion, such that the larger the ratio, the higher the context represented. But this criterion can be challenged in many ways. For example, consider a scatter plot representing a hundred data points and a concept map with 5 nodes and links. While the scatter plot represents more data, it could be argued that the concept map can holistically represent more about the data with less visual encoding. The question remains: which one has more context? The literature

shows other ways to represent, arguably, more context in a visualization. Some examples from literature overlay pictures or word clouds on geo-visualizations [33]. This way, the visualization is associated with broad representation of the context it represents. But how can the context within pictures or word clouds be measured? Is a word cloud more or less contextual than a picture? What does it mean to measure the amount of context, especially given the multiple co-existing contexts that relate to every visualization?

These questions regarding the level of context within a visualization informed our research as we conducted interviews of qualitative researchers. While we did not ask these questions explicitly, we wondered whether participants would talk about some visualizations representing more context than others. This theme did not emerge during our interviews. However, as described in this thesis (Section 4.5), participants used visualizations in conjunction with the context *outside* the visualizations. In other words, while the visualizations did not evidently include *more context*, but they were used *with context* and *within context*.

As noted in our survey of the literature, and from the responses to our call-for-participation, we observed that qualitative research uses fewer visualizations than what we have seen in a cursory review of the literature of other forms of analysis. Verdinelli and Scagnoli [9] quantify this observation in their systematic survey of three qualitative research journals (see Section 3.2.2), which found that only 27% of the articles they surveyed used visual displays (including tables). We wonder: what are the reasons behind this limited use? Is the qualitative nature of the research an analysis a contributing factor?

Additionally, in our analysis of the dichotomy between qualitative-ness and quantitative-ness, we posit that their coexistence in visualizations provides opportunities to support mixed-method research. Our interviews did not provide sufficient insight to examine this conjecture, as

only two of our participants used mixed-methods (see Table 1). However, P4, who works in a team which uses mixed-methods, stated that she depended on visualizations to understand the quantitative work of her collaborators. We wonder: what is the relationship between and among visualizations, triangulation, and mixed-methods in research?

5.2 LIMITATIONS

The first limitation of our work is that it investigates the qualitative-ness of visualizations within the constraints of qualitative research: (a) we categorize the engagements between visualizations and qualitative forms of data and analysis *within the literature of qualitative research*; and (b) we *interview qualitative researchers* to learn about how they use visualizations. However, the question we pose at the beginning of our exploration (i.e., how can we reason about qualitative-ness in visualization) might have both *answers* and *implications* beyond qualitative research. Qualitative characteristics of visualizations might be explored by asking questions, forming hypotheses and testing them in different fields of design and research outside of qualitative research. (We present an alternative approach in Section 5.1). Further, visualizations with qualitative characteristics, and qualitative approaches to visualization, can be applied in many design and use scenarios that are independent of qualitative research (or independent of research in general). Further research could explore these areas of inquiry.

Another limitation of our research pertains to our third approach in which we conducted interviews to investigate how qualitative researchers use visualizations. We frame our findings from this investigation as characteristic of qualitative approaches to visualizations in analysis. We support the link between our findings and qualitative approaches by grounding these findings in the literature of qualitative research. However, our findings from this investigation could also be relevant to the use of visualization in processes of inquiry and analysis in general. In order to

investigate whether these characteristics are *distinct* and *particular to qualitative research*, further research is needed. This further research could *compare* the use of visualizations across multiple forms of analysis (e.g., quantitative analysis in addition to the qualitative).

5.3 In Conclusion

How can we reason about qualitative-ness in visualization? This question motivated our curiosity and drove our exploration in this thesis, especially noting that the literature of visualization is predominated by quantitative discourse [11] [10]. Our pursuit of the nature of qualitative-ness and its possible characteristics in visualization resulted in three approaches.

Our first approach addressed the tendency to frame what is qualitative as an opposite of what is quantitative. This tendency is exhibited in a dichotomy of discourse on methods and analysis between qualitative and quantitative approaches. This seemed to extend to our initial questioning of qualitative characteristics in the field of visualization specifically. We challenged this dichotomy in our thesis, showing that qualitative and quantitative characteristics coexist in every visualization and that, even though a visualization might have more of either characteristic, an exploration of qualitative-ness that is based on dichotomy is simplistic and might even be counterproductive. As Holliday describes it, defining qualitative approaches as opposite to quantitative approaches is "an unadventurous way to begin" [20, p. 1].

Our second approach, using Cope and Elwood's definition of Qualitative GIS [39], theorizes that the qualitative-ness of visualization is a feature of the engagements between visualization and qualitative forms of data and analysis. We surveyed the literature to learn about the various ways in which qualitative research intersects with visualization and organized those ways into three categories of engagements: (1) using visualizations to generate qualitative data

(i.e., visual methods); (2) using qualitative data to generate visualizations; and (3) using visualizations in the process of qualitative data analysis.

Seeing how the engagements we studied in our second approach were not voluminous (reflecting a vacancy in the literature of visualization and qualitative research), we conducted 7 semi-structured interviews with qualitative researchers to learn how they use visualizations to support their qualitative analysis activities. The findings from our interviews pertained specifically to *how* researchers approached and interacted with visualizations, rather than *what* visualizations they used. We observed that researchers (a) exhibited knowledge of many data points within a visualization with a degree of familiarity that reflects their closeness to the data in the process of collection and analysis (*knowing the data*); (b) synthesized their reading and interpretation of the visualization with elements from the analysis context (*synthesizing and explaining*); and (c) reached for more data within the analysis context to explain and further investigate patterns that become evident through a visualization (*reaching for more*). These observations led us to formulate an understanding of the position of visualizations as embedded in the contexts that they abstract, such that the qualitative researchers viewed them within context and interpreted them in conjunction with other elements within context.

The qualitative-ness of a visualization can be a feature of its engagement with qualitative research: a visualization depicting qualitative data and a visualization that is used in the process of qualitative analysis can be considered qualitative.

Qualitative-ness can, also, be a feature of the beholder's approach to the visualization: as qualitative researchers used visualizations, they showed closeness the data visualized, they synthesized visual encodings with contextual understanding, and they reached for more data within the context of the visualization.

Ultimately, a visualization in qualitative research is read and used within its context and it is interpreted in conjunction with elements from this context.

BIBLIOGRAPHY

[1] D. Socha and K. Sutanto, "The “Pair” as a Problematic Unit of Analysis for Pair Programming," in *Proceedings of the 2015 International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE ’15)*, 2015.

[2] D. Socha, "BeamCoffer Dataset: Professional Software Developers Collaborating in the Wild.," University of Washington Bothell CSS Technical Report, TR-09-15, Bothell, 2015.

[3] C. Geertz, "Thick description: Toward an interpretive theory of culture," *Readings in the philosophy of social science*, 1994.

[4] J. Mason, Qualitative Researching, London: SAGE Publications, 1996.

[5] B. Jordan and A. Henderson, "Interaction Analysis: Foundations and Practice," *The Journal of the Learning Sciences*, pp. 39-103, 1995.

[6] J. v. Wijk, "The Value of Visualization," *Visualization*, 2005.

[7] N. Yau, Visaualize This. The Flowing Guide to Design, Visualization and Statistics, Indianapolis: Wiley Publishing, 2011.

[8] M. Kemp, Visualizations. The Nature Book of Art and Science, Berkeley: University of California Press, 2000.

[9] S. Verdinelli and N. Scagnoli, "Data Display in Qualitative Research," *International Journal of Qualitative Methods*, 2013.

[10] S. Henderson and E. Segal, "Visualizing Qualitative Data in Evaluation Research," *New Directions for Evaluation*, 2013.

[11] A. J. Onwuegbuzie and W. Dickinson, " Mixed Methods Analysis and Information Visualization: Graphical Display for Effective Communication of Research Results," *The Qualitative Report*, 2008.

[12] E. Tufte, The Visual Display of Quantitative Information, Cheshire: Graphics Press, 2001.

[13] J. Tukey, Exploratory Data Analysis, Addison-Wesley Publishing Company, 1977.

[14] S. Few, "Data Visualization for Human Perception," in *The Encyclopedia of Human-Computer Interaction*, Interaction Design Foundation.

[15] T. Munzner, "Process and pitfalls in writing information visualization research papers," *Information visualization*, pp. 134-153, 2008.

[16] J. Corbin and A. Strauss, Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, Thousand Oaks, CA: Sage Publications, 2014.

[17] H. Russell Bernard, A. Wutich and G. Ryan, Analyzing Qualitative Data: Systematic Approaches, London: SAGE Publications, 2010.

[18] M. Miles and A. Huberman, Qualitative Data Analysis, London: SAGE Publications, 1984.

[19] A. J. Onwuegbuzie, W. B. Dickinson, N. L. Leech and A. G. Zoran, "Toward More Rigor in Focus Group Research: A New Framework for Collecting and Analyzing Focus Group Data," *International Journal of Qualitative Methods*, vol. 8, no. 3, 2009.

[20] A. Holliday, Doing and Writing Qualitative Research, London: SAGE Publications, 2007.

[21] Office of Planning and Community Developement - City of Seattle, "Population and Demographics," [Online]. Available:

http://www.seattle.gov/dpd/cityplanning/populationdemographics/. [Accessed 9 January 2017].

[22] S. Few, "Selecting the Right Graph for Your Message," *Intelligent Enterprise,* 2004.

[23] S. Few, "Quantitative vs. Categorical Data: A Difference Worth Knowing," *DM Review,* April 2005.

[24] F. Mosteller and J. Tukey, Data analysis and regression, Boston: Addison-Wesley, 1977.

[25] E. Tufte, Envisioning Information, Cheshire: Graphics Press, 1990.

[26] F. Bentley and J. Murray, "Understanding Video Rewatching Experiences," in *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video,* 2016.

[27] K. Krippendorff, Content Analysis: An Introduction to Its Methodology, London: SAGE, 2012.

[28] D. Holten, "Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data.," *IEEE Transactions on visualization and computer graphics,* pp. 741-748, 2006.

[29] T. Dwyer and Y. Koren, ""Dig-CoLa: directed graph layout through constrained energy minimization"," *Information Visualization,* 2005.

[30] C. McNaught and P. Lam, "Using Wordle as a Supplementary Research Tool," *The Qualitative Report,* 1 May 2010.

[31] F. Heimerl, S. Lohmann, S. Lange and T. Ertl, "Word cloud explorer: Text analytics based on word clouds," in *IEEE Hawaii International Conference on System Sciences*, 2014.

[32] F. Viegas, M. Wattenberg and J. Feinberg, "Participatory visualization with wordle.," *IEEE transactions on visualization and computer graphics,* vol. 15, no. 6, 2009.

[33] J. Jung, "Code clouds: Qualitative geovisualization of geotweets," *The Canadian Geographer,* pp. 52-68, 2015.

[34] Y. Wu, T. Provan, F. Wei, S. Liu and K. Ma, "Semantic-preserving word clouds by seam carving," *Computer Graphics Forum,* vol. 30, no. 3, 2011.

[35] J. Chuang and D. Ramage, "Stanford Dissertation Browser," 2008. [Online]. Available: http://nlp.stanford.edu/projects/dissertations/browser.html. [Accessed 10 January 2017].

[36] T. West, Thinking Like Einstein. Returning to Our Visual Roots with the Emerging Revolution in Computer Information Visualization, Amherst: Prometheus Books, 2004.

[37] T. Munzner, Visualization Analysis and Design, CRC Press, 2014.

[38] S. Stevens, "The psychophysics of sensory function," *American Scientist,* pp. 226-253, 1960.

[39] M. Cope and S. Eldwood, Qualitative GIS, A Mixed Methods Approach, London: SAGE Publications, 2009.

[40] P. Ayoub, S. Wallace and C. Zepeda-Millán, "Triangulation in Social Movement Research," in *Methodological Practices In Social Movement Research*, Oxford, Oxford University Press, 2014, pp. 67-96.

[41] H. Eckstein, "Case Studies and Theory in Political Science," in *Handbook of Political Science,*, London, SAGE Publishing, 1975, pp. 94-137.

[42] J. Gerring, Case Study Research, Principles and Practices, Cambridge: Cambridge University Press, 2007.

[43] D. Sloane, "Visualizing Qualitative Information," *The Qualitative Report,* 2009.

[44] M. Pavlovskaya, "Qualitative GIS," *Wiley-AAG the international encyclopedia of geography: people, the earth, environment, and technology*, 2016.

[45] N. Schuurman, "Trouble in the heartland: GIS and its critics in the 1990s," *Progress in human geography*, pp. 569-590, 2000.

[46] T. Harris and D. Weiner, "Empowerment, marginalization, and" community-integrated" GIS," *Cartography and Geographic Information Systems*, pp. 67-76, 1998.

[47] T. A. Slocum, R. B. McMaster, F. C. Kessler and H. H. Howard, Thematic Cartography and Geovisualization, Pearson Education, Limited, 2013.

[48] J.-K. Jung, "Critical GIS Twenty Years After Friday Harbor Meeting: Critical Intervention to Epistemology, Ontology, Methodology, and the Social Implications of GIS," *Journal of the Association of Korean Geographers*, 2013.

[49] N. Obermeyer, "The evolution of public participation GIS," *Cartography and Geographic Information Systems*, pp. 65-66, 1998.

[50] M.-P. Kwan, "Feminist Visualization: Re-Envisioning GIS as a Method in Feminist Geographic Research," *Annals of the Association of American Geographers*, 2002.

[51] S. T. Rahim, Z. Kougen, T. Saidu and P. Yunhe, "Capabilities of multimedia gis," *Chinese Geographical Science*, p. 159–165, 1999.

[52] M. Shiffer and L. L. Wiggins, "The union of GIS and multimedia," in *Profiting from a Geographic Information System*, Colorado, GIS World Book, 1993, pp. 336-341.

[53] M. Kwan, "Gender differences in space-time constraints," *Area*, vol. 32, no. 2, pp. 145-156, 2000.

[54] L. Knigge and M. Cope, "Grounded Visualization: Integrating the Analysis of Qualitative and Quantitative Data through Grounded Theory and Visualization," *Environment and Planning A*, p. November, 2006.

[55] J.-K. Jung, "Community Through the Eyes of Children: Blending Child-Centered Research and Qualitative Geovisulization," *Children's Geographies,* 2015.

[56] S. Pink, L. Kurti and A. I. Afonso, Working Images. Visual Research and Representation in Ethnography, London: Routledge, 2004.

[57] M. Strong and L. Wilder, Viewpoints. Visual Anthropologists at Work, Austin: University of Texas Press, 2009.

[58] S. Pink, "Visual Methods," in *Qualitative Research Practice*, London, SAGE, 2004.

[59] S. Pink, Advances in Visual Methodology, London: SAGE Publishers, 2012.

[60] M. Banks and D. Zeitlyn, Visual Methods in Social Research, London: SAGE, 2015.

[61] P. Reavey, Visual Methods in Psychology: Using and Interpreting Images in Qualitative Research, Routledge, 2012.

[62] C. Keller, J. Fleury, A. Perez, B. Ainsworth and L. Vaughan, "Using Visual Methods to Uncover Context," *Qualitative Health Research,* pp. 428-436, 2008.

[63] A. Comi, N. Bischof and M. Eppler, "Beyond Projection: Using Collaborative Visualization to Conduct Qualitative Interviews," *Qualitative Research in Organizations and Management: An International Journal*, 2014.

[64] F. Yauner and A. Tennant, "Art/Design Work as a Tool for Primary Data Collection in the Field of Health Research," Newcastle Upon Tyne.

[65] M. Umoquit, P. Tso, T. Varga-Atkins, M. O'Brien and J. Wheeldon, "Diagrammatic Elicitation: Defining the Use of Diagrams in Data Collection," *The Qualitative Report,* 2013.

[66] J. Wheeldon, " Is a Picture Worth a Thousand Words? Using Mind Maps to Facilitate Participant Recall in Qualitative Research," *The Qualitative Report,* 2011.

[67] J. P. Wheeldon and J. Faubert, "Framing Experience: Concept Maps, Mind Maps, and Data Collection in Qualitative Research," *International Journal of Qualitative Methods,* vol. 8, no. 3, 2009.

[68] L. Berends, "Embracing the Visual: Using Timelines with In-depth Interviews on Substance Use and Treatment," *The Qualitative Report,* vol. 16, no. 1, 2011.

[69] M. K. Thygesen, "Utilizing a New Graphical Elicitation Technique to Collect Emotional Narratives Describing Disease Trajectories," *The Qualitative Report,* vol. 16, no. 2, 2011.

[70] K. Charmaz, Constructing Grounded Theory, London: SAGE Publications, 2006.

[71] D. Layder, Doing Excellent Small-Scale Research, London: SAGE Publications, 2013.

[72] P. Hiller, "Visualizing the Intersection of the Personal and the Social Context - The Use of Multi-layered Chronological Charts in Biographical Studies," *The Qualitative Report,* vol. 16, no. 4, 2011.

[73] N. Baugh, A. McNallen and M. Frazelle, "Concept Mapping as a Data Collection and Analysis Tool in Historical Research," *The Qualitative Report,* 2014.

[74] S. Beckhaus, S. L. Brugger and K. Wolter, "Collect and map it all: the artifact map, a tool for complex context analysis," in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, New York, 2012.

[75] J. Cidell, "Content Clouds as Exploratory Qualitative Data Analysis," *AREA, Royal Geographic Society,* 2010.

[76] M. Horney, "Interactive Data Visualization in Qualitative Research," in *SIGGRAPH Computer Graphics*, 1994.

[77] S. Hai-Jew, Enhancing Qualitative and Mixed Methods Research with Technology, Hershey: Information Science Reference, 2015.

[78] A. Fouse and J. Hollan, "DataPrism: a Tool for Visualizing Multimodal Data," in *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research (MB '10)*, New York, 2010.

[79] A. Fouse, N. Weibel, E. Hutchins and J. Hollan, "ChronoViz A System for Supporting Navigation of Time Coded Data," in *Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*, New York, 2011.

[80] N. Weibel, A. Fouse, E. Hutchins and J. Hollan, "Supporting An Integrated Paper-Digital Workflow for Observational Research," in *Proceedings of the 16th International Conference on Intelligent User Interfaces (IUI '11)*, New York, 2011.

[81] P. Bazeley and K. Jackson, Qualitative Data Analysis with NVivo, London: SAGE Publications, 2013.

[82] J. Gamez, "What qualitative charts are there?," Zendesk, 17 12 2015. [Online]. Available: https://dedoose.zendesk.com. [Accessed 2 3 2017].

[83] "Move forward with queries and visualizations," QSRInternational, [Online]. Available: http://help-nv11.qsrinternational.com. [Accessed 2 3 2017].

[84] J. Hermanowicz, "The Great Interview," *Qualitative Sociology*, pp. 479-499, 2002.

[85] H. Rubin and I. Rubin, Qualitative Interviewing: The Art of Hearing, Thousand Oaks: SAGE , 2005.

[86] L. Richards, Handling Qualitative Data: A Practical Guide, London: SAGE Publications, 2014.

[87] M. Sandelowski, "Sample Size in Qualitative Research," *Research in Nursing and Health*, 1995.

[88] N. G. G. Fielding and R. M. Lee, Computer Analysis and Qualitative Research, London: SAGE Publishers, 1998.

[89] L. Richards, "Closeness to Data: The Changing Goals of Qualitative Data Handling," *Qualitative Health Research ,* vol. 8, no. 3, 2016.

[90] A. Bryman, M. Stephens and C. Campo, "The importance of context: Qualitative research and the study of leadership," *The Leadership Quarterly*, vol. 7, no. 3, 1996.

[91] G. Miller and R. Dingwall, Context and Method in Qualitative Research, London: SAGE Publishing, 1997.

[92] P. Dourish, "What we talk about when we talk about context," *Personal and ubiquitous computing*, 2004.

[93] A. Hinton, Understanding Context. Environment, Language, and Information Architecture, Sebastopol: O'Reilly, 2015.

[94] P. Dourish, "Seeking a foundation for context-aware computing," *Human0Computer Interaction*, pp. 229-241, 2001.

[95] Lamping and R. Rao, "The hyperbolic browser: A focus+ context technique for visualizing large hierarchies," *Journal of Visual Languages & Computing*, 1996.

[96] S. K. Card, J. D. Mackinlay and B. Shneiderman, Readings in Information Visualization: Using Vision to Think, San Francisco: Morgan Kaufmann Publishers, 1999.

[97] B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," in *Proceedings of the 1996 IEEE Symposium on Visual Languages*, Washington, DC, 1996.

[98] C. Stolte, D. Tang and P. Hanrahan, "Polaris: A system for query, analysis, and visualization of multidimensional relational databases," *IEEE Transactions on Visualization and Computer Graphics*, 2002.

APPENDIX A

The following table expands Table 2 in Section 4.4. It provides excerpts that illustrate and support the major themes within the transcript of our interviews. The (●) symbol indicates the presence of a theme in relation to the participant's approach to the corresponding visualization in the table (empty cells indicate that the theme is not present.).

P	Visualization	Knowing the Data	Synthesizing and Explaining	Reaching for More
1	Network view		●	
		P1 explains how network view enables the visualization of connectedness between elements from the analysis context: *"Visualization is an important part of it. How you really see not only just raw data, but in a way that doing that through analysis, right? Making the relationship, like, this is related to this. Oh this is related, this is connected. This is the cause of this."*		
	Geo-visualization of census Data			
	Multimedia geo-visualization	●	●	●
		P1 links photographs taken by his participants to geo-visualizations. When looking at the geo-visualization, P1 remembers which of his participants provided the photograph. As P1 analyzes this visualization, he synthesizes the location of the photograph taken by participants with other types of data about that location: *"Sometimes we really want to [know], where is it? Because when we know the location, we can kinda compare it to what's going on there. Is it really a safe place by other crime analysis?"* P1 *reaches for more data* outside the visualization to learn more about the photographs and the participants: *"When you click it, its image can be linked to the textual data: what they talked about, the particular area, and audio and video. So again, it's not just for the mapping or visualization, but with the visualization it also allows us to access the database."*		
	Geo-visualization of racial distribution			
	3D heat map	●	●	
		The heat map showed the neighborhoods of Buffalo, NY, and the distribution of grocery stores within those neighborhoods. As P1 showed		

		the visualization, he <u>remembered</u> certain locations, bus stops, and grocery stores: *"When I lived here, I know that, I actually stayed in Buffalo East side. And then, not many people have cars, and very much rely on public transportation: train, bus. And I was actually one of them. They're carrying a lot of plastic bags, cold winter time."* P1 <u>synthesized</u> locations with racial distribution: *"Buffalo is very much a segregated city. The most segregated city in the US. Buffalo East side is completely black community. And the West side it's more diverse. But also, we start to see the lower West side, it becomes more Hispanic community."*		
	2D heat map		●	
		The heat map showed the neighborhoods of Buffalo, NY, and the distribution of grocery stores within those neighborhoods. P1 <u>synthesized</u> the areas with no grocery stores with racial distribution: *"I'm thinking this map…. But, now, when I look at that black population, it's so obvious to me that areas that are not part of that big grocery store service area is, it looks like Buffalo Eastside."*		
	Geo-visualization of transit connections	●	●	●
		The visualization showed the travel time to grocery stores using public transit. P1 <u>remembered</u> travel times for some of the locations in the visualization. He <u>synthesized</u> those times from memory with those indicated in the visualization and disagreed with the visualization: *"Looks like it actually covers most of the areas in Buffalo. Which was actually big surprise for me. Because when I was living here I had to wait for the bus for a very long time. But, it looks like it covers all area."* To inspect the accuracy of this visualization. P1 <u>reached for more data</u> by visiting some of those locations: *"I visited these two grocery stores. I'm adding qualitative research because I'm visiting, I'm going to the field."*		
	Cartogram			
	Collage of map, visual, and visualization			
	Geo-visualization of crime rates	●	●	
		The visualization showed the crime rates within neighborhoods. P1 <u>remembered</u> the crime rate and the perception about some of the locations in the visualization. He <u>synthesized</u> his memories with those crime rates indicated in the visualization and disagreed with image created by the visualization: *"When I was already came here as internship student, orientation [they showed a similar visualization and] they're saying that: 'don't even think to go to Buffalo Eastside.' It was part of those orientations; it's so dangerous area. But I spent four, five years there and I'm here."*		

		"That's why I was shocked when I first joined this district, oh my god, I never expected people [living there] would meet monthly and discuss how we improve our neighborhood."		
	Code cloud			
	Word cloud of geo-tweets			●
		As P1 inspected a word cloud of geo-tagged tweets, he <u>reached for more data</u> to inspect the contexts within which each word is used: *"But it was actually interesting that even this one more thing. So some student from the UW library said 'I need to call in sick today so I can watch all my shows', 'I'm sure this concert was sick as hell', 'sick socks bro'. Sick, different meanings. It needs to be treated differently then."*		
2	Timeline sketches	●	●	●
		P2 <u>knows the data</u>: *"It is funny, the other thing about these visual artifacts. It gives me very vivid memories of the interview. So this one was with an artist and she was very bohemian…"* P2 <u>synthesizes</u> the visual encoding within the timeline with memories about the interviewee who drew the timeline: *"This also, she had black, which is all negative emotions. It is really funny because the circles are people in her life and how she felt about them at different points and you will notice they are very close here and very close here, but when she was going through her darkest periods she felt very far away from them.* P2 <u>reaches for more data</u> outside the timeline sketches in the audio or transcript of the interview where the sketch was drawn: *"So with data triangulation what I do is I never look at their sketch without either listening to the transcript or reading the transcript."*		
	Affinity diagram		●	
		P2 and her collaborators read the content of the diagram differently because they <u>synthesized</u> them with their own experiences: *"And it turned out that the cohorts had two very different takes on one thing [in the diagram] and it was because they were in different times in life and had different priorities. But we were able to say in the paper these people talk about the same thing but they talk about it in very different ways."*		
	Code co-occurrence		●	●
		P2 and her collaborators discuss and interpret code co-occurrence <u>in synthesis</u> with what codes mean and why they created them: *"So [my collaborators] will say 'oh man code A and code B really overlap and I wasn't thinking about that when I was coding but I think that might be a thing.' Then, we talk about why that might be a thing or are these codes the same thing maybe."* P2 and her collaborators <u>reach for more data</u> to investigate the patterns they see in the code co-occurrence table: *"We'll say 'Well, are any of these patterns worth tracking down?' If there are no patterns, 'why is that?' So I'll often start with the co-*		

		occurrence table when I meet with the coding team and then you can double click where you see no co-occurrence or you see lots of co-occurrence."		
	Word cloud			
4	Word cloud			
	Code co-occurrence	●	●	●
		P4 and her collaborators use code co-occurrence table to debug and refine their coding scheme. Knowing the data, as they look at some of the problematic codes in the table, they remember aspects of their fieldwork that resulted in the patterns they see in the table: *"Because we were both deeply in the field, or as close to the data as you can be when you have this much data to keep in your head."* *"Some of these schools were very chaotic, like in the times they were making these transitions. These were low performing schools trying to get better, and so especially in the early years there's just a lot of noise in the data."* P4 and her collaborators discuss and interpret code co-occurrence in synthesis with what codes mean and why they created them: *"We're learning some of our codes don't have a ton of analytic value if there's 80,000 things in them, so [the code co-occurrence table] has been helping us do some of that thinking around refining the coding scheme."* *"Then I'd run co-occurrence tables to say, "Okay, well this code has 1500 things, but it's co-occurring with X, Y, and Z in the following ways. So let's think about what that code is truly telling us. Is it completely redundant with these three things? Of the places where it's not co-occurring, is that actually telling us something unique?"* To investigate problematic codes in the table, P4 and her collaborators reach for more data: *"As it turns out we just have 1500 elements that don't go with anything else. Make note of those places that I think require deeper dive. "* *"Generally, I will talk with [my collaborator] first and we will come up with what the data mining strategy is, and then yeah, it often involves either going back to the primary documents to figure out..."*		
	Network view			
	Tableau (various visualizations)			
	Dynamic map of research results			
	Flow chart			
5	Network view			●
		P5 explains the function of the network view for navigation because it enables reaching for more data:		

		"Even just take it as an opportunity to make a visualization that you could then go from your visualization back to your data."		
	Code distribution		●	
		P5 synthesizes the distribution of her analytic codes with what they mean in her analysis context: *"I've got the super codes obviously the most."* *"So I was sort of [using the code distribution] throughout and then I did as a sort of check on. It's a credibility check where you go back through your code or you go back through whatever and be like: 'Did I do anything really stupid?'"*		
	Code co-occurrence		●	
		P5 uses code co-occurrence for theory-building. She synthesizes the co-occurrence of codes with her deductions: *"So this code is basically capturing every time that there is interaction between public and government. And I know from the hypotheses that I'm working with now that I'm interested in knowing how much of the time does that interaction accompany learning.* *So I could use the co-occurrence table and look at it and be like, okay, public-government interaction and public-government learning actually has a reasonably high co-occurrence."*		
	Code primary document table	●		
		P5 knows the data within the code primary document table. She remembers some of the interviews and surveys corresponding to the data points within the table: *"There's just a point where you've spent a long time with it and.. I even just remember what the interviews were because I've spent so long with them that I can look at that code primary document table and I know what every one of those columns is, I can remember who that column is, and so to some extent when I see that they flagged a public-government learning I can often remember what the thing they were describing was, because that's one of the things that comes from translating your own interviews. You just get to know them really well."*		
	Concept diagram			●
		P5 draws the concept diagram on paper. She, then, reaches for more data in her QDA (Atlas.ti) to learn about the concepts she is drawing: *"That's where I might interact with Atlas, where I'm like, 'OK, who are all the actors who I've been describing here?' I would code people with what type of person they were. Like, what kinds of roles are being listed here, where do they fit into these maps?"*		
6	Code co-occurrence		●	●
		P6 and his collaborators synthesize the patterns in the code co-occurrence visualization with the literature: *"So like this one as you would expect, they talked a lot about the learning context but also in relation to the school environment. And so that's an example where it's not enough, that it doesn't appear to be enough that a teacher knows his or her subject and can teach it well. If*		

		the learning environment of the school as a whole is hostile or indifferent, middle school students won't learn." As P6 uses code co-occurrence, he reaches for more data to investigate the patterns he sees: *"Then you have this list that it pulls instantly from all 55 [interview transcripts]."* *"So you can come to an assertion like that based on kind of all of the excerpts that are represented here."*		
	Code presence table	●		●
		P6 knows the data represented in the table. As P6 looks at the table, he remembers why one of the transcripts was split into two rows: *"I remember this one. This is the only one where there was part one, part two. There was a fire alarm. So we had to stop."* P6 synthesizes P6 reaches for more data about a an analytic code across all transcripts or within one transcript, in order to create a narrative for one or multiple themes: *"So then we look at those excerpts. Then it's either, so you have two ways: One is to pull all of those excerpts with that code. To see across the group, what was the impact that those students sensed that it had on their pathway to dropping out?* *The other is to look at one of those transcripts and to read through it to see where the retention occurred and what it was like for that individual."*		
	Code cloud		●	●
		P6 synthesizes the visual encoding of codes in the cloud with the literature and with understandings from analysis: *"But what I like is that the things that tend to get emphasized by popular media anyway are as small as they should be. For example, pregnancy. I mean obviously that'd have to be small, because at least half of our participants couldn't have gotten pregnant. Just there's that."* P6 reaches for more data to study the context in which a code was applied: *"... That in turn reflects what was said. Because you know they talked about school. And I can see that we can. And if we click on it, I think that will go to all the excerpts."*		
	Concept diagram	●		
		P6 created a diagram of actors and relationships that were studied, known, and understood by him: *"So the first part is just about building relationships and so I found that literacy coaches were part mentor and part advocate. And so those two roles were balanced across these stages."*		
	Line graph		●	
		P6 used a line graph based on quantitative survey data and synthesized it with his interview data:		

		"From [the line graph] we could show that the low point was here, right before dropping out, and that it ticked back up a little bit because the students who reported more favorable experiences later in high school were specifically students who had dropped out and then gone into a re-entry program, or alternative school that they liked much better. We explained that, but it really shows that it's liking elementary school, gradual slide, and then dropping out. But then those students who came back liked the programs they came back into."
	Bar chart	<table><tr><td></td><td>●</td><td></td></tr></table>
		P6 was surprised by the disparity that is evident in the bar chart between the attitudes of males and females towards math. He explained and reasoned about this disparity by <u>synthesizing</u> with the literature and with his interviews: *"The math one was more of a surprise. We expected students to be discouraged by mathematics, and research suggests that girls tend to be more discouraged, if that's the right way to say it, by mathematics in high school. But we were surprised when we put in the data that the discrepancy was so large. In that case, the visualization shows the dramatic gap. They should be even, because all of those kids dropped out. They should be pretty close as to how they saw themselves entering math, but it appears from what they said that the girls were much more discouraged by math, or they were less afraid to say it."*
7	Interaction diagram	<table><tr><td>●</td><td>●</td><td>●</td></tr></table>
		P7 <u>knows some of the data</u> within the diagram. As P7 explained the diagram, he remembered the names of several people represented in the diagram even though their names were not shown in the diagram: *"This is two people at one pairing station: David and Henry. They're working together. They're pairing for that period of time. And then, Henry stays and David R. comes and they pair for a while."* P7 <u>synthesized</u> the visualization with data he collected outside the visualization and he reasoned about why the visualization does not correspond to the data: *"But it still doesn't make sense when I look at it. So maybe something's wrong with my statistics. But that was the data when I did it and I checked it multiple times. And It looks like 20 percent of the time they're not actually working by themselves."* P7 notes that the static nature of the visualization did not allows him to quickly reach for more data. He explains how this motivated him and his collaborators to create an interactive visualization that uncovers more about the underlying data: *"That was the next question: can we make this interactive? Can we start using this as a visual analytics tool?"*