Empirical Data in Disaster Recovery: A Data Pipeline to Investigate a Pandemic's Impact on Community Mobility

Matthew Martell

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

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Youngjun Choe, Chair

Ji-Eun Kim

Prashanth Rajivan

Program Authorized to Offer Degree:
Industrial and Systems Engineering

University of Washington

Abstract

Empirical Data in Disaster Recovery:
A Data Pipeline to Investigate a Pandemic's Impact on Community Mobility

Matthew Martell

Chair of the Supervisory Committee:
Youngjun Choe

Disaster events are becoming an increasingly common part of life for people across the world. Climate and weather-based disasters are on the rise, in addition to higher probabilities of future disease emergence. One of the main ways we understand past and current events and work to mitigate the effects of future events is through the use of data. Data, when used appropriately, can help us understand the possible impact of disaster events and our mitigation efforts. Data's impact is maximized when data sets and methods are shared. Unfortunately, there is a dearth of understanding of what data are available in the disaster research field, leading to slow progress in community resilience efforts.

This dissertation provides a greater understanding of the data being used in the disaster recovery space, and generates a new data set on community recovery from COVID-19 using a novel method. First, we conduct a review of data use in lifeline infrastructure restoration modeling. The review covers publications from the 1980s to 2019. It describes both the types of data used and their features, in addition to breakdowns of how different modeling approaches obtained and utilized data differently. The review concludes with some discussion of future research directions for the field and data management best practices, most notably data publication and documentation for fully reproducible methods.

Next we create methods for utilizing street-view images to study community mobility during disaster events. The framework is an open-source data pipeline for street-view images.

It demonstrates the data management best practices from the review while showing it is possible to gather insights into the effects of community events and public policy on foot traffic using the street-view images. The framework is usable to detect any type of object where an appropriate computer vision algorithm is readily available, and has applications in disaster recovery, urban planning, and even demographics estimation. Our implementation uses the images to count pedestrians and uses the count as a metric for community mobility. Additionally, the entire data set and a sample code base are available for reproducibility, or use in other research efforts.

The final chapter of this dissertation implements a distance estimation algorithm on the street-view image data set to understand how outdoor social distancing practices changed over the course of the pandemic. The analysis includes vaccine availability, weather, and time of the week as predictors in addition to socioeconomic factors at the census tract level. In addition to looking at the city of Seattle as a whole, this analysis also subsets the data into various places of interest such as schools or hospitals to understand the distancing trends at these locations.

In Memoriam

Carl Armbrust, my grandfather, who taught me the value of an education.

Acknowledgements

Thank you so much to my dissertation committee members Dr. Prashanth Rajivan and Dr. Ji-Eun Kim, who volunteered their time to help me through this process and provided valuable feedback at every point. Thank you to my family, especially my parents and siblings, Alan, Susan, Jennie, Jeffrey, and Johnny, who supported my many different career tracks over the years, and always encouraged me to put my own happiness first. Thank you to all the UW ISE graduate students, who made B14 a fun place to be in the early months, helped keep me sane during the height of the pandemic, and welcomed my back after my leave of absence. Thank you to my many friends and mentors in Illinois and Minnesota, who have maintained those relationships no matter the distance. Lastly, special thanks to Dr. Youngjun Choe, who guided me every step of the way, chaired my dissertation committee, and most importantly made sure I did not give up on this dream.

TABLE OF CONTENTS

Page

List of Figures . . . iii

List of Tables . . . v

Chapter 1: Introduction . . . 1
- 1.1 Motivation . . . 1
- 1.2 Research Objectives . . . 3
- 1.3 Organization . . . 7

Chapter 2: Review of Empirical Quantitative Data Use in Lifeline Infrastructure Restoration Modeling . . . 8
- 2.1 Introduction . . . 8
- 2.2 Methods . . . 9
- 2.3 Findings . . . 11
- 2.4 Future Directions for Research . . . 23
- 2.5 Summary . . . 35

Chapter 3: Open-Source Data Pipeline for Street-View Images: A Case Study on Community Mobility During COVID-19 Pandemic . . . 40
- 3.1 Introduction . . . 40
- 3.2 Methods . . . 42
- 3.3 Results . . . 48
- 3.4 Discussion . . . 50
- 3.5 Conclusion . . . 58

Chapter 4: Outdoor Social Distancing Behaviors Changed During a Pandemic . . . 59
- 4.1 Introduction . . . 59
- 4.2 Methods . . . 62

4.3 Results 67
4.4 Discussion 71

Chapter 5: Conclusion 80

Appendix A: Appendix 83
A.1 Additional Regression Results for Data Pipeline 83
A.2 Additional Sample Image Outputs for Social Distancing Algorithm 87
A.3 Capitals Regression Results - Number of Distances Under 9 ft per Image . . 90
A.4 Capitals Regression Results - Proportion of Distances Under 9 ft 97
A.5 Regression Results - All Distances Per Image 104

Bibliography 105

LIST OF FIGURES

Figure Number | Page

1.1 Billion Dollar Disaster Events Over Time in the U.S. Data source: [197]. . . . 2

2.1 Publication trends over time. 12

2.2 Breakdown of the reviewed literature. 15

2.3 Relating modeling approach to lifeline system and hazard type. 16

2.4 Interdependency in the reviewed literature. 25

2.5 Graphical representation of the proposed data management methodology. . . 34

3.1 Sample images from the pedestrian detection data pipeline. The left image is an original 360° image from a data collection run. The image on the right is the right-hand side of the original image after orthorectification and pedestrian detection (both sides of the image are processed separately). There are two pedestrians that were detected by the algorithm (in red bounding boxes). . . 44

3.2 Flowchart of the data processing pipeline. The parts of the flowchart in gray occur on NHERI DesignSafe-CI, while the right-hand part in blue is done on the Frontera cluster. 46

3.3 Time series data of the total detections per image (solid blue line, left axis), and detections per image for the subset of detections sharing an image with at least 4 others (orange dashed line, right axis). As the survey dates are irregular, all dates are included in the figure. Please note that the axis for total detections per image does not start at 0. This was done purposefully to facilitate comparison between the trends of the two graphs. 49

3.4 A depiction of our own detections per image data (blue, dashed; right axis) against Google Community Mobility data (orange, solid; left axis). The Pearson correlation between the two data sets is 0.387. The Google Community mobility data is aggregated at King County, WA, while our data covers a survey route within Seattle, which belongs to King County. As the dates of surveys were irregular (e.g., due to weather conditions), all dates are included in the figure. 54

4.1 Sample output of the pedestrian detection and social distance estimation algorithm. There are 2 pedestrians, with an estimated 14.38 ft between them. Additional sample images are available in the supplementary material. 61
4.2 Summary data from the outdoor social distancing analysis. The blue line represents the number of distances under 9 ft per image. For example, 1.0 means that there is one pair of pedestrians on average per image, whose distance is under 9 ft. 63
4.3 Cumulative vaccination totals over time in King County, Washington, where Seattle is the largest city, show diminishing willingness to get vaccines. Data collected from the King County Department of Public Health [104] 72
4.4 The left side graph shows the death rates per 100,000 citizens for different racial groups in King County, WA [103]. (NHPI, Native Hawaiian or Pacific Islander; AIAN, American Indian or Alaska Native). The right side graph shows the expected proportion of distances under 9 ft per image across our entire data set during the summer, in a lower-income census tract, on a weekday, before the vaccine became publicly available, based on our regression results. It is shown here that the more white areas did not have a statistically significant difference from the less white areas. 74

A.1 Sample output of the pedestrian detection and physical distancing algorithm. This image contains two pedestrians near each other on one side of the street, about 30 feet away from a pedestrian across the street. 87
A.2 Sample output of the pedestrian detection and physical distancing algorithm. This image contains two pedestrians walking nearby each other with their backs to the camera. 88
A.3 Sample output of the pedestrian detection and physical distancing algorithm. This image contains two pedestrian walking nearby each other walking parallel with the vehicle. 89

LIST OF TABLES

Table Number Page

2.1 List of every publication included in review, in order of publication year. E - electricity, W - Water, G - Gas, Tel - Telecommunications, Tr - Transportation, WW - Waste Water, ESC - Emergency Supply Chain, F - Fuel, M/U - Multiple/Unspecified, OW - Other Wind. 37

3.1 OLS Regression Results for Detections per Image. The first three non-intercept terms represent indicator variables for the different seasons, with fall being the baseline. The Vaccine Available term represents a binary variable for whether the COVID-19 initial vaccination series was publicly available or not. Weekend is a binary variable for whether the data was collected on Saturday or Sunday. The four Income Bracket terms are indicator variables for the median income level of the census tract where the data was collected. The income brackets are defined in our methods. Lastly, the More than 55.5% White term is an indicator variable for if the census tract in question had a populace that is more than 55.5% White. Full documentation for the Python package used to make this output is available from the developers [188]. 51

3.2 OLS Regression Results for Detections per Image for the detections subset sharing an image with at least 4 others. Coefficients are defined the same as in Table 3.1. 52

4.1 OLS Regression Results - Distances under 9 ft per Image. Please note that as there is some overlap that occurs between images, the coefficients here can only be interpreted relative to each other. Stating that the vaccine effect is 3 times as great as the summer effect in the same direction is useful, but the absolute interpretation of these coefficients' sizes is not. Full documentation for the Python package used to make this tabular output is available from the developers [188]. 68

4.2 OLS Regression Results - Distances under 9 ft per Image - Summary of Capitals Analysis. The numbers reported are the regression coefficients. Significance level: 0.05*, 0.01**, or <0.001***. 69

4.3 Difference-in-proportions test results for the proportion of distances under 9 ft *at each capital*, when compared to 0.264, the proportion of distances under 9 ft across the *entire* data set. 69
4.4 OLS Regression Results - Proportion of Distances Under 9 ft. The coefficients here can be directly interpreted as the relative change in the proportion of distances under 9 ft when a given predictor changes from 0 to 1. 70
4.5 OLS Regression Results - Proportion of Distances under 9 ft - Summary of Capitals Analysis. The numbers reported are the regression coefficients. Significance level: 0.05*, 0.01**, or <0.001***. 71

A.1 OLS Regression Results for Detections per Image for the detections subset sharing an image with at least 1 other. 84
A.2 OLS Regression Results for Detections per Image for the detections subset sharing an image with at least 2 others. 85
A.3 OLS Regression Results for Detections per Image for the detections subset sharing an image with at least 3 others. 86
A.4 OLS Regression Results - Transit Stops - Number of Distances Under 9 ft . . 90
A.5 OLS Regression Results - Faith-Based Organizations - Number of Distances Under 9 ft 91
A.6 OLS Regression Results - Hospitals - Number of Distances Under 9 ft 92
A.7 OLS Regression Results - Medical Clinics - Number of Distances Under 9 ft . 93
A.8 OLS Regression Results - Museums - Number of Distances Under 9 ft 94
A.9 OLS Regression Results - Parks - Number of Distances Under 9 ft 95
A.10 OLS Regression Results - Schools - Number of Distances Under 9 ft 96
A.11 OLS Regression Results - Transit Stops - Proportion of Distances Under 9 ft 97
A.12 OLS Regression Results - Faith-Based Organizations - Proportion of Distances Under 9 ft 98
A.13 OLS Regression Results - Hospitals - Proportion of Distances Under 9 ft . . 99
A.14 OLS Regression Results - Medical Clinics - Proportion of Distances Under 9 ft 100
A.15 OLS Regression Results - Museums - Proportion of Distances Under 9 ft . . 101
A.16 OLS Regression Results - Parks - Proportion of Distances Under 9 ft 102
A.17 OLS Regression Results - Schools - Proportion of Distances Under 9 ft . . . 103
A.18 OLS Regression Results - Distances Per Image. This table contains regression results for a regression containing ALL distances, not just those under 9 ft. 104

Chapter 1

INTRODUCTION

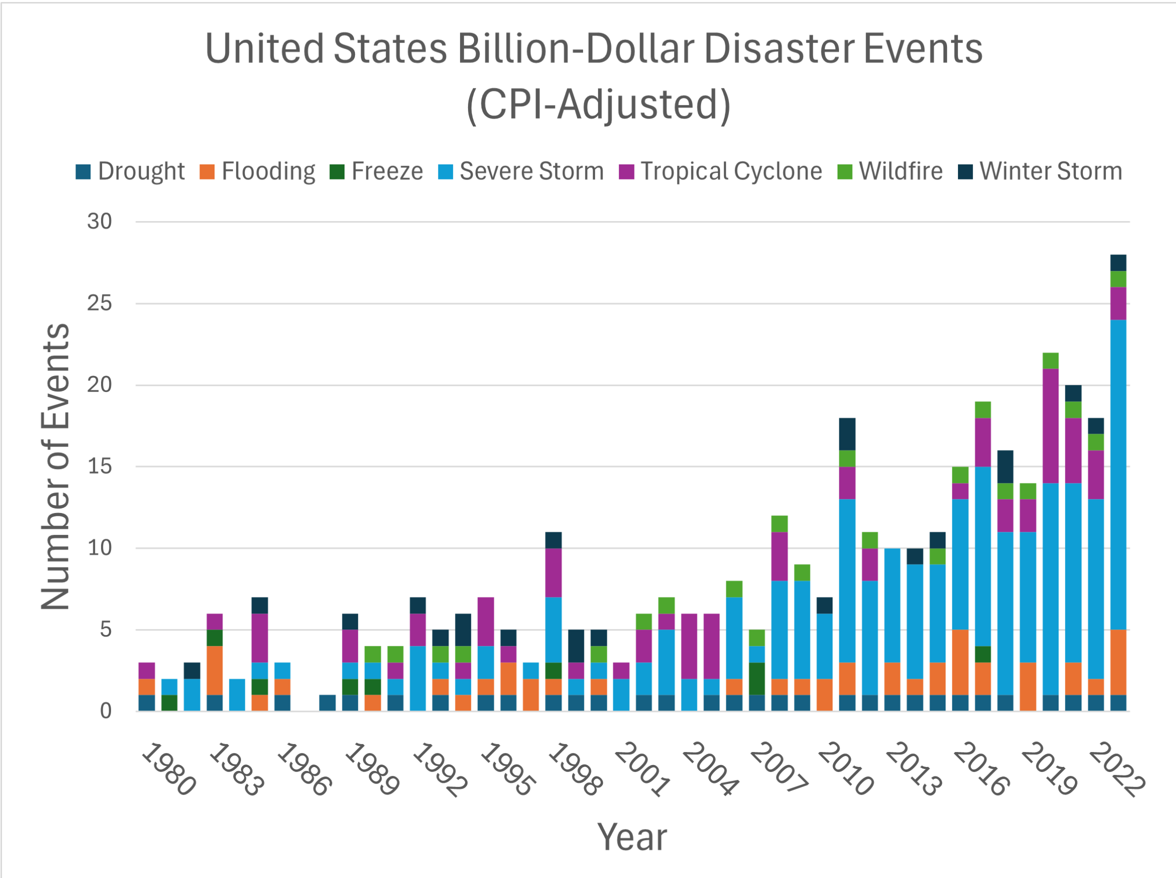
1.1 Motivation

Cities face significant risks from disaster events, including extreme weather events, heat waves, and transmissible disease. Examples from recent years in the U.S. include natural disaster such as the 2017 Hurricane Harvey, as well as the COVID-19 public health disaster. The risk of disaster events occurring is only increasing, with climate change being a key driver of both weather-based disasters [102] and vector-borne disease [121]. As can be seen in Figure 1.1, the number of CPI-adjusted billion dollar disaster events in the US has increased drastically over time. In order to better prepare for disaster events, it is necessary to understand what factors make communities more successful in disaster mitigation at all stages, including preparedness, response, and recovery.

In particular, recovery from disasters is considered one of the less-understood phases of the disaster cycle [132, 200]. Disaster recovery is a broad term that encompasses social, economic, built and natural environments, and it is largely accepted that recovery implies bringing each of these facets to predisaster levels or better [101, 113]. In addition to national efforts, in recent years, several state governments in the US created initiatives to better plan for disaster events [160, 221]. These initiatives in turn inspired the creation of the National Institute of Standards and Technology (NIST) Community Resilience Planning Guide [150] as a model for other areas of the US.

While there were many strides made in the last decade for disaster recovery efforts in the U.S., one challenge that has continually impeded progress is the lack of, or perceived lack of, data related to disaster events and recovery from those events [33, 133]. Particularly at higher level of spatial or temporal resolution, emergency management organizations are lacking in

Figure 1.1: Billion Dollar Disaster Events Over Time in the U.S. Data source: [197].



data and tools to understand recovery patterns [91, 108, 142]. Data can be difficult to access due to reasons such as security, liability, competition laws, confidentiality and privacy [163]. In addition, the volume of data required to model some aspects of disaster recovery is a major barrier to research [178].

In spite of these challenges, disaster recovery research continues to move forward, and there are more and more potential data sources available every year as the amount of information on the web continues to increase [43]. The challenge remains, however, to identify the best possible data sources, and utilize them for disaster recovery research. This dissertation aims to help answer these questions by understanding what data sources have been used historically through a literature review of data use in lifeline infrastructure restoration modeling. Additionally, we identify a promising data source with untapped potential in street view imagery, and develop a data pipeline for collecting and processing this data. Lastly, we use this data pipeline to help understand one facet of recovery from the COVID-19 pandemic, social distancing behaviors.

1.2 Research Objectives

The main research objective of this dissertation is to advance a holistic understanding of data usage in disaster recovery research. Specifically, this is accomplished by the following:

1. Review empirical quantitative data usage in lifeline restoration modeling to understand the history of data usage in this sub-field of disaster recovery research, and provide future thematic and methodological directions for research.

2. Develop and implement methods for collecting longitudinal street view image data for disaster recovery research, including a data pipeline for extracting pedestrian count data from the images.

3. Implement a distance estimation algorithm on a large scale street view image data set to conduct a case study on longitudinal social distancing behaviors during the COVID-19 pandemic.

1.2.1 Review of empirical quantitative data use in lifeline restoration modeling

Lifeline infrastructure restoration modeling is a data-intensive process, frequently requiring system topologies, component geographical locations, and emergency procedural knowledge for the lifeline in question. Here, lifelines are defined as a subset of critical infrastructures vital for communities to function, including systems such as electricity and natural gas [210]. Given the data-intensive nature of the field, experts have called for uniform data collection methods [163], and a community of practice around the broader field of disaster recovery, including shared data sets [132]. Knowledge about available data, where it was collected from, and how it was used to model lifeline restoration will allow the field to advance forward in its data practices, creating higher quality models in a shorter period of time. This review summarizes the history of empirical data use in this field, and highlights common relationships between modeling approaches, hazard type, lifeline system, and data set features.

In order to limit the scope of the review, only publications that contain a quantitative restoration model of a lifeline system, based on empirical data, are included. With these restrictions, we identified 54 publications that utilized a broad variety of methods and data sources. This review demonstrates promising future thematic directions include modeling interdependent systems, something currently done by only 13 of the reviewed studies, and engagement through benchmarking testbeds, virtual communities that support emergency management research without compromising potentially sensitive information. Future methodological directions included making improvements in model validation and data management practices, as well as utilizing untapped data sources in the lifeline infrastructure restoration modeling space.

1.2.2 Open-source data pipeline for street-view images: a case study on community mobility during COVID-19 pandemic

One data source with some history in disaster recovery modeling is repeat photography [27]. Image data is considered valuable in a variety of research domains because it allows for a comprehensive audit of a scene [111]. In particular, Street View Images (SVI) are increasingly being used for research purposes, as they are frequently free to access and can be used to answer questions in a host of research domains [41]. Challenges with SVI data include its spotty availability in some parts of the world, such as Africa, and more importantly, its temporal instability. As most SVI data is either crowdsourced, as in the case of Mapillary [124] or collected by a corporate entity, such as Google Street View, researchers historically have not had influence on when, where, or how often this data is collected. This limits the applications of SVI data, particularly in disaster recovery research, where both image location and capture time are vital to their usefulness.

Repeat photography was used in the disaster recovery domain to measure recovery by capturing images at predetermined locations over a set time interval [27]. This allowed recovery to be tracked at those specific locations, and then estimated by interpolating between the points of interest. While extremely valuable, this method is time-consuming in terms of both data collection and manual processing of images. It is reasonable to hypothesize that we could glean similar insights from repeat SVI data collection. With modern advances in computer vision, it would also be possible to save substantial time on data processing, despite capturing several orders of magnitude more in image data when compared with repeat photography.

In this chapter, we develop an open-source data pipeline for street view imagery and demonstrate its potential value for disaster research. We begin with over 4 million 360° images captured around the city of Seattle along a predefined survey route over a period of 3 years during the height of the COVID-19 pandemic. We then orthorectify the images, and apply a pre-trained pedestrian detection algorithm, yielding a spatio-temporal dataset of pedestrian traffic. As a proof of concept, we analyze the pedestrian traffic data to gain insights into potential drivers of behavior, including vaccine availability, seasonal trends, and socioeconomic traits of different parts of the city. Lastly, we validate the model outputs against known results in the literature using mobile phone and survey data. Future work includes further analysis of the SVI data in a variety of fields, including disaster recovery modeling.

1.2.3 Outdoor Social Distancing Behaviors Changed During a Pandemic

One of the key non-pharmaceutical interventions against the COVID-19 pandemic was social distancing. Early in the pandemic, the United States government recommended avoiding gatherings of more than 10 people, and to work from home if possible. One important element of social distancing is maintaining a larger distance than normal between yourself and others.

The United States Center for Disease Control and Prevention recommended maintaining a physical distance of at least six feet between people, both indoors and outdoors. As time has gone on, the emphasis on these government restrictions has lessened, or the restrictions have been lifted. Similarly, people were less likely to follow these recommendations as time went on, particularly after becoming vaccinated [93, 194].

These behaviors, tied directly to the concept of mobility, can serve as a metric for measuring community recovery from disaster events [91]. Mobility and particularly social distancing behaviors during the COVID-19 pandemic were a heavily researched topic, with both mobile phone data and self-reported survey data as two extremely common data sources [73, 81, 129, 140, 222]. While the findings from these studies are extremely useful, the data sources used have some limitations. Aggregated mobile phone location data is useful for telling the general location of a pedestrian (e.g. if they are visiting a crowded location like a grocery store), but not the location in relation to others. Thus, phone data is not suitable for understanding social distancing behaviors. Survey data can be used to assess social distancing behaviors, but it relies on self-reported metrics, and can be victim to cognitive biases such as recall.

In this chapter, we provide a new data source for measuring social distancing behaviors over time. We utilize the outputs from the SVI data pipeline in chapter 3, and a distance estimation algorithm from Salazar [184], to create a longitudinal social distancing data set for the city of Seattle. We then analyze the longitudinal data to understand community mobility patterns and social distancing behaviors during a three-year period. Given the geospatial nature of the data, this includes analyzing specific points of interest, and different geographic areas. We are able to confirm key findings from previous studies using other data sources, and provide additional perspective on COVID-19 related inequities in the US. Future work can include algorithmic improvements, and analyzing the gap between social distancing recommendations and pedestrian behavior.

1.3 Organization

This dissertation is organized as follows. Chapter 2 presents a review article on the use of empirical quantitative data in lifeline restoration modeling [126], which is published in *Natural Hazards Review*. Chapter 3 describes a data pipeline for street view imagery, and its applications for tracking pedestrian traffic during the COVID-19 pandemic. The work presented is based on a manuscript that is accepted for publication in *PLOS One* [127]. Chapter 4 applies a distance estimation algorithm to the detections data from Chapter 3 to model outdoor social distancing behavior as a metric for disaster recovery, and is based on a working manuscript. The chapters build on each other to advance the overall objectives of the dissertation. The dissertation concludes with a summary of contributions and future research directions in Chapter 5.

Chapter 2

REVIEW OF EMPIRICAL QUANTITATIVE DATA USE IN LIFELINE INFRASTRUCTURE RESTORATION MODELING

2.1 Introduction

Recovery from disasters is considered to be one of the less-understood phases of the disaster cycle [132, 200]. Disaster recovery is a broad term with many facets, including social, economic, built, and natural environments. It is largely accepted that it implies bringing each of these facets to predisaster levels or better [33, 101, 113]. We follow most closely the definition of Lindell [113]. Disaster recovery is a phase of the emergency management cycle that usually overlaps with emergency response. A subsection of disaster recovery research is lifeline restoration modeling. *Restoration* refers to the short-term patching up of essential services to help facilitate longer-term recovery [101, 113, 119]. Lifelines are a subset of critical infrastructures vital for communities to operate [210], namely, electricity, natural gas, telecommunication, transportation, water, wastewater, and liquid fuel [5]. Understanding how these systems are restored allows for more informed community resilience planning efforts [150, 161]. We can understand lifeline restoration processes better through modeling.

The lack of, or perceived lack of, empirical data is one of the primary challenges for the growth of the lifeline restoration modeling field [33, 133]. Ouyang [163] identifies difficult to access data and lack of precise data as key problems for modeling lifeline systems. Lifeline modeling requires many data, frequently including system topologies, component geographical locations, and emergency procedures used by the lifeline system's owners. Data access is difficult for reasons such as security, liability, competition laws, confidentiality, and privacy. Rinaldi et al. [178] also identified the volume of data required to model lifeline systems as a major challenge in the field. Ouyang [163] called for a standardized data collection method to remedy data issues, wheras Miles et al. [132] called for a community of practice to develop around the broader field of disaster recovery modeling, including development of shared data sets. Consistent data collection and management strategies would allow for many more data reuse opportunities and greater access into the field for new researchers.

The need for a consistent approach to handling data in lifeline restoration modeling is apparent. It is necessary to understand the history of data usage in the field. This study reviews the usage of empirical quantitative data to model lifeline infrastructure restoration and provide recommendations for future directions of the field. Section 2.2 discusses high-level trends seen in the literature and the literature search methods used. Section 2.3 breaks down the literature by modeling approach for an in-depth look at how various approaches utilize empirical quantitative data. Section 2.4 discusses topics related to lifeline restoration modeling, such as model validation and testing methods, modeling interdependent systems, benchmarking testbeds and data management best practices.

This review shows relationships between modeling approach, hazard type, lifeline system, and data set features. Additionally, it identifies trends in the field related to modeling interdependencies in the restoration process and alternative data sources, including benchmarking testbeds. The most significant contribution of this review, which separates it from existing reviews such as Ouyang et al. [163] or Miles et al. [132], is its focus on data and data management. To our knowledge, no other review in the literature has this focus. Our review shows the breadth of data features and data sources used in the literature. It can guide researchers regarding the many kinds of data that could be used to model lifeline infrastructure restoration. This focus on data emphasizes the importance of data collection and sharing to scholars and practitioners and hopefully encourage more of them to collect and share data.

2.2 Methods

This section details our methods for identifying publications to include in the review, what data items we collected from each publication, and some high-level trends from the literature as a whole. We identified initial publications to include in the review by searching Web of Science. Web of Science was chosen for the search as it includes the curated reputable journals in relevant disciplines, indexed such as in Science Citation Index Expanded and Social Sciences Citation Index. This allowed us to focus on the peer-reviewed studies that meet a certain scholarly standard amid the recent increase of spurious journals. Keywords used across our searches included terms related to the phases of recovery, different hazard types, and different lifeline systems. The full list of search terms is lifeline, infrastructure, water, wastewater, electricity, power, gas, transportation, telecommunications, outage, restoration, reconstruction, recovery, disaster, hurricane, ice storm, tornado, earthquake, and data.

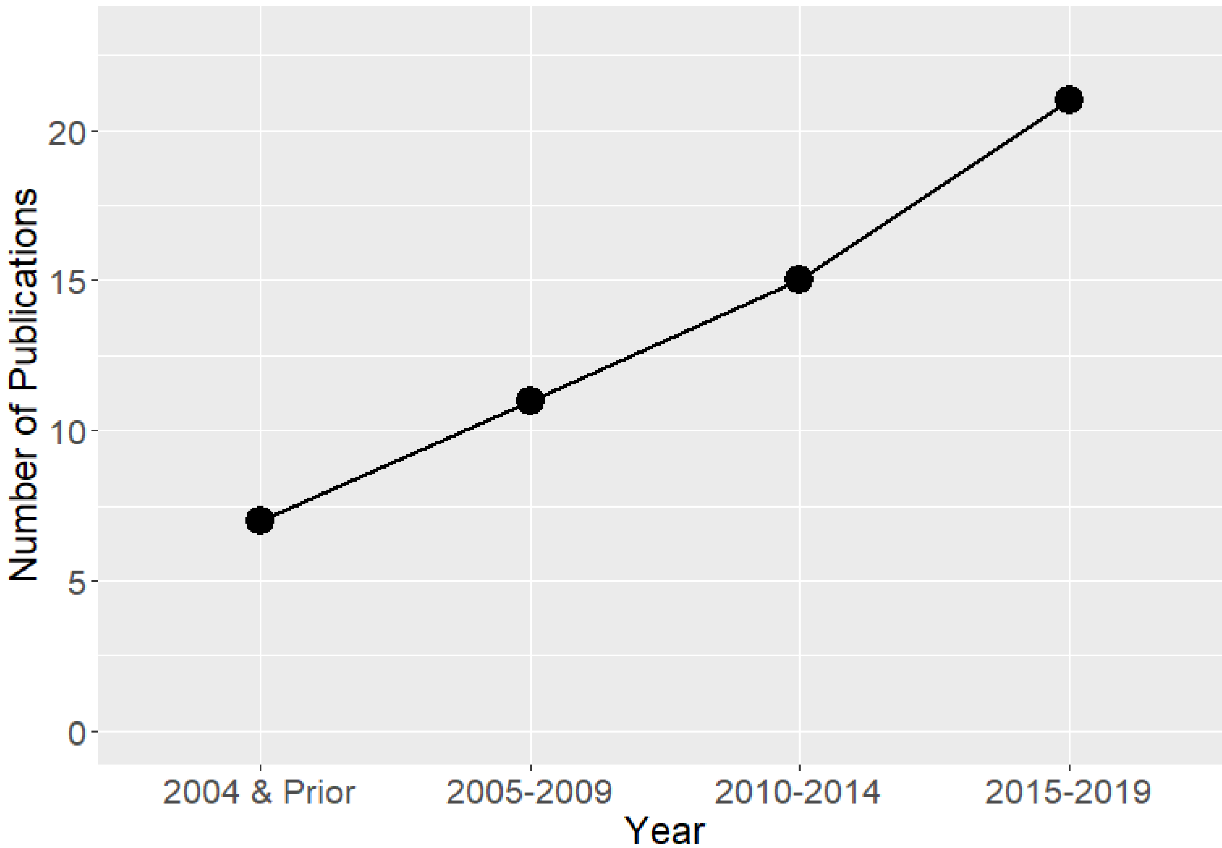
From these searches, publications were included only if they included a quantitative restoration model, that model was for one or more lifeline systems, and the model was grounded in empirical quantitative data in some way. Any publications published in 2019 or earlier were eligible for inclusion. Using the initial qualifying publications from Web of Science, we found older publications that fit the inclusion criteria using backward snowballing. Backward snowballing is a technique for searching the literature by proceeding backward in time through references of known publications to find older sources on a topic [61]. In total, we identified 54 publications that met the inclusion criteria for this study. Publications were identified and reviewed by the first author to determine suitability for inclusion.

As there are inconsistencies in the literature regarding the usage of key terminologies such as restoration, recovery, and response [132], and we only reviewed publications in English, this list may not be exhaustive. However, the inclusion criteria create a representative set of publications on the topic so the study's findings and insights are grounded in major literature trends. A list of all publications included in this review can be found in Table 1. The data items collected from each publication included the modeling approach(es), the hazard(s) of interest, the modeled lifeline system(s), whether the publication considered interdependencies in the restoration process, and the country of origin of the data.

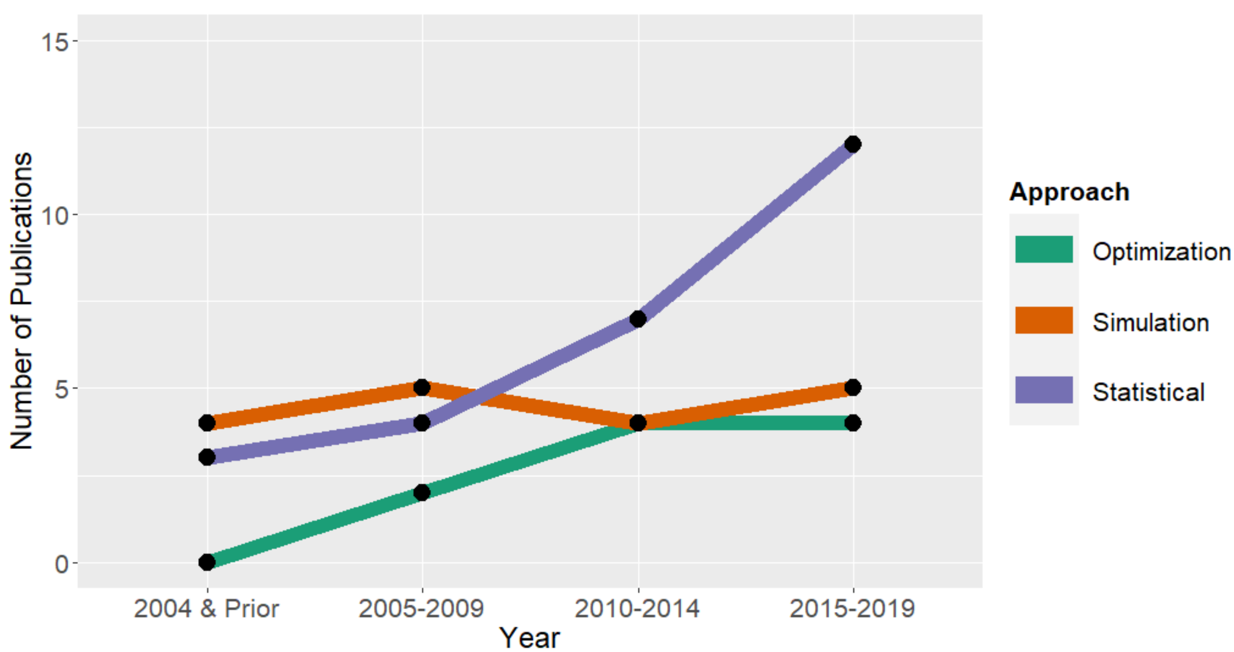
The literature analyzed for this study was a subset of disaster recovery and modeling literature. It is useful to identify some excluded publications to illustrate the boundary of the reviewed literature. Nejat and Ghosh [151] used empirical data to model housing recovery, but their work is excluded from this review because housing is not considered a lifeline. Similarly, publications that model greater community recovery, or other non-lifeline sectors, were not included in this study [11,134]. Works that collect restoration data without building a restoration model such as Nojima and Maruyama [156] were also not included. Additionally, publications that work with qualitative data only, such as expert judgments [34], were not included. Although qualitative data are useful for building quantitative models, the methods for collecting/generating qualitative data are substantially different from those for quantitative data and would be best served with their own review. A large body of literature omitted from this study concerns the power service restoration problem defined by the Institute of Electrical and Electronics Engineers (IEEE), because they use a specific technical definition. The problem is also known as the Fault Isolation and Service Restoration problem. Solutions to this problem try to find the fastest way to isolate a fault in a power distribution network while minimizing the number of healthy out-of-service areas [125]. There are reviews of the literature in this area including Ćurčić et al. [48] and Liu et al. [116], so these articles provide more information on this problem. Many publications in this domain used electricity infrastructure data, so they are a potentially valuable data source. Making exclusions of the above types allows us to keep our scope narrow while still having a significant body of research to review.

2.3 Findings

An initial finding from this review is that lifeline restoration modeling is a growing field. Figure 2.1a shows the marked increase in publications over time. The sharp increase in publications over the last ten years (2010-2019) coincides with the proliferation of statistical models of lifeline restoration. Figure 2.1b shows the change in modeling approaches over time. Statistical modeling grew markedly in the last ten years compared to other modeling approaches. This trend may be related to changes in the amount of available data and what data are being used. The availability of outage/restoration data has likely increased with



(a) Number of publications over time on modeling lifeline restoration using empirical quantitative data.



(b) Number of publications over time by modeling approach.

the increasing number of weather-induced disasters [102]. This increase contrasts with the availability of lifeline-specific data (e.g., topology of a networked system) typically used by simulation and optimization approaches. This type of data has not experienced the same trend in accessibility as outage/restoration data since it requires collaboration with utility companies. Whereas statistical models can use publicly available community attributes, such as demographic information or economic data, as predictors for outage duration, optimization or simulation approaches require lifeline-specific data to model the restoration process. Thus, the growth in statistical models is only natural. The data usage patterns of each modeling approach are discussed in more detail in Section 2.3.

Data availability is not the only factor that affects modeling decisions. Earthquake hazard research is a historically more organized and well-funded research domain than other hazards research. This is exemplified by major earthquake engineering research centers such as the Mid-America Earthquake Center, Pacific Earthquake Engineering Research Center, and the Multidisciplinary Center for Earthquake Engineering Research (MCEER). MCEER, formerly known as the National Center for Earthquake Engineering Research, alone produced hundreds of publications, some of which involve lifeline restoration modeling [146]. Two of the most extensive past restoration modeling and data collection efforts are MCEER projects that involved collaborations with the Los Angeles Department of Water and Power (LADWP) and Memphis Light, Gas and Water Division (MLGW). Both partnerships resulted in multiple publications, so earthquake-related models are heavily represented in the literature as seen in Figure 2.2a. Furthermore, a significant body of literature assumes an initial damaged state without specifying a hazard type, or considers multiple hazards, to make those models more generalizable (Figure 2.2a).

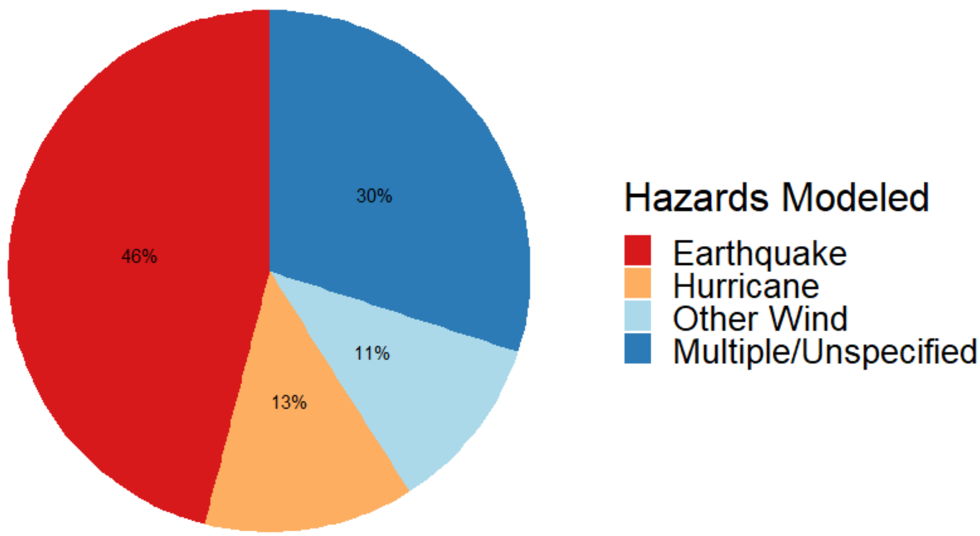
We separated lifeline restoration modeling into three categories for our analysis: optimization, simulation and statistical modeling. Although these categories are broad, there are still enough differences in data usage between them. To facilitate our discussion of data management practices, this section discusses each modeling approach and the common data-usage practices within them. Each modeling approach subsection has three parts. Part 1 focuses on data set features (i.e., what types of data are being used), part 2 focuses on notable data sets (defined as any data set that was used in more than one publication), and part 3 focuses on data sources (i.e., from where the data were obtained). Statistical modeling approaches are the most common, followed by simulation, and then optimization (see Figure 2.2b).

There are clear connections between the modeling approaches and the types of data used. Optimization models most often consider multiple lifeline systems and hazard types, whereas statistical typically are linked to electricity restoration and simulation models typically are linked to earthquakes (see Figures 2.3a and 2.3b). Optimization models are procedural and emphasize generalizability, so they tend to use data sets representing multiple systems and hazards. Statistical approaches to modeling electricity restoration are common because power outage data are more common than outage data for other lifelines. Electricity restoration models often are constructed using outage data and any data used as a predictor (e.g., electricity system features, hazard characteristics, or socioeconomic data about the surrounding community). Simulation modeling of postearthquake restoration is common because of the MCEER research program. The long-term MCEER partnership with the Los Angeles Department of Water and Power (LADWP) yielded high-resolution simulation models of post-earthquake restoration, resulting in multiple publications.

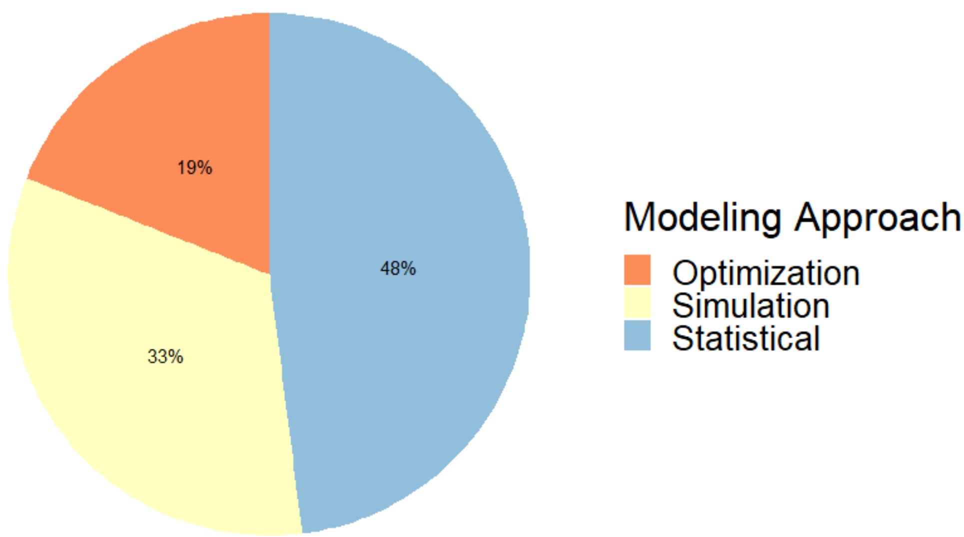
2.3.1 Simulation

Overview

Simulation models have the longest history of any method in the lifeline restoration modeling domain, dating back to the 1980s [96, 97]. Simulation modeling was the second most common modeling approach in the literature reviewed. In terms of data usage, simulation models are typically based on lifeline-specific data such as a connected graph representation of the system, individual component repair times, and available repair resources (e.g., maintenance crews). High-fidelity simulation models require detailed data about all parts of the

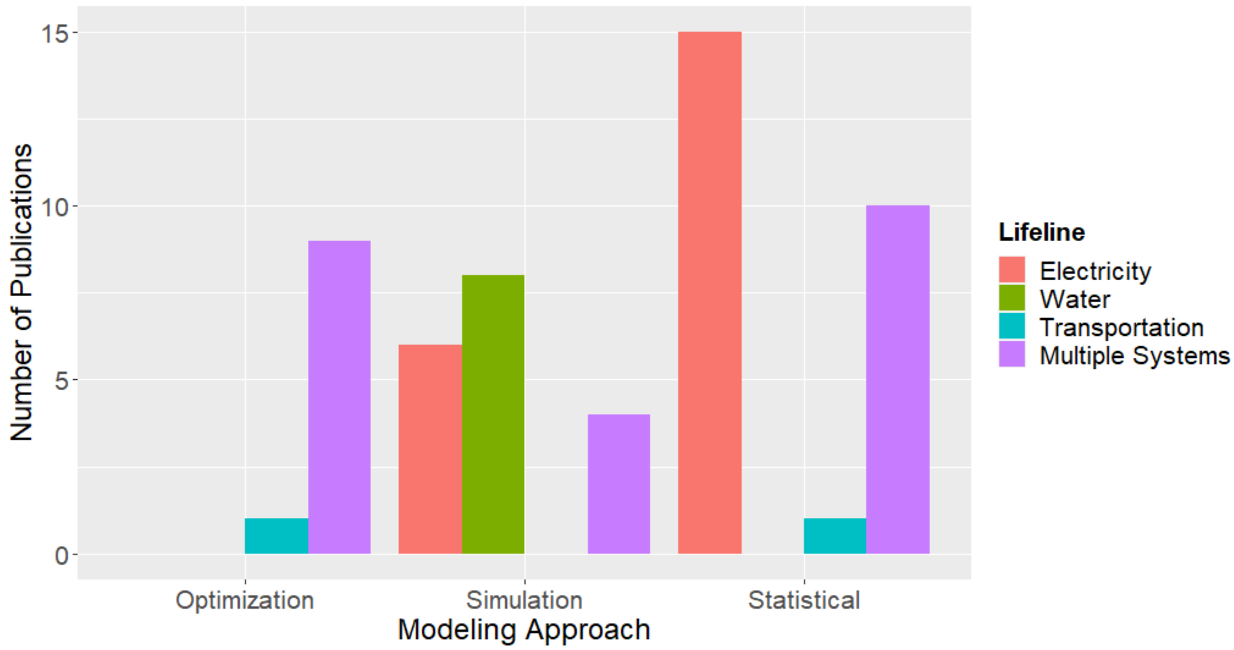


(a) Breakdown by hazard type. ‘Other Wind’ includes ice storms and tornadoes.

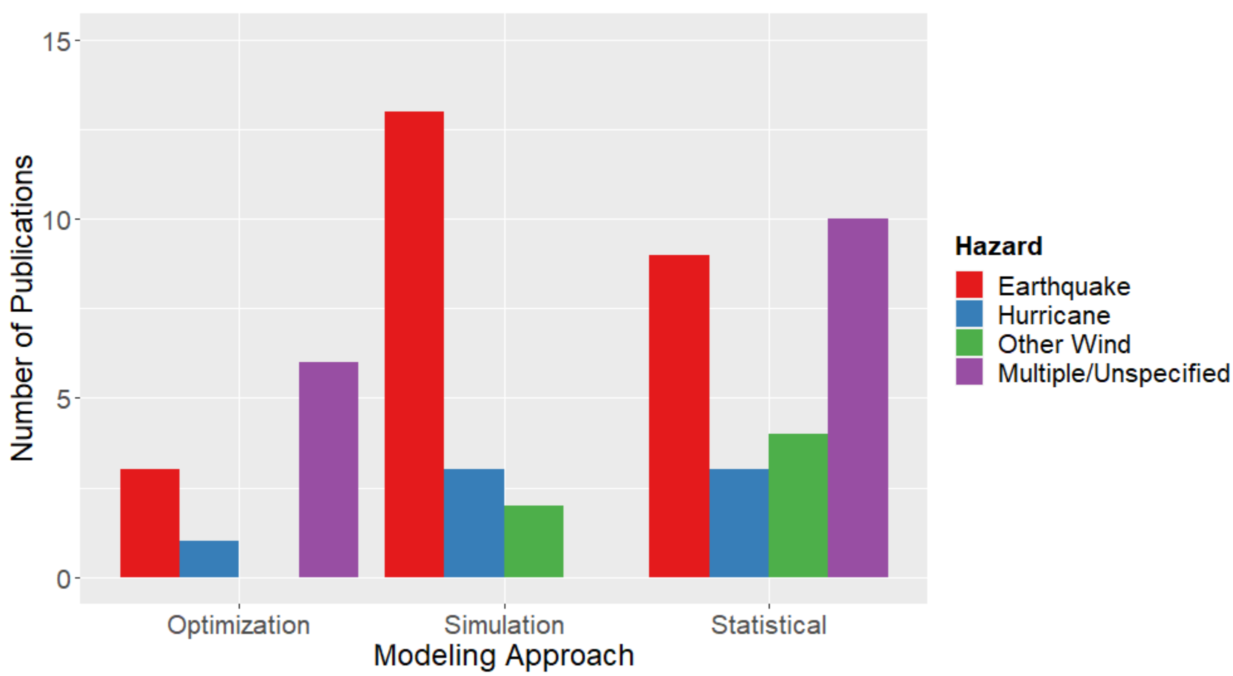


(b) Breakdown by modeling approach.

Figure 2.2: Breakdown of the reviewed literature.



(a) Breakdown of the reviewed literature by modeling approach and lifeline system.



(b) Breakdown of the reviewed literature by modeling approach and hazard type.

Figure 2.3: Relating modeling approach to lifeline system and hazard type.

restoration process, so some of the largest data sets in terms of the number of features are found in this section.

Data Set Features

Data for simulation models come from many different sources. In spite of this, there is a high level of overlap in the features of the data sets. Every simulation-based publication reviewed used lifeline infrastructure data in some capacity. Lifeline systems are commonly represented as connected graphs [29, 39, 96, 174]. Component failure rates are frequently obtained from other works [20,23,96,205]. Another common data set feature is repair crew information such as repair rate, efficiency, and number of crews [20,21,28,29,120,208,227]. Lastly, information about restoration from an outage event primarily is used for model validation and testing. Validation and testing methods are discussed more in Section 2.4.

Notable Data Sets

Two data sets with a large number of features used for simulation modeling are those used to model the restoration of the LADWP systems [20, 21, 28, 29, 208, 209, 227]. The data sets for these publications are the result of extensive collaboration with LADWP. The publications are from two separate projects, one for water restoration [20, 21, 208, 209] and one for power restoration [28,29,227]. The data sets include detailed network representations of the respective lifelines, locations of the various resources necessary for repair work, expected behavior of repair crews, and each repair resource's availability. Additionally, restoration and initial damage data from the 1994 Northridge earthquake serve as the basis for model validation.

Luna et al. [120] studied water supply system restoration from earthquakes using discrete event simulation and a colored Petri nets approach. They use the data set of Isoyama and Katayama [96]. The data set includes the network representation of Tokyo's trunk water supply system, damage probabilities for system components, repair crews, trucks, replacement pipes, and excavators. The authors compare their model against Isoyama and Katayama [96]; however, they do not use baseline restoration data to test the model.

Data Sources

Many data sources are used in simulation modeling studies, although some publications do not identify an original source for their data sets. Sun et al. [205] use an IEEE Bus Test Case for their network data as well as data from HAZUS and previous works for component fragility functions. Several studies in this area [20, 21, 28, 29, 208, 209, 227] collaborate directly with LADWP and collected extensive data sets through interviews and reviewing emergency response plans. Other data sources include HAZUS, S&P Global Platts (a provider of information for commodities markets), public utility data, and government disaster reports. Easily the most common data source in the simulation literature is previous publications [23, 37, 120, 164–166]. Isumi et al. [97] use damage and restoration reports from local government and utility companies. Lastly, Google Earth is an infrequent but inventive data source for identifying lifeline facility network structure [86, 174].

2.3.2 Optimization

Overview

The purpose and data usage of optimization modeling studies differ from those of the other two modeling approaches. The purpose of an optimization model is typically to identify an efficient restoration sequence. In contrast, the purpose of simulation modeling often is to understand a restoration process in greater detail, whereas the purpose of statistical modeling often is to predict outage duration. Optimization models also distinguish themselves from other approaches by more frequently modeling interdependencies between lifelines through model constraints.

From a data usage perspective, optimization models do not put as strong of an emphasis on using empirical data. Compared to other modeling approaches, optimization models are typically focused on proving a theoretical result, which explains the lack of emphasis on data. Real-world data are not strictly necessary to prove a theoretical result, such as optimality, or to show computation times. This is how data sets, such as the one used by Lee et al. [109]

arise, which generate a realistic representation of several lifelines using empirical quantitative and qualitative data together.

Data Set Features

Optimization models are similar to simulation models in that they focus their modeling efforts on the lifeline systems and restoration processes themselves. This focus leads to data sets that take the form of connected graph representations of lifelines. These representations include location and capacity of supply nodes, node-arc lifeline interdependencies, flow capacities, flow costs, and repair costs.

Notable Data Sets

Lee et al. [109] is one of the more frequently cited optimization restoration modeling studies, and the data set they created is reused in multiple other publications [30, 158, 159]. They used data from the U.S. Census, New York City Metropolitan Transit Authority, a local electric company, and Verizon to represent the lifelines in lower Manhattan. This representation includes physical layout, supplies, demands, capacities, interdependencies, and origin-destination information for the transportation and telecommunications networks.

Nurre et al. [158] use the same data set for lower Manhattan as Lee et al. [109], in addition to collecting data about New Hanover County, North Carolina. The New Hanover County data set includes representations of electricity systems, wastewater systems, and emergency supply chain infrastructures. This data set was created with the infrastructure systems' managers and a county emergency manager. All systems are represented as connected graphs; restoration strategies were implemented using the input of emergency and utility managers. Sharkey et al. [191] also used this New Hanover County data set. Iloglu and Albert [95] used a different data set from New Hanover County, representing the road network, locations of fire and rescue stations, and locations of demand for emergency services.

González et al. [75, 76] tested their models on a data set representing Shelby County, Tennessee. The data set contains network representations of the power, water and gas systems of the county. This data set stems from an extensive partnership with a utility company, in this case, MLGW. This partnership yielded a feature-rich data set used in many subsequent studies. It dates back to an MCEER project with many contributors. This data set is discussed in more detail in Section 2.3.3.

Data Sources

The data sources for optimization models are similar to those of simulation models but less varied. Collaboration with lifeline management organizations to get data is a common method [109,158,212,228]. Researchers also consistently make use of data sets collected from prior studies [30, 75, 76, 158, 159, 191], frequently other lifeline restoration modeling efforts. There is less emphasis on data collection and usage than in other modeling approaches, as optimization models are frequently theoretical. Overall, optimization approaches use a similar, but smaller, set of data sources than simulation approaches.

2.3.3 Statistical Models

Statistical models are the most frequently used and most varied of the three modeling approaches. The goal of such a model is usually to generate a restoration time estimate (e.g., it will take four days for the lifeline to be 90% functional), or a restoration probability (e.g., there is an 80% probability the lifeline has 90% functionality in three days). The statistical modeling approaches include curve fitting [168], survival analysis [15, 50, 138], various machine learning techniques [144, 149] and econometric models [122], among others.

With the widest variety of approaches, statistical models also encompass the widest variety of data set features and sources. A commonality amongst the statistical models is lifeline restoration data used for model fitting and model validation and testing. Some larger data sets, in terms of features and the number of disaster events, include power restoration after several hurricanes in the U.S. Gulf Coast region [148, 149] and a data set for power restoration after hurricanes and ice storms for three power companies covering North Carolina, South Carolina, and Virginia [50, 115, 177].

Data Set Features

A common feature of statistical modeling data sets is the use of restoration data from historical disaster events. Sometimes this takes the form of time series restoration data and other times a single data point representing the percentage restoration for a particular geographic area. Lifeline data sets are also used in many studies. Common features for power system data sets include the number of poles, transformers, switches and lines in each grid cell of a spatial data layer [115, 148, 149]. Other common data set features include hazard data such as wind speed, rainfall, and ice accretion and geographic data such as land cover and soil depth [50, 79]. Several studies use socioeconomic data [115, 137], including demographics, population density and poverty rates. Other data set features include commodity trade data and climate data, such as mean annual precipitation [79, 122, 144].

Notable Data Sets

In two studies, Nateghi et al. [148, 149] use a data set representing the Gulf Coast region of the U.S. The data set includes estimates of wind gust speed, duration of wind speed exceeding 20 m/s, land cover, soil moisture, antecedent precipitation and mean annual precipitation. Power system-related features included numbers of poles, transformers, and switches; length of overhead and underground lines; and number of impacted customers. These features were mapped to 3.66 km by 2.4 km grid cells. Restoration information is available for three hurricane events. This multi-event nature combined with the number of features makes this data set one of the largest in the power restoration literature.

One of the most commonly used data sets for modeling various aspects of disaster recovery is that used in Liu et al. [115]. This data set was used for many publications, some of which were not directly modeling lifeline restoration [114], and others extending existing restoration modeling work [177]. The data set includes outage data from three utility companies in the North Carolina area for six hurricanes and eight ice storms. The data set was collected at the county level for land cover, number of customers affected, type of device affected, population density, outage start time compared to start time of the first outage, estimated wind speed, seven-day rainfall and ice accretion. Reed [177] used a subset of this data set and data from the 1999 French winter storms for their model.

Yu and Baroud [231] also utilized a data set from Shelby County, Tennessee. Their data set comprises outage data from fifteen storms for MLGW between 2007 and 2017. Shelby County and MLGW provided data for research in the past that resulted in many extensive works, most notably an MCEER project in the 1990s [36, 193]. The data set presented in Chang et al. [36] comprises layouts for water, electricity and natural gas systems, restoration data for the 1994 Northridge earthquake, utility usage data, census data, and economic data.

Several studies aggregated data from many events worldwide to build their models, and others that focused on a specific geographic area for restoration data. Díaz-Delgado Bragado [53] built a database of restoration data for 31 earthquake events from around the world from 1923 to 2015, considering water, power, gas and telecommunications systems and uses it to fit gamma cumulative distribution functions. Monsalve and de La Llera [139] also compiled earthquake restoration data, encompassing six different earthquakes and various infrastructure systems. Kammouh et al. [100] also collected worldwide earthquake restoration data, for 32 earthquakes. Zorn and Shamseldin [235] collected restoration data from 18 events, including earthquakes, hurricanes, and other types of disasters, for electricity, water, gas, and telecommunications systems. Duffey [60] collected power restoration data for 13 disaster events between 2012 and 2018 using"power tracker" or "outage map" websites.

Nojima and Sugito [154, 155] and Nojima and Kato [152, 153] collected data sets from Japanese earthquake events as the basis for their models. These data sets include seismic intensity from the Japan Meteorological Agency, spatially distributed population data and network vulnerability data for water and gas systems. Restoration data sets for electricity, water, and gas systems also were used. The data sets are collected from the 1995 Hyogoken-Nambu earthquake and the 2011 Tōhoku earthquake.

MacKenzie and Barker [122] used publicly available U.S. outage data collected by the U.S. Department of Energy through form OE-417, along with state population data. The data set includes duration, location (state), and cause of the outage between January 2002 to June 2009. Barker and Baroud [10] and Barabadi and Ayele [9] used the same data set, while Mukherjee et al. [144] utilized a larger data set of OE-417 submissions, containing information from January 2000 to July 2016. They used state-level population data, climate data from the U.S. National Oceanic and Administrative Administration, electricity consumption patterns from the U.S. Energy Information Administration, Urban/Rural and Land/Water percentages from the U.S. Census Bureau, and state-level economic characteristics from the U.S. Bureau of Economic Analysis.

Data Sources

Direct collaboration with utility companies is again a common data source [36, 46, 137, 148, 149, 168]. Nateghi et al. [148, 149] supplemented their utility-provided data with data collected from a commercial weather forecasting service and the National Land Cover database. Mitsova et al. [137] collect additional data from the American Community Survey for their model. Several modelers got their data sets from public U.S. government data sources [9, 10, 122, 144]. The most common data source is data sets from previous studies, such as the worldwide restoration data sets in Zorn and Shamseldin [235] and Kammouh et al. [100]. Using a novel approach, Duffey [60] used "outage tracker" websites to gather restoration data after multiple disasters. Sources outside the U.S. are used in several works [9, 15, 138]. Public outage reports are used in Duffey and Ha [59].

2.4 Future Directions for Research

We identified several important topics for discussion from this literature review. These topics can be broken into thematic and methodological directions. The two thematic directions we identified were modeling interdependent systems and engagement through benchmarking testbeds. The methodological directions are alternative data sources, model validation and testing, and data management best practices. These topics have relevance to the future directions of the lifeline restoration modeling literature.

2.4.1 Thematic Directions

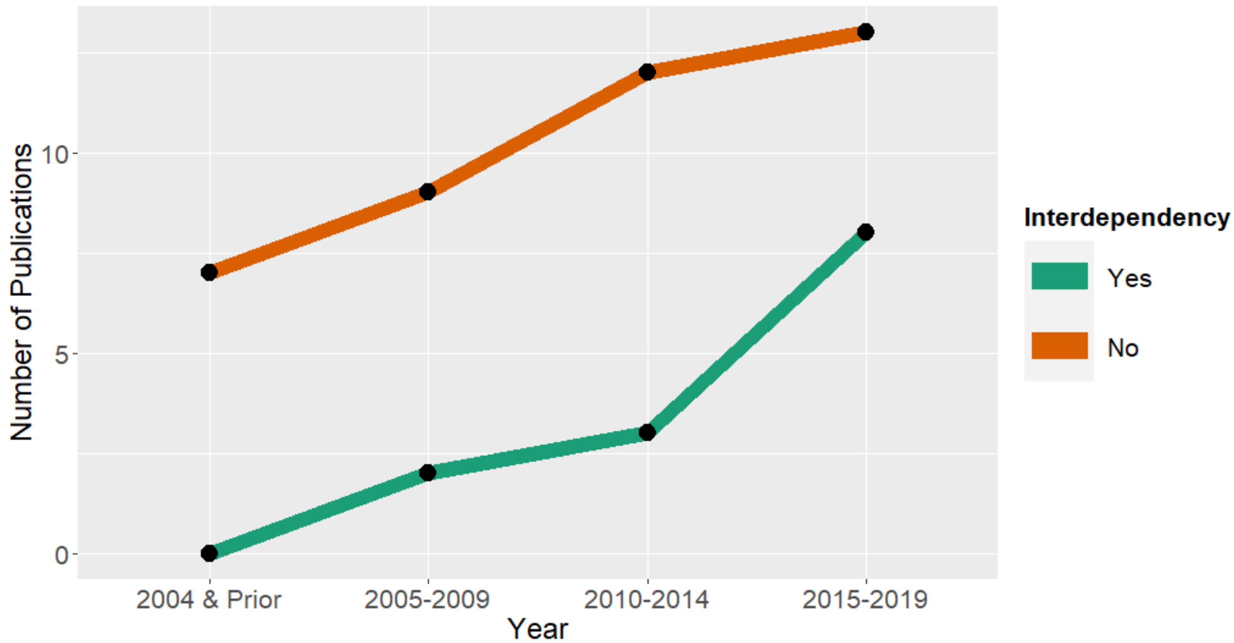
Modeling Interdependent Systems

Interdependency is increasingly recognized as an important factor to consider while modeling lifeline restoration, as seen in Figure 2.4a. Lifeline systems are interdependent by nature. For example, power generators require water for cooling and electricity is needed for water pumps to function. Quantifying these interdependencies regarding restoration is an ongoing challenge for modelers, but one that is actively being worked on by researchers.

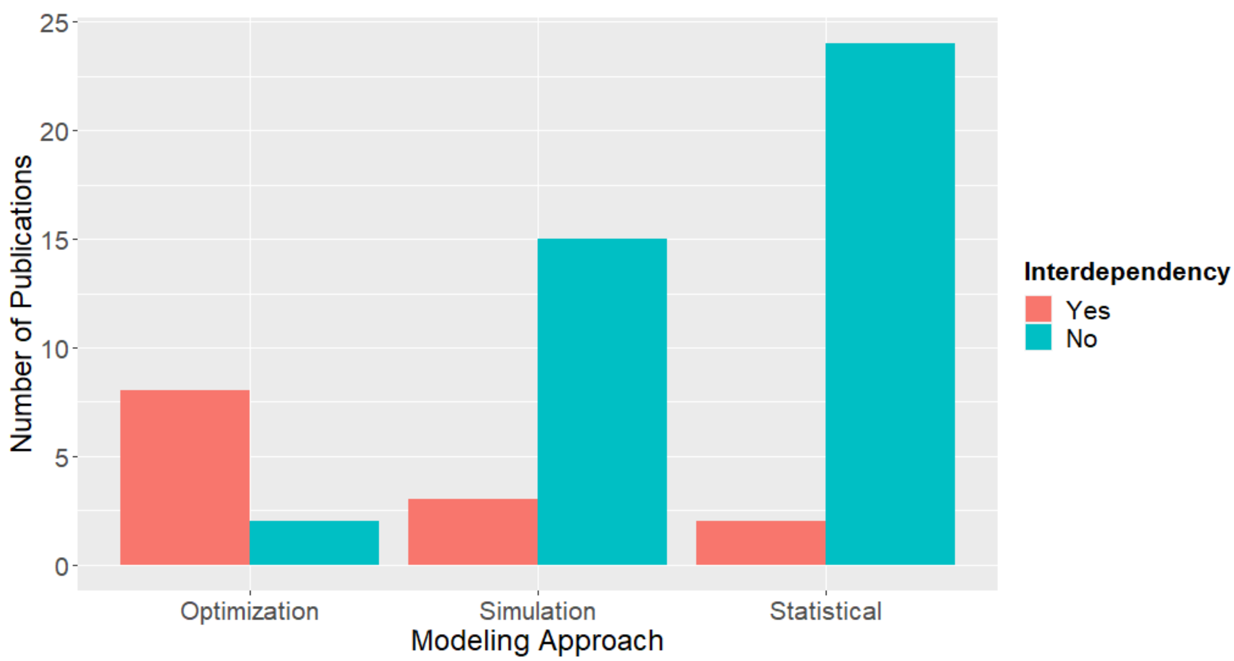
There was an increase in studies of cascading failures [89, 217, 225] in recent years, and there is a broad recognition that lifelines are restored in an interdependent fashion [178, 192]. In contrast, our review shows that only about 25% of the reviewed literature consider interdependencies directly. Optimization models have the longest history incorporating interdependencies in their models, as seen in Figure 2.4b. The rest of this section discusses a few of the methods used to model interdependent restoration in the reviewed literature and promising approaches that, to our knowledge, have yet to be applied in a restoration modeling context.

Lee et al. [109] is the oldest instance of modeling interdependent infrastructure restoration in the reviewed literature. They considered power, telecommunications, and transportation systems in modeling five types of interdependency: input dependence, mutual dependence, shared dependence, exclusive-or dependence and co-located dependence. The authors include interdependencies as constraints in their problem formulation. Çavdaroğlu et al. [30] utilized the same data set but also determined an optimal restoration sequence for the lifeline systems. Their objective was to maximize the functionality of the lifeline services over the restoration period by balancing unmet demand costs and operating costs. They also model restoration interdependencies through their model constraints.

Yan and Shih [228] modeled transportation restoration and emergency relief distribution together. While not a model of interdependent *lifeline* restoration, their work demonstrated a way to model restoration of interdependent *systems*. They used a multi-objective optimiza-



(a) Number of publications over time that model interdependent restoration.



(b) Breakdown of publications by modeling approach on interdependent restoration.

Figure 2.4: Interdependency in the reviewed literature.

tion model to minimize the length of time for restoration and subsequent relief distribution. The authors noted the connection between the transportation system and the ability to distribute relief.

MacKenzie and Barker [122] utilized the dynamic inoperability input-output model (DIIOM) to include interdependency in their restoration model. Interdependencies in the DIIOM are quantified using commodity flow data from the U.S. Bureau of Economic Analysis. MacKenzie and Barker applied the model to estimate restoration from power outages. Theirs is the earliest nonoptimization approach to modeling restoration interdependency in the reviewed literature. He and Cha [86] extended the DIIOM to calculate facility-level interdependencies as opposed to system-level interdependencies in the traditional DIIOM. This facility-level approach captures interdependencies not only *across*, but also *within* systems.

Other more recent models used a variety of approaches for modeling interdependent restoration [75, 76, 139, 174]. Monsalve and de La Llera [139] calculated a daily restoration rate for each lifeline in their model based on the lifeline type, its interdependencies, and an additive Gaussian error term. The authors utilized a least-squares criterion that minimizes the difference between the expected value of the model and the data to estimate model parameters, including lifeline interdependencies. Their model assumes that the restoration rate of a given lifeline depends on the functionality of other lifeline systems but not on their restoration rates. González et al. [75, 76] defined four types of interdependencies: logical, physical, cyber, and geographic. They accounted for these interdependencies through the constraints of their optimization model. Ramachandran et al. [174] included interdependency in their simulation model by including constraints that some tasks cannot start until others finish (e.g., power lines cannot be repaired until the road to access those lines is free of debris).

A series of studies utilized time-series restoration data and cross-correlation functions to quantify the interdependency between two lifelines [40, 58, 107]. This method of quantifying interdependencies has not been incorporated into a restoration model, but there is the potential to do so. We believe that it could be applied in an approach similar to that of

Monsalve and de La Llera (described above) [139].

Engagement through Benchmarking Testbeds

Over the last few years, there has been significant progress creating benchmarking testbeds for recovery modeling. Two examples are Customizable Artificial Community (CLARC) County, created by Loggins et al. [119], and Centerville, created by the Center for Risk-Based Community Resilience Planning at Colorado State University [44]. CLARC County is a GIS data set representing an artificial hurricane-prone community of 500,000. The data set contains demographic and geographic data typically reported for U.S. census tracts and physical locations and characteristics of components of civil and social infrastructure systems along with their interdependencies. The data set exists to support infrastructure and emergency management research without compromising potentially sensitive information. The Centerville community resilience testbed is a virtual city, representing a typical middle-class city in the Midwestern U.S. that is susceptible to tornadoes and earthquakes. Buildings, transportation systems, electric power, and water systems are represented in the data set, and socioeconomic features based on American Community Survey data for Galveston, Texas, and income data from Fort Collins, Colorado.

These testbeds are conducive to recovery research, as they allow for complete, although synthetic, data sets to be used to test and compare recovery models. The two examples provided here also show that testbeds can be constructed in various ways, ranging from being completely synthetic to being based on empirical data from a single source or an amalgamation of sources. The areas represented by the example testbeds are different, one being an individual city, while the other a U.S. county. No matter the construction, these testbeds can provide value as boundary objects for comparison, if nothing else. Given how recent these efforts are, it is unclear if the development of testbeds affects the use and collection of empirical data.

The difficulty of collecting extensive data sets for lifeline restoration modeling is well documented in the reviewed literature. Loggins et al. [119] mention an extensive data collection process they attempted for New Hanover County, NC, and how the difficulties they experienced led them to create CLARC county. The likelihood of developing complete data sets on all lifelines in a community *and* lifeline restoration data from a disaster event in that community is low. Even if such a data set were to be developed, security concerns might prevent it from ever entering the public domain. This makes testbeds the logical next step for the developing large-scale, highly detailed optimization and simulation models of interdependent recovery. However, this does not eliminate the need for the collection of empirical data.

The data collection of Loggins et al. for New Hanover County informed the creation of CLARC county [119], and Centerville [44] was created from an amalgamation of several empirical sources. Data availability can and sometimes should inform modeling approaches depending on modelers' objectives, although models built with no empirical data can still provide useful insights and create new knowledge (e.g., what-if analysis, facilitation of discussion, and education). Examples of data availability informing model choice include the work done with LADWP. The authors had access to a feature-rich lifeline-specific data set, which made a detailed simulation model feasible. Another example of data availability informing modeling efforts/direction is the work of Mukherjee et al. [144]. The authors had access to publicly available data at the state level, making possible a broader statistical model. Having the data set publicly available means others can duplicate and extend this work. There are also many examples of benchmarking in the literature where authors extend the modeling efforts of previous work using the same data set and compare the results.

2.4.2 Methodological Directions

Alternative Data Sources

Some studies that fall outside of this review's inclusion criteria still deserve mention for their usage of data sources not seen in the reviewed literature. McDaniels et al. [128] and Chang et al. [35] characterized lifeline failure interdependencies using manual content analysis of newspapers and technical reports. In contrast, Lin et al. [112] used natural language processing to analyze newspaper stories from New Zealand after the Canterbury earthquakes to track long-term recovery. Doubleday et al. [56] used daily bicycle and pedestrian activity as an indicator of disaster recovery. Burton et al. [27] used repeat photography to characterize recovery after Hurricane Katrina. Brown and Pinkerton [24] used synthetic data for power system elements. Chang et al. [34] used expert elicitation to characterize lifeline resilience. Expert elicitation plays an important role in statewide resilience initiatives [160, 221], and in the development of the Federal Emergency Management Agency's HAZUS [67].

All of the above approaches do not rely on empirical data directly related to lifelines or lifeline restoration. In particular, approaches such as those seen in Doubleday et al. [56] are promising because they make use of empirical quantitative data that has not been used in the restoration modeling space. If data sets of this nature can be linked to lifeline restoration data, the total amount of restoration data sets available would increase. Partially synthetic power system data sets such as those used by Brown and Pinkerton [24] also show how new data can be generated to meet an existing need. There is already a significant body of literature related to generating synthetic power system data sets [17, 18, 71].

The approach of Lin et al. [112], using natural language processing to generate recovery data, has the potential to create many new data sets that could be used for restoration modeling. While their analysis is focused on long-term recovery, a similar approach could be used for modeling shorter-term restoration, perhaps using a different source such as Twitter data [135, 173, 236]. Expert elicitation is another method that can be used to develop restoration models. Models based on expert judgment can apply techniques such as Cooke's method [45] to create a systematic approach for eliciting expert knowledge when empirical data sets are unavailable or inaccessible.

Validation and Testing Methods

As a precursor to this section, we want to acknowledge that model validation is a contested concept with many definitions, recommendations, and best practices across disciplines [181, 223]. There is a tendency to think that every model was developed to predict, and thus every model should be validated using out-of-sample testing. However, there are many reasons for modeling outside of prediction [64]. Because this review discusses the use of empirical quantitative data, reviewed publications used data for model calibration, validation, or application through a case study. Acknowledging that out-of-sample testing is not applicable or feasible for every modeling study, this subsection discusses what out-of-sample validation and testing techniques have been used in the field to date.

Statistical models have the widest variety of out-of-sample validation and testing approaches. Cross-validation is used in a few models for parameter fitting or model comparison [79, 122, 148, 231]. Some modelers split their data sets into training and test sets by withholding information from some disaster events [50, 115, 149]. Park et al. [168] fit a curve to restoration data from one event and compared the fitted parameters to that of another event.

Several out-of-sample validation methods have been used by simulation models as well. The projects that partnered with LADWP [20, 21, 28, 29, 208, 209, 227] used restoration data from the 1994 Northridge earthquake. The studies performed the validation by setting model input parameters (e.g. number of repair crews) equivalent to the Northridge conditions and comparing the simulated restoration time to the actual restoration time from the Northridge event. Other studies that compare their model output to restoration data included Isumi et al. [97] and He and Cha [86]. A comparison of model output for a theoretical disaster event and restoration data from a similar disaster event in a different location [174] is one of the more inventive validation methods seen in the literature.

There are no optimization models in the reviewed literature that were tested out of sample. However, given the nature of an optimization model, this should not come as a surprise. The goal of optimization is usually to perform better than the status quo. Thus the restoration time estimates from an optimization model nearly always are below real-world restoration times. The contribution of an optimization model is typically a new model formulation [158, 228] or solution approach [76].

Overall, the reviewed literature encompassed a wide variety of validation and testing approaches. Although we encourage the use of out-of-sample model evaluation, we understand that this is not always possible, nor does it always make sense. However, as models continue to become more generalizable and data more available, we hope to see more out-of-sample evaluations take place in this field.

A Data Management Methodology for Reproducibility in Disaster Recovery Modeling

The research community benefits from reproducibility, which is fostered by detailed metadata and data publication. Within the analyzed publications, data descriptions frequently were lacking, and data sets rarely were published. Although it is not always possible to publish data sets for a wide variety of reasons (e.g., security and privacy), this reduces the reproducibility of any research using those data sets. As data becomes increasingly prevalent across research domains, there are more advocates for increased accessibility of data sets. Gentleman and Lang [72] go so far as to call for the publication of "reproducible research compendiums" which include the final paper, as well as the data set, software, and any other items necessary to reproduce the research. They acknowledged that this is not feasible for all research but maintained that publishing as much information as possible is worthwhile.

The findability, accessiblity, interoperability and resuse of digital assets (FAIR) guiding principles for scientific data management and stewardship [224] represent a minimal set of domain-independent data management principles. Specifically, these principles state that research objects [13] should be findable, accessible, interoperable, and reusable, both for human-driven and machine-driven activities. These principles purposefully are kept minimal to make the barrier to entry as low as possible to allow for easy implementation, even in cases where data sets cannot be published in their entirety or at all. With these principles in mind, the remainder of this section is devoted to proposing a methodology for writing the data description section of a research paper in the disaster recovery modeling domain and best practices for data publication.

González-Barahona and Robles [77] discussed the reproducibility of empirical software engineering studies and identified elements of said studies with an impact on reproducibility. We adapted their ideas to fit the disaster recovery modeling domain. We recommend all disaster recovery modeling publications using empirical data include a data description section, with at least the following components:

1. Data source(s). Where did the data set come from? This should be as specific as possible. Even if the only thing an author can share is "an unnamed utility company from the U.S. Southeast", that is still worthwhile for a reader to know. Where possible, links or citations to the original data source(s) should be included.

2. Retrieval method. How was the data set collected from the source? Examples include downloading a CSV or GIS file from a government website, receiving data via email, and using a web scraper.

3. Raw data set and metadata. Can the data set be shared or is it publicly available? If yes, we recommend linking to an online repository that includes a description of the format and features of the data set. If not, we recommend providing the metadata of the data set and the proper procedure to gain access to the data set if there is one. We recommend following the Dublin Core [57] standards for metadata. These standards include a cross-disciplinary list of properties for use in resource descriptions designed to be machine-processable that suit modern metadata needs.

4. Data processing methods. For example, what transformations were performed on the data set to adapt it into a usable form? Were features standardized? How were missing entries handled?

5. Processed data set and metadata. What are the form and features of the processed data? Processed data sets should be stored separately from the raw data set in an online repository with a link for others to access. Obtaining a Digital Object Identifier (DOI) for the data set is a best practice after publishing it on an online repository

or in a peer-reviewed journal (e.g., [143]). Version-controlled documentation can track problems in the data sets as they are found and corrected. As with the raw data set, if the processed data set cannot be published, we recommend to at least provide the metadata.

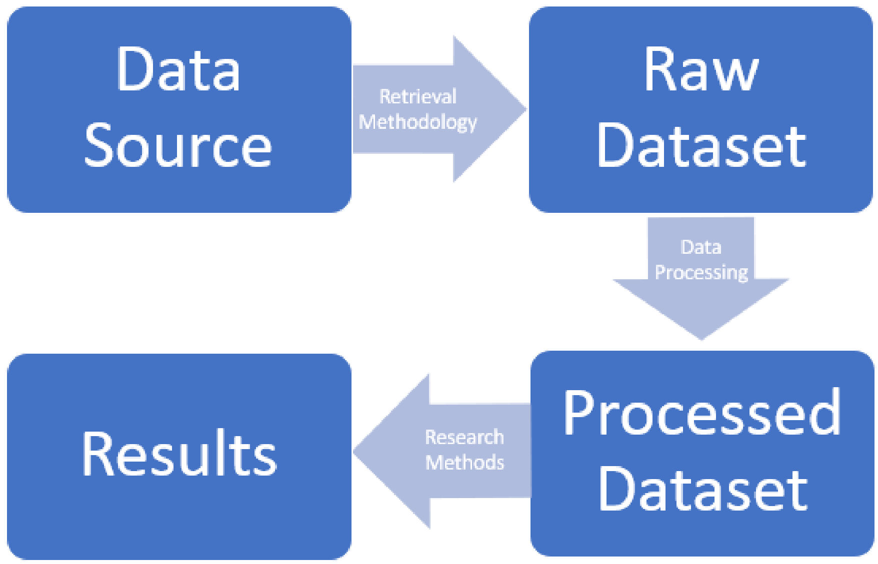
6. Processed data set's summary statistics. If the data set is numerical, statistics can be presented in a table. If the data set is only a network representation of an infrastructure system, a graphical representation may be sufficient. If this information cannot be shared, it can be clearly stated with a reason (e.g., national security).

This methodology is visually depicted in Figure 2.5. One example of a brief, comprehensive data description in an academic publication is from Yan and Shih [228]:

> The roadway-network information includes the roadway segments and intersections in Nantou County, the location of repair points and work stations, and the location of supply and demand points. The emergency repair resources include the work teams for each station and the average time required for a work team to repair each repair point. Note that, in practice, when scheduling the roadway repairs, information on the time needed to repair every repair point is given by the engineers. Several engineers estimate the repair time in advance based on experience and the level of damage. Decision-makers then use the average time (as was done in this research) to set the repair time. ... There are 46 intersections, 24 repair points, 8 demand points, 9 work stations, 24 work teams, 5 distribution centers, and 196 time unit lengths (3 days is the time length, with a time unit of 15 min), in the tests.

This description does not contain all the elements in our proposed data management methodology, but it shows what can be done with a small amount of space in a research publication. The authors do mention a specific data source in their acknowledgments section.

Figure 2.5: Graphical representation of the proposed data management methodology.



When it comes to publishing data sets, following the FAIR principles [224] should be the minimum standard we strive to achieve. An emphasis on these principles is enhancing the ability of machines to find and access data sets automatically. We also would like to emphasize this point and recommend that researchers in the disaster recovery modeling domain publish their data sets on digital platforms that are easily machine-accessible. One domain-specific example of this is the National Science Foundation's Natural Hazards Engineering Research Infrastructure, DesignSafe-CI [175]. DesignSafe-CI is a cyberinfrastructure environment for research in natural hazards engineering. The features of this cyberinfrastructure include data sharing and publication, integrated data analysis tools, high-performance computing access, and collaboration tools. DesignSafe-CI has detailed data publication guidelines and guidance for writing data management plans available.

We recognize that it is not always possible to share all the information in our proposed methodology due to privacy or security concerns. Even under this constraint, it is still important to make clear data management practices for research. If the data set is private, one can still provide relevant metadata within the limitations of the data set provider. This creates an opportunity for future research to collect a different data set with the same features and apply the method used in the original study. Overall, there are many opportunities for increased data sharing and higher standards for reproducibility in the field of disaster recovery modeling. The proposed methodology further supports a community of practice around data management in disaster recovery modeling. Using resources like DesignSafe-CI is one way to make the disaster recovery modeling community of practice a reality.

2.5 Summary

The data sources and data features used by lifeline restoration modelers vary across modeling approaches and there is no standardized methodology for utilizing data in this research domain. This review highlights various data sets that have been used in the past to model lifeline restoration to help build a community of practice within the broader field of disaster recovery modeling. We propose a set of best practices for managing and writing about data sets used for disaster recovery modeling. These best practices rely on data access and availability. Practitioners in emergency management or infrastructure operation can significantly aid disaster recovery researchers by collecting and sharing any recovery-related data with them. A stronger partnership between the research and practicing communities is necessary for the field to move forward.

Our review shows that direct collaboration with utilities and publicly available data, usually from the government, were two of the most common data sources in the literature. Data sets are frequently reused over time to provide additional insights with new/updated modeling approaches. We discussed the usage of benchmarking testbeds as an alternative way to develop and test recovery models where relevant data sets are unavailable. Expert elicitation and large-scale text data sets were identified as additional alternative data sources. Overall, this review demonstrates the wide variety of data sources available to modelers. Some limitations of our review include that the review may not have identified every publication in this domain. This exclusion could happen due to using only one database (i.e., Web of Science) for identifying publications and the inconsistencies regarding the usage of terms such as restoration, recovery, and response. Additionally, nearly two thirds of the publications in the review used data from the U.S. Two potential causes for this include the availability of data in other countries and our inclusion only of articles published in English.

Our intent for the proposed data management practices is to cause more data sets to become publicly available. These data practices can guide modelers without much experience in disaster and hazard research to enter the research domain and open doors for disaster and hazard researchers to build models with more data than that to which they previously had access. With more and more data available, the goal of a generalizable model of interdependent restoration could be realized, with communities around the world as the beneficiaries.

Table 2.1: List of every publication included in review, in order of publication year. E - electricity, W - Water, G - Gas, Tel - Telecommunications, Tr - Transportation, WW - Waste Water, ESC - Emergency Supply Chain, F - Fuel, M/U - Multiple/Unspecified, OW - Other Wind.

Reference	**Approach**	**Lifeline**	**Hazard Type**	**Country**
Isoyama and Katayama [96]	Simulation	W	Earthquake	Japan
Isumi et al. [97]	Simulation	E, W, G	Earthquake	Japan
Chang et al. [36]	Statistical	E, W, G	Earthquake	U.S.
Brown et al. [23]	Simulation	E	OW	U.S.
Cooper et al. [46]	Statistical	E	M/U	U.S.
Chang et al. [37]	Simulation	W	Earthquake	U.S.
Nojima and Sugito [154]	Statistical	E, W, G	Earthquake	Japan
Nojima nad Sugito [155]	Statistical	E, W, G	Earthquake	Japan
Park et al. [168]	Statistical	E	Earthquake	U.S.
Çağnan et al. [29]	Simulation	E	Earthquake	U.S.
Liu et al. [115]	Statistical	E	OW	U.S.
Lee et al. [109]	Optimization	E, Tel, Tr	M/U	U.S.
Xu et al. [227]	Simulation	E	Earthquake	U.S.
Çağnan and Davidson [28]	Simulation	E	Earthquake	U.S.
Reed [177]	Statistical	E	OW	U.S.
Tabucchi and Davidson [209]	Simulation	W	Earthquake	U.S.
Yan and Shih [228]	Optimization	Tr, ESC	Earthquake	Taiwan
Brink et al. [20]	Simulation	W	Earthquake	U.S.
Tabucchi et al. [208]	Simulation	W	Earthquake	U.S.
Nateghi et al. [149]	Statistical	E	Hurricane	U.S.

Continued on next page

Table 2.1 – *Continued from previous page*

Reference	**Approach**	**Lifeline**	**Hazard Type**	**Country**
Luna et al. [120]	Simulation	W	Earthquake	Japan
Ouyang et al. [165]	Simulation	E	Hurricane	U.S.
Brink et al. [21]	Simulation	W	Earthquake	U.S.
Nurre et al. [158]	Optimization	E, WW, ESC	M/U	U.S.
Nojima and Kato [152]	Statistical	E,W,G	Earthquake	Japan
MacKenzie and Barker [122]	Statistical	E	M/U	U.S.
Çavdaroğlu et al. [30]	Optimization	E, Tel	M/U	U.S.
Duffey and Ha [59]	Statistical	E	M/U	U.S., Sweden
				Belgium, France
Nateghi et al. [148]	Statistical	E	Hurricane	U.S.
Barker and Baroud [10]	Statistical	E	M/U	U.S.
Tuzun Aksu and Ozdamar [212]	Optimization	Tr	M/U	Turkey
Nojima and Kato [153]	Statistical	E, W, G	Earthquake	Japan
Nurre and Sharkey [159]	Optimization	E, Tel	M/U	U.S.
Sun et al. [205]	Simulation	E	Earthquake	U.S.
Zorn and Shamseldin [235]	Statistical	E, W, G, Tel	M/U	New Zealand
				Worldwide
Ramachandran et al. [174]	Simulation	E, W, Tel, F	OW	U.S.
Sharkey et al. [191]	Optimization	E, W, WW, Tel	Hurricane	U.S.
Ouyang and Wang [166]	Simulation	E, G	Hurricane	U.S.
Bessani et al. [15]	Statistical	E	M/U	Brazil
González et al. [76]	Optimization	E, W, G	Earthquake	U.S.
Díaz-Delgado Bragado [53]	Statistical	E, W, G, Tel	Earthquake	Worldwide
Mojtahedi et al. [138]	Statistical	Tr	M/U	Australia
Davidson et al. [50]	Statistical	E	OW	U.S.

Continued on next page

Table 2.1 – *Continued from previous page*

Reference	**Approach**	**Lifeline**	**Hazard Type**	**Country**
González et al. [75]	Optimization	E, W, G	Earthquake	U.S.
Mitsova et al. [137]	Statistical	E	Hurricane	U.S.
Choi et al. [39]	Simulation	W	Earthquake	South Korea
Mukherjee et al. [144]	Statistical	E	M/U	U.S.
Barabadi and Ayele [9]	Statistical	E, Tr	M/U	Iran, U.S.
Kammouh et al. [100]	Statistical	E, W, G, Tel	Earthquake	Worldwide
Iloglu and Albert [95]	Optimization	Tr, ESC	M/U	U.S.
He and Cha [86]	Simulation	E, W, Tel	Hurricane	U.S.
Monsalve and de La Llera [139]	Statistical	E, W, G, Tel	Earthquake	New Zealand Japan, Chile
Yu and Baroud [231]	Statistical	E	OW	U.S
Duffey [60]	Statistical	E	M/U	Ireland, U.S. New Zealand

Chapter 3

OPEN-SOURCE DATA PIPELINE FOR STREET-VIEW IMAGES: A CASE STUDY ON COMMUNITY MOBILITY DURING COVID-19 PANDEMIC

3.1 Introduction

One way to assess recovery from disasters is through repeat imaging [27, 88, 130]. Repeat imaging provides primary data that can be compared across different disasters and, with modern technology, efficiently collected over time. One type of imagery that is becoming increasingly popular data source for research purposes is street-level imagery [41]. Between 2009 and 2020, more than 200 publications utilized street-level imagery from corporate sources in urban research [41]. Out of all these sources, Google Street View's Street View Images (SVI) were the most popular among academics [41, 74, 111]. Uses for SVI data include tracking disaster recovery [88], estimating demographics [70], evaluating the built environment [6], measuring pedestrian volume [230], among many other applications [157, 179, 183, 219].

While SVI data can provide many useful insights for researchers, it is not without its flaws. For corporate-collected images such as Google Street View, or Tencent Street View the availability of images depends on where the companies decide to collect data, while the accessibility of these images hinges on the companies' data provision policies. For example, there is no Google Street View service in most parts of Africa. An alternative to corporate-collected images are crowdsourced SVI databases such as Mapillary [124]. These crowdsourced images sometimes may have better coverage or temporal resolution than Google Street View, at the cost of varying image quality, field of view, and positional accuracy [111, 123]. Perhaps the largest challenge with SVI data is its temporal instability. Updates to these image datasets at specific locations are infrequent, especially in rural areas [41, 198, 203]. Additionally, images frequently are not collected at a consistent time of day, or season, even within the same city. These issues make existing SVI data unreliable for temporal studies.

Typically, temporal studies involving image data use images (or video) from fixed locations. This data is used to do things such as evaluate disaster recovery [27], monitor ecological change [52], or measure urban flooding [83]. Data from fixed cameras is also used to count people [216]. The challenge with these methods is that they are fixed-location. In order to collect spatial image data for these methods, frequently a large team is required to traverse areas on foot. This challenge, along with existing SVI data's temporal issues, demonstrate the potential value of collecting longitudinal SVI data.

Our main contribution is demonstrating the feasibility of collecting longitudinal SVI data. We demonstrate this through the creation of a complete data pipeline for conducting pedestrian counts using car-based street-level imagery. The pipeline accepts raw video collected by the camera as an input and outputs a record of each pedestrian detection and their locations (latitude and longitude). This approach allows for analysis of mobility patterns with high spatial resolution and a short lag time. It alleviates the quality and field of view inconsistencies that come with crowdsourcing SVI data [111,123], generates data that is not corporately owned, eliminates the temporal instability challenge of both kinds of data [41,198,203], while still maintaining the advantages of SVI data over fixed-location methods [27].

Specifically, we use this pipeline to generate and analyze video from 37 video-collection runs in the city of Seattle, Washington, USA from May 2020 through July 2023. The video data was converted into over 4 million high-resolution images, with each data-collection run representing about 1.5 TB of image data. We used the images to create a record containing the location of each detected pedestrian, cross-referenced to the relevant GEOID [26]. To detect pedestrians in the still images, our pipeline leverages the state-of-the-art convolutional neural network, Pedestron [85]. We used the cascade_hrnet architecture benchmarked on the CrowdHuman data set [190]. Our methods and dataset represent a first of its kind longitudinal collection and application of SVI data for research purposes.

As a secondary contribution, we provide a case study based on the video data collected throughout the COVID-19 pandemic. We examine the effect of vaccine availability and local demographics on pedestrian detections, while accounting for weekly and yearly seasonality. Community mobility became a key metric during the height of the COVID-19 pandemic as government officials worked to halt the spread of the virus [94, 186], and is also a common metric for disaster research [91]. Two of the largest and most widely used data sets for community mobility during this time were the Google Community Mobility Reports [2] and Apple Mobility Trends Reports [1]. Researchers used this data to study the incidence of COVID-19 in the US [167] and the effectiveness of government lockdown policies [98, 118], among other topics. Issues with these two data sets include mandatory opt-in, use of specific map applications, a lack of independent verification, and no long-term data availability guarantees [68, 98, 167, 222]. Our findings demonstrate the utility of our data processing pipeline as an alternative for tracking community mobility over time and show the potential for its use in a variety of research domains.

3.2 Methods

3.2.1 Data Collection

We collected our data as a part of the Seattle street-level imagery campaign, an ongoing series of video surveys for the purposes of documenting mobility throughout the COVID-19 pandemic. During each survey, a vehicle equipped with a 360° video camera is driven along a pre-defined route through Seattle while collecting video data and GPS metadata. The route incorporates broad neighborhood/area canvassing designed to collect data useful to multidisciplinary researchers as well as capital transects. Full details on the route design are available in Errett et al. [65]. The capital transects specifically target capitals (social, cultural, built, economic, and public health) which are theorized to be closely tied to community resilience [131]. Specific canvassing areas and capitals within Seattle were chosen to ensure a representative sample of the overall population of Seattle [65]. While the drivers try to make the surveys as consistent as possible, occasionally exogenous factors caused deviations from standard protocols. For example, during three of the surveys (05-29-2020, 06-18-2020, and 06-26-2020), protests over the murder of George Floyd caused parts of the survey route to be unnavigable.

After consulting with the University of Washington Human Subjects Division, it was determined that this study was not considered human subjects research and would not require IRB approval. The data we captured was people in public places, where they cannot expect personal privacy. As an added precaution, all data for this study was published through Mapillary [124], which automatically obscures faces.

3.2.2 Data Processing Pipeline

After video collection, the raw data is segmented into image data. The images are subsampled from video frames so that they are collected about every 4 meters. The images are then uploaded into the DesignSafe-CI Data Depot [176]. From DesignSafe, the images are transferred to the TACC Frontera high-performance computing cluster [201]. We completed all file transfers between the two services using Globus [38]. Without access to these services, or similar ones, the storage and computing requirements for this project would be intractable.

On Frontera, orthorectification is performed to the images, then pedestrian detection is performed on the orthorectified images. The orthorectification transforms the images from a single image in the equirectangular projection to two images in the rectilinear (gnomonic) projection [229]. Pedestrians are detected on each of the new images using a convolutional neural network (CNN) based on a pre-trained model from the Pedestron repository [85]. Our data represents a highly challenging detection task, as there is great variation in lighting, backgrounds, human poses, levels of occlusion and crowd density from image to image and run to run. The Cascade Mask R-CNN architecture in the Pedestron repository performed well on the CrowdHuman data set, representing a similar challenge to our data [190]. All testing and use of the CNN was performed using GPUs on the Frontera cluster. An example image after undergoing orthorectification and pedestrian detection is shown in Fig 3.1.

Figure 3.1: Sample images from the pedestrian detection data pipeline. The left image is an original 360° image from a data collection run. The image on the right is the right-hand side of the original image after orthorectification and pedestrian detection (both sides of the image are processed separately). There are two pedestrians that were detected by the algorithm (in red bounding boxes).

Using one GPU node on Frontera, with four NVIDIA Quadro RTX 5000 GPUs, the entire process takes about 3 seconds per original 360° image. Given the 4 million images we collected, this takes about 3,300 hours of computing time. While this is not a small number, when running in parallel, the whole process can be completed in a manner of days. In comparison, a human taking 10s per orthorectified image to count all the pedestrians would take over 22,000 hours to complete the same task. File compression/decompression for file transfer also takes a substantial amount of time. Since we used DesignSafe as our main data storage platform, we had to transfer files to/from the Frontera supercomputer to perform our pedestrian detection. To avoid overloading the file transfer system, we compressed the images from each run into a tar file prior to transferring the files to Frontera. This file compression/decompression can take several hours per run, but can be performed in parallel with the detection algorithm since they are on different systems. After compression, file transfer using Globus [38] takes minutes.

In post-processing, the pipeline filters out low-confidence detections (defined as any detection with less than 80% confidence) and associates the remaining high-confidence detections to U.S. Cenus Bureau GEOIDs [26]. We arrived at this confidence level after tuning for the precision and recall of the CNN classifier. Specifically, the pipeline filters based on the output of the second to last layer of the CNN, known as a *softmax layer*. For a $k-$class classification problem, the softmax layer will output a $k-$dimensional probability vector, where each i^{th} entry of the vector gives the probability that the original input to the CNN belongs to class i.

The final stage of post-processing is GEOID matching, where latitude and longitude metadata are cross-referenced to disjoint geographic regions (e.g. U.S. census tracts or block groups) and their respective GEOID codes. The cross-referencing code assumes the availability of shapefiles describing the geometry of the geographic regions. Aggregating the pedestrian detections according to U.S. Census Bureau GEOIDs [26] is necessary for analyses using sociodemographic data collected by the census. Additionally, the pedestrian detections can easily be cross-referenced with custom geometry defined using popular geographic information system software, such as the capitals data used in route construction and our analysis.

Following the GEOID matching step, the pedestrian detections data is written to a tabular format file (e.g. comma separated values). This file is an "analysis-ready" data product, in the sense that it is readable by most popular statistical analysis software (R, SPSS, Stata, etc.) and can be easily merged with other datasets using the GEOID column(s). A visual depiction of the entire pipeline is seen in Fig 3.2. Full code and a manual for following our process is available at `https://github.com/marte292/rapid-data-pipeline`.

3.2.3 Case Study: Community Mobility in Seattle during the COVID-19 Pandemic

Data Processing

All analysis is performed using the Python programming language version 3.11 [215]. The initial data product as outlined in the previous section is a list of detections, alongside the date of collection, geolocation, and GEOID. We also utilized a similar list of the images

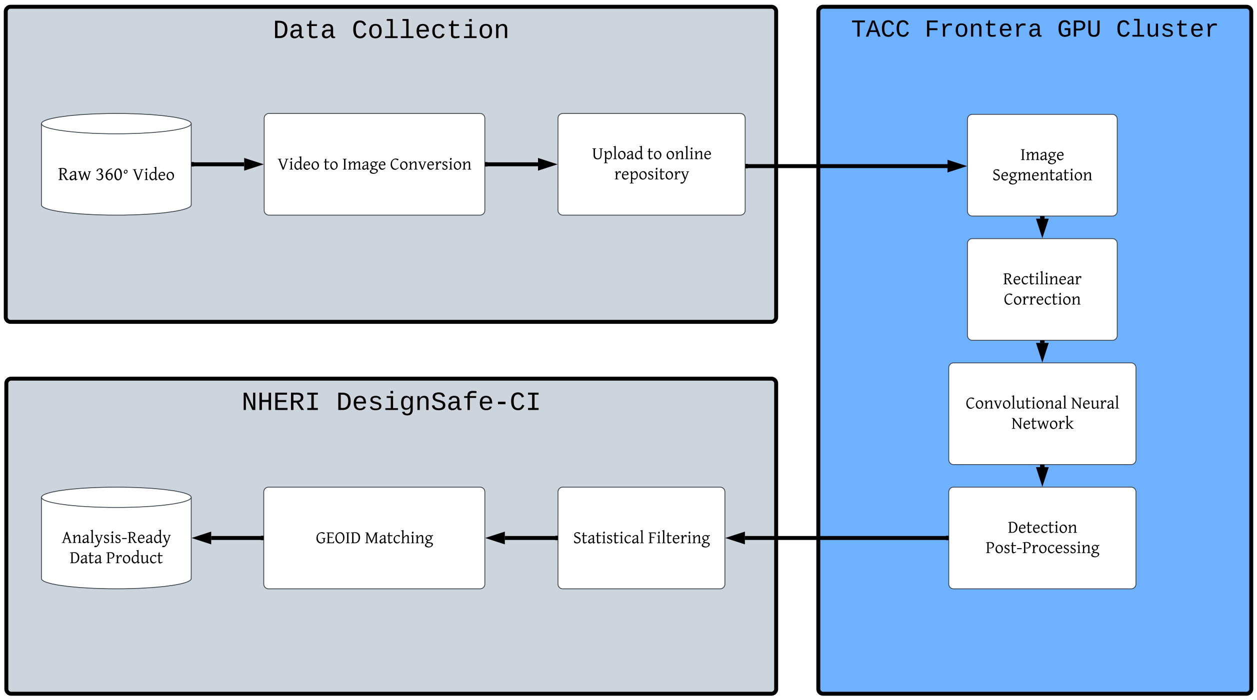


Figure 3.2: Flowchart of the data processing pipeline. The parts of the flowchart in gray occur on NHERI DesignSafe-CI, while the right-hand part in blue is done on the Frontera cluster.

themselves with the same features. The last dataset we utilized is the median household income data and racial demographic data from the 2019 American Community Survey (ACS) 5-year estimates. We aggregated the detections and image data for each data collection survey at the census tract level, then matched each census tract's total number of detections and images to its respective demographic and income data.

We utilized the data from 36 of the 37 surveys, omitting data from 10-29-2020. A heavy rain event caused the survey to be stopped early due to poor video quality. For each survey, we divided the number of detections in each census tract by the number of images collected in the tract to create a normalized 'detections per image' metric. This is a necessary step as the number of images in each tract may change survey to survey due to circumstances outside our control, such as construction or community events altering the route.

The last step in data processing was to transform some of our data to be represented by categorical variables. The date of each survey was coded both as either a weekend or weekday, and by the season. The date was also coded as either being before, or after the date that vaccines became publicly available. Income data was coded to be one of 5 levels that were used during route design. These brackets were $48,274 and below, $48,275 to $80,819, $80,820, to $110,536, $110,537 to $153,500, and $153,501 and above. Lastly, the proportion of the census tract's population that identifies as non-white was coded as an indicator variable, with '1' corresponding to areas that are 55.5% white or more. We determined this threshold using Jenk's natural breaks optimization. This left us with a dataset of 3171 observations to be used for analysis. Each observation represented a census tract with a detections per image value, as well as values for each of the categorical variables defined above.

Initial Regression Analysis

Using the processed data, we conducted a regression analysis to understand the relationships between predictors of interest and pedestrian traffic. Based on the known literature, we hypothesized that season, day of the week, COVID-19 vaccine availability, income level, and demographics all would have an impact on pedestrian traffic. We implemented a linear regression model to identify which of these factors are identified as statistically significant ($\alpha = .05$). We chose this modeling approach for its simple interpretability, as our modeling goal is to describe. The regression model is detailed below:

$$\begin{aligned} Y = \beta_0 + \beta_1 \times I_{vaccine} + \beta_{2..4} \times C_{season} + \beta_5 \times I_{weekend} \\ + \beta_{6..9} \times C_{incomelevel} + \beta_{10} \times I_{demographicindicator} + \epsilon, \end{aligned} \tag{3.1}$$

where Y is the detections per image for a date/census tract combination; $I_{vaccine}$ is an indicator for if the vaccine was available on that date; C_{season} is a categorical variable with 3 levels for summer, winter, and spring; $I_{weekend}$ is an indicator for if it is the weekend or not; $C_{incomelevel}$ is a categorical variable with 4 levels for the 4 income brackets above the lowest bracket; $I_{demographicindicator}$ is an indicator variable for if the population is 55.5% white or more. β_0 is the baseline detections per image on a weekday, not in the summer, with the vaccine unavailable, in a census tract at the lowest income level and a population less than 55.5% white. β_1 represents the change in detections per image from the vaccine becoming available, and $\beta_{2..4}$ represent the change for different seasons. β_5 represents the change from a weekday to the weekend, and $\beta_{6..9}$ represent the change to other income brackets. Lastly, β_{10} represents the change in detections per image to from an area that is less than 55.5% white to an area that is more.

In addition to the above analysis, we subset the data by only looking at detections that occurred in an image with at least one other detection. Then we calculated detections per image again, and fit the above model again with the new response variable. This same process was followed for detections with at least two, three, and four other detections in the same image. The goal of these analyses was to see if there were different trends for larger groups of people when compared with the entire data set.

3.3 Results

3.3.1 Data pipeline

Our main contribution, the open-source data pipeline, is publicly available on `https://github.com/marte292/rapid-data-pipeline`. The repository contains a process manual with step-by-step instructions on how to implement the data pipeline in Python [215]. The required Python libraries and system requirements are provided. Additionally, we provide enough code for future researchers to implement the pipeline on their own systems, with their own file structure. The pipeline is capable of processing terabytes of image data and outputting an analysis-ready data product in a matter of days (using high-performance computing, such as a single GPU node on Frontera, an academic supercomputer) with minimal human input.

3.3.2 *Case study*

Using data from the Seattle street-level imagery campaign, we calculated the number of detections per image across all data collection surveys. Fig 3.3 shows the detections per image for each survey, as well as the detections per image for the subset of detections sharing an image with at least 4 others. Fig 3.3 also displays the timestamp of COVID-19 vaccines becoming publicly available in Washington state.

Figure 3.3: Time series data of the total detections per image (solid blue line, left axis), and detections per image for the subset of detections sharing an image with at least 4 others (orange dashed line, right axis). As the survey dates are irregular, all dates are included in the figure. Please note that the axis for total detections per image does not start at 0. This was done purposefully to facilitate comparison between the trends of the two graphs.

Fig 3.3 depicts the trends over time for detections per image and detections sharing an image with at least 4 others. While both graphs exhibit similar trends overall, notably after vaccine rollout the graph of detections sharing an image with at least 4 others exceeds the graph of detections per image in all cases. The spike in detections seen in June 2020 is due to the large scale protests of police brutality that took place in Seattle in the aftermath of George Floyd's murder.

The full results of the linear regression model for total detections per image are displayed in Table 3.1. They show that the season being summer is the only significant seasonal effect. Additionally, the income bracket is a significant predictor, with wealthier areas seeing less pedestrian traffic. Finally, a census tract having a population greater than 55.5% white is a significant positive predictor. All other variables are not significant, including vaccine availability.

For the regression models using a subset of data, the results are similar to the initial model. All models have the same significant predictors as the initial model. The model using the detections sharing an image with at least one other also had the weekend as a borderline significant, negative predictor. The models using detections sharing an image with at least 3 and 4 others had vaccine availability as a significant, positive predictor. The full results of the linear regression model for detections per image with at least 4 others are displayed in Table 3.2, with all other regression models available in the supporting information.

3.4 Discussion

3.4.1 Comparison to Google Community Mobility Data

Given the ability to measure community mobility through pedestrian counts, there is potential value of our pipeline for social sciences and public health research [94, 186]. At an individual level, higher physical activity is known to predict better physical [8, 170] and mental health [171, 214, 234], and is associated with higher self-reported satisfaction and quality of life [14, 145]. In an aggregate sense, mobility is theorized to be an intermediate variable through which socioeconomic deprivation affects vulnerability to infectious disease [162, 233], resilience to disasters [92], and exposure to environmental hazards [110].

Dep. Variable:	Detections_per_Image	**R-squared:**	0.086
Model:	OLS	**Adj. R-squared:**	0.083
Method:	Least Squares	**F-statistic:**	29.71
No. Observations:	3171	**Prob (F-statistic):**	3.20e-55
Df Residuals:	3160	**Log-Likelihood:**	-4456.8
Df Model:	10		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	1.1242	0.075	14.914	0.000	0.976	1.272
Spring	0.0783	0.053	1.471	0.141	-0.026	0.183
Summer	0.2527	0.055	4.599	0.000	0.145	0.360
Winter	-0.0046	0.058	-0.079	0.937	-0.118	0.109
Vaccine Available	0.0061	0.036	0.172	0.863	-0.064	0.076
Weekend	-0.0775	0.052	-1.483	0.138	-0.180	0.025
Income Bracket 2	-0.4689	0.081	-5.795	0.000	-0.628	-0.310
Income Bracket 3	-0.8688	0.079	-11.041	0.000	-1.023	-0.714
Income Bracket 4	-0.9938	0.086	-11.540	0.000	-1.163	-0.825
Income Bracket 5	-1.3752	0.116	-11.893	0.000	-1.602	-1.148
More than 55.5% White	0.6416	0.054	11.820	0.000	0.535	0.748

Table 3.1: OLS Regression Results for Detections per Image. The first three non-intercept terms represent indicator variables for the different seasons, with fall being the baseline. The Vaccine Available term represents a binary variable for whether the COVID-19 initial vaccination series was publicly available or not. Weekend is a binary variable for whether the data was collected on Saturday or Sunday. The four Income Bracket terms are indicator variables for the median income level of the census tract where the data was collected. The income brackets are defined in our methods. Lastly, the More than 55.5% White term is an indicator variable for if the census tract in question had a populace that is more than 55.5% White. Full documentation for the Python package used to make this output is available from the developers [188].

Dep. Variable:	Five_Or_More_Peds_per_Image	**R-squared:**	0.059
Model:	OLS	**Adj. R-squared:**	0.056
Method:	Least Squares	**F-statistic:**	19.78
No. Observations:	3171	**Prob (F-statistic):**	7.09e-36
Df Residuals:	3160	**Log-Likelihood:**	4286.3
Df Model:	10		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0393	0.005	8.213	0.000	0.030	0.049
Spring	0.0006	0.003	0.180	0.857	-0.006	0.007
Summer	0.0100	0.003	2.870	0.004	0.003	0.017
Winter	-0.0039	0.004	-1.057	0.291	-0.011	0.003
Vaccine Available	0.0093	0.002	4.109	0.000	0.005	0.014
Weekend	0.0009	0.003	0.270	0.787	-0.006	0.007
Income Bracket 2	-0.0289	0.005	-5.626	0.000	-0.039	-0.019
Income Bracket 3	-0.0468	0.005	-9.373	0.000	-0.057	-0.037
Income Bracket 4	-0.0541	0.005	-9.889	0.000	-0.065	-0.043
Income Bracket 5	-0.0688	0.007	-9.376	0.000	-0.083	-0.054
More than 55.5% White	0.0309	0.003	8.960	0.000	0.024	0.038

Table 3.2: OLS Regression Results for Detections per Image for the detections subset sharing an image with at least 4 others. Coefficients are defined the same as in Table 3.1.

In light of this body of literature, we argue that the use of pedestrian counts to assess mobility could be a differentiating factor in researching social and health inequity. One extremely common source of mobility data during the COVID-19 Pandemic has been Google Community Mobility Reports [2] and Apple Mobility Trends Reports [1]. While there have been improvements in recent years [202], there are known representation and self-selection biases with existing mobility data captured by smartphones and other internet-based data collection methods [4, 99, 117, 136, 182].

Given the large number of publications using smartphone data as the foundation for their work, a natural question is how our data compares to smartphone mobility data. Comparison between our data set and the still publicly available Google Community Mobility Reports data can reveal some of the similarities and differences between the two data sets [2]. Google Community Mobility data is reported at the county level in the United States. Since Seattle is in King County, Washington, the King County data is what we use to draw the comparison.

Google Community Mobility data does not provide raw mobility numbers, but rather is reported as a percentage change from the five-week period of Jan 5–Feb 6, 2020. This data is collected from smartphones running the Android operating system with location history turned on, which is off by default. The data is baselined by day of the week, so data from a given Monday is compared to the median of the five Mondays in the baseline window to calculate a percent change. Additionally, it is unclear how exactly Google quantifies mobility. It is mentioned that it combines number of visitors to a location with amount of time spent in that location, but no specifics beyond that are provided.

Google mobility data is broken down into different categories. The category that most closely aligns with one of the categories used in our analysis is parks. Although Google's data classifies parks as official national parks and not the general outdoors, it does not indicate how it accounts for city or state parks. Our own data for park locations is based on the City of Seattle's official classifications.

Fig 3.4 shows a comparison of our detections per image data against Google Community Mobility data. Note that not all surveys are included because Google Community Mobility data stopped being provided on October 15, 2022. Overall, the trends between the two data sets are remarkably similar, lending further credibility to our data collection procedure. The more notable differences in the graph are from the months of November 2020 through August 2021, where the Google mobility data shows a larger drop followed by an increase in community mobility than was visible through our own data.

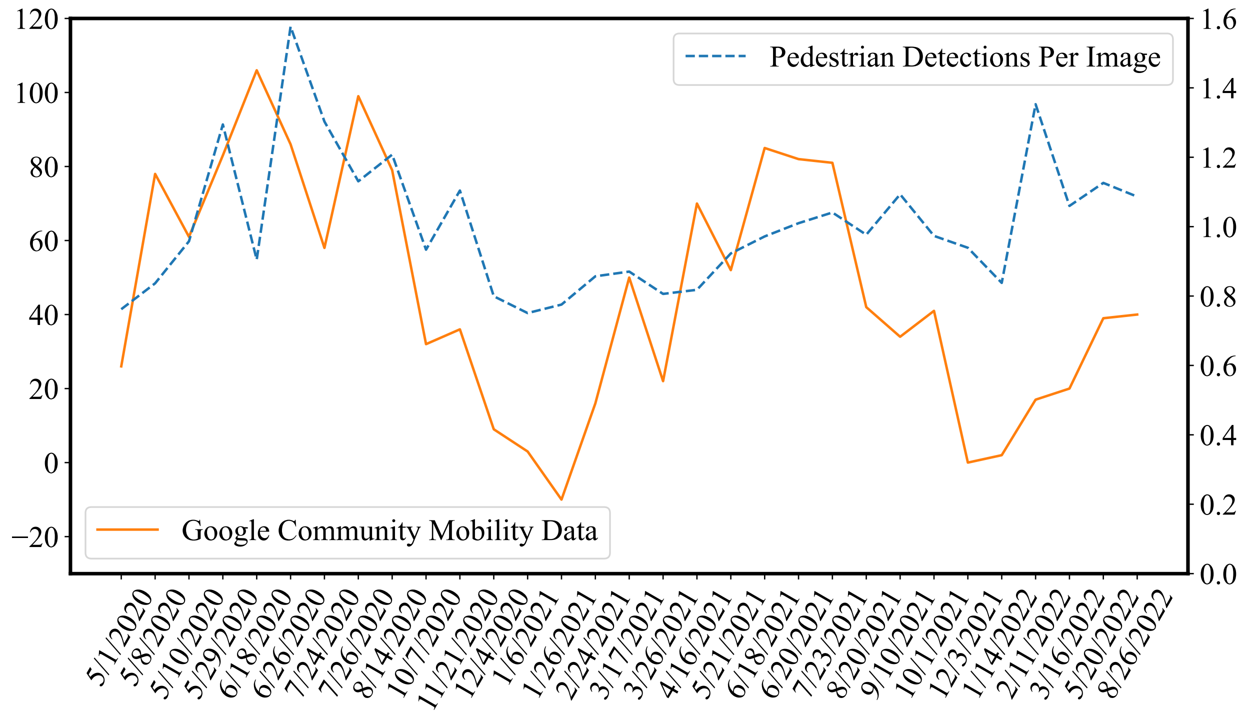


Figure 3.4: A depiction of our own detections per image data (blue, dashed; right axis) against Google Community Mobility data (orange, solid; left axis). The Pearson correlation between the two data sets is 0.387. The Google Community mobility data is aggregated at King County, WA, while our data covers a survey route within Seattle, which belongs to King County. As the dates of surveys were irregular (e.g., due to weather conditions), all dates are included in the figure.

One plausible explanation for this is the upwards sampling bias that occurs when using smartphone data [19, 211]. Our data set captures anyone on the street, including individuals experiencing homelessness, who are less likely to have smartphones. This population was on the streets throughout the entirety of the pandemic, so they were consistently captured by our data collection efforts. This consistent baseline pedestrian count could lead to a lesser response to vaccine rollout and winter weather in our own data in comparison with Google's. Additionally, there is a known income gap in both vaccination rates and smartphone ownership [12, 31]. This gap could drive the increase in the Google Mobility data during vaccine rollout.

3.4.2 Implications, Limitations, and Extensions

Our results show that it is possible for researchers to collect and analyze longitudinal SVI data. The presented methods can be used to collect and process SVI data from 8 hours worth of video in a manner of days. This time will only further decrease with faster data processing infrastructure and methods. These methods will allow novel longitudinal SVI data to be collected for research in a variety of application areas.

The results of the case study also bear further discussion. We demonstrated expected relationships between seasonal effects like day of week and weather on pedestrian traffic. Additionally, we showed that pedestrian traffic is inversely proportional to income, a known result during the COVID-19 pandemic, as lower income households are constrained in their capacity to work from home or take time off of work [62, 222]. Our results also showed that more white areas had higher on average pedestrian counts. This could be due to known trends, such as areas with larger non-white populations being more likely to stay home in response to government restrictions [195] and participate in other risk-reducing practices such as wearing a mask [87], or just due to local trends, as racial mobility trends tend to vary between cities [32]. These findings are consistent across all of our models, both looking at the entire data set, and the subsets examining pedestrians sharing an image. These results validate our method with respect to established literature, and provide a quantitative confirmation of results that had previously been found using cell phone data.

One new finding from our case study is that while overall pedestrian counts did not respond to vaccine availability, the subset of pedestrians who were in larger groups (4+ people in an image) did. Likely, the reason we did not see a response to the vaccine in the aggregate data is because our data only captures people who are outdoors. There is data that shows that outdoor pedestrian activity varied across cities, frequently increasing at recreation locations like trails, during the early days of the pandemic [55, 106]. Given these increases at some locations, a return to 'normal' pedestrian traffic may not mean an increase, but rather a change in traffic patterns. Our data captures this by showing that there was a significant increase in larger groups of people after the vaccine became available. This implies that people were more willing to be near each other outdoors after they had been vaccinated.

While the data pipeline presented here does represent a method for generating a novel data product, there are implementation challenges worth further discussion. For data collection, in addition to the time required to drive the route limiting the places of interest the route could reach, there were also many tradeoffs that had to be made when designing the route itself [65]. Despite having our survey route carefully designed to assess a representative sample of the Seattle population, some bias in route design is unavoidable. Since the route design included data from the American Community Survey aggregated at the census tract level, there is an implicit assumption of spatial homogeneity of the population within each census tract. Such bias is a manifestation of the well-known modifiable areal unit problem [69]. Since the majority of the route was primarily based on locations of interest throughout the city, this concern is somewhat mitigated.

In terms of processing, the pre-trained model we used required a substantial amount of high-performance computing time, and at times the data product generated was so large as to be unwieldy. Given the challenge our data set represents, using a model designed to be generalizeable is necessary to attain good detection results. As many state-of-the-art models perform substantially worse out of sample, we had to be careful to choose a model that was designed to perform well in this situation, at the cost of slower computing times [84]. Another unforeseen challenge was regular updates to the video camera's software to process and segment the video data into images. Consistent image formatting was vital for the data processing pipeline to function, so regular quality checks are necessary to make sure the images are processed properly.

The data product created, pedestrian detections, has some limitations as well. First, our method only captures pedestrians who are outdoors and near enough to the street to be captured via camera. This means that our data set does not include people who are indoors at these locations of interest, or who are too far from the street to be seen by camera. While the changes over time in pedestrian traffic we observed are still meaningful, it is important to recognize they don't capture everything. Similarly, our data cannot be interpreted as the actual number of pedestrians on the street. There is overlap in the image data, even when subset at 4 meter intervals and cropped during orthorectification. The orthorectified images only represent about 25% of the originals. However, this natural cropping is not enough to avoid the image overlap and further cropping would risk information loss. Pedestrians that appear in the foreground of one image may end up in the background of another. There are also several known instances of cyclists keeping relative pace with the street-view vehicle for several blocks, resulting in numerous detections. These issues are easy to circumvent in analysis by comparing the relative number of detections, although at the cost of interpretability.

Even with the above limitations, the data pipeline presented in this paper can be directly applied or adapted to be used in a number of contexts. Potential applications of longitudinal SVI data in assessing the built environment [198], broad urban research [16, 41, 111], and health research [183] have been well-documented, as the temporal instability of existing SVI data is discussed as a limitation in all of these fields. Beyond this, it is possible to estimate population demographics [70], and other neighborhood-level statistics [80,203] using SVI data. As our ability to quickly and accurately parse scenes using computer vision improves [54], potential application areas will only increase in number.

Another field where longitudinal SVI data could contribute a lot is disaster research. There is a substantial body of research dedicated to empirical methods for modeling various aspects of disaster recovery [126]. Our methods could be applied in this field to quantify recovery using pedestrian detections as a metric for community mobility, or another metric assessing the built environment as appropriate. Similar work has been done using repeat photography after Hurricane Katrina [27] but our methods represent a substantial increase in generated data, allowing for a wider range of analyses. Spatial video data collection for disaster reconnaissance has also been done [47], but involves manual assessment of the captured video. Our methods demonstrate that a fully-automated approach is possible, which would allow for more frequent data collection at a lower cost.

3.5 Conclusion

This chapter describes the creation of the first open-source SVI data pipeline for longitudinal analysis. Regression analysis based on the resulting longitudinal SVI data showed that pedestrian traffic patterns changed in response to the availability of the COVID-19 vaccine, thereby demonstrating the data pipeline's usefulness in research and practice. In particular, we showed that there were statistically significant increases in groups of people in proximity to each other after the vaccine became publicly available. Our data also captured expected trends in pedestrian traffic based on annual seasonality and socioeconomic factors. Our results demonstrate the feasibility and value in collecting SVI data as part of a longitudinal study. Longitudinal SVI data is capable of providing valuable insights in a variety of fields of study. Future work includes applications of our methods in broader public health research, disaster recovery research, and other fields of study that can benefit from longitudinal SVI data. Potential methodological directions include study-specific route design process improvements and newer pedestrian detection approaches, as further progress is made in this area.

Chapter 4

OUTDOOR SOCIAL DISTANCING BEHAVIORS CHANGED DURING A PANDEMIC

4.1 Introduction

In response to the COVID-19 pandemic, national and local governments declared emergencies and encouraged social distancing and mask wearing to reduce the spread of the virus [129]. Three months after the first confirmed case on U.S. soil in the state of Washington [90], the United States government declared in March 2020 that gatherings with more than 10 people should be avoided, and strongly encouraged people to work from home if possible. Additionally, the state of Washington closed schools, sit down restaurants, and entertainment facilities and issued a statewide stay at home order. The state also required face masks in public places starting in June 2020. The United States Centers for Disease Control and Prevention (CDC) recommended a physical distance of at least six feet both indoors and outdoors to prevent the spread of the virus. During the early days of the pandemic, some of the largest superspreading events in the US occurred when these recommendations were violated by large groups of people [49]. The CDC recommendations were not relaxed until August 2022, although most states officially 'reopened' well before then, including Washington State on June 30, 2021.

While it is known that it is easier for the disease to spread indoors when compared to outdoors, there are still many examples of outdoor transmission occurring, sometimes at a large scale [25, 172, 206]. Studies that have attempted to understand how outdoor transmission occurs and what factors reduce transmission risk showed that there is a strong risk of outdoor transmission under the right circumstances. Hang et al. [82] studied outdoor transmission in a street canyon environment and showed that outdoor transmission risk is highest in downwind areas. Fan et al. [66] likewise studied outdoor transmission in a street canyon environment using simulation methods. They recommended an outdoor distance of 4 meters when relative humidity and winds are low. Lastly, Kia et al. [232] studied outdoor transmission risk on the University of Houston campus, focusing on areas where air is not quickly ventilated. They showed that there were ‘hot spots’ on campus where air could take as long as 1,000 seconds to ventilate. Given the risk of outdoor transmission, it is useful to understand the public’s outdoor social distancing behaviors during the pandemic. Outdoor social distancing behaviors can serve as a proxy metric for recovery from COVID-19 and allow disaster researchers insight into key moments in the recovery timeline.

There are many studies that analyze community mobility and congregating behaviors using mobile phone data or self-reported survey data [73,81,129,140,194,222]. One finding from these studies is that pedestrians were less likely to strictly follow COVID-19 risk-reducing behaviors, including avoiding congregating after getting vaccinated [93, 194]. Another is that there are gaps in social distancing behaviors between different socioeconomic and racial groups [222]. While the findings of the above studies are extremely useful for understanding people’s congregating behaviors, there is a gap in understanding how people are adhering to interpersonal distance recommendations. Instead of measuring physical distances, these studies typically use foot traffic at specific locations of interest, and survey responses on social distancing behaviors as measures of adherence. Additionally, while efforts are made to avoid it, there are known representation issues associated with mobility data captured by cell phones [117, 182], and sampling bias associated with surveys. Lastly, these studies do not specifically track outdoor pedestrian behavior, but rather are focused on more general social distancing behaviors. This leaves a knowledge gap for understanding how the public practiced social distancing while outdoors.

In this study, the large-scale, longitudinal data set of street-view imagery captured in the city of Seattle, as part of the study described in Martell et al. [127], is used to track trends in social distancing over time. Physical distance between pedestrians is estimated using an algorithmic approach [184]. The end result of this process is a set of images, with the pedestrians identified, and the distances between them estimated, as seen in Fig 4.1, allowing for an empirical analysis of social distancing behaviors. The data set used in this study extends from May 2020 to July 2023, and can be used to track overall trends in community mobility in addition to social distancing using the estimated distances generated from the street view imagery. Our main contribution is a first of its kind method for empirically studying longitudinal outdoor social distancing behaviors. Given that different communities within the same geographic area can recovery differently from disaster events [7], the geospatial nature of our data, allowing for subsets of the data at specific locations of interest to be studied, is quite valuable. We can compare different geographic areas such as census tracts to understand the different drivers of social distancing behaviors. Secondary contributions include empirical confirmation of results from qualitative studies on social distancing behaviors, and additional perspective on COVID-19 related inequities in the US.

Figure 4.1: Sample output of the pedestrian detection and social distance estimation algorithm. There are 2 pedestrians, with an estimated 14.38 ft between them. Additional sample images are available in the supplementary material.

4.2 Methods

This section describes the methods used, including data collection, hypothesis generation, and statistical methods. The original data set in this work is featured in Martell et al. [127]. We do not describe the process for the generation of that data set here, but rather focus on the new features we added, most notably the physical distance estimates.

4.2.1 Data Description and Processing

The primary data set utilized in this study is the longitudinal street view image data set from Martell et al. [127]. The data set consists of 37 street-view surveys of the city of Seattle, beginning in May 2020 and ending in July 2023 (typically 2–4 weeks apart, as shown in Fig 4.2). There are over four million time-stamped, location-tagged images across the 37 surveys. 36 of the 37 data collection surveys were used in this study, as a heavy rain event caused a survey to be stopped early on 10-29-2020. The route design is featured in Errett et al. [65]. As described in Martell et al. [127], we created a data pipeline that took the 360-degree street view images and identified pedestrians in them using the Pedestron algorithm [85]. The output of this pipeline was a set of bounding boxes (each of which is a set of 4 coordinates surrounding a pedestrian, as seen in Fig 4.1) including additional metadata such as latitude, longitude, GEOID [26] of the census tract where the image was captured, various demographic data related to that census tract from the 2019 American Community Survey. Other data features include day of the week, and season of the year, both directly derived from the image timestamp. After consulting with the University of Washington Human Subjects Division, it was determined that this study was not considered human subjects research and would not require IRB approval. The data we captured was people in public places, where they cannot expect personal privacy. As an added precaution, all data for this study was published through Mapillary, which automatically obscures faces.

In addition to the data pipeline outputs, this study makes use of publicly available geolocation data on community capital transects from the city of Seattle. These capitals are

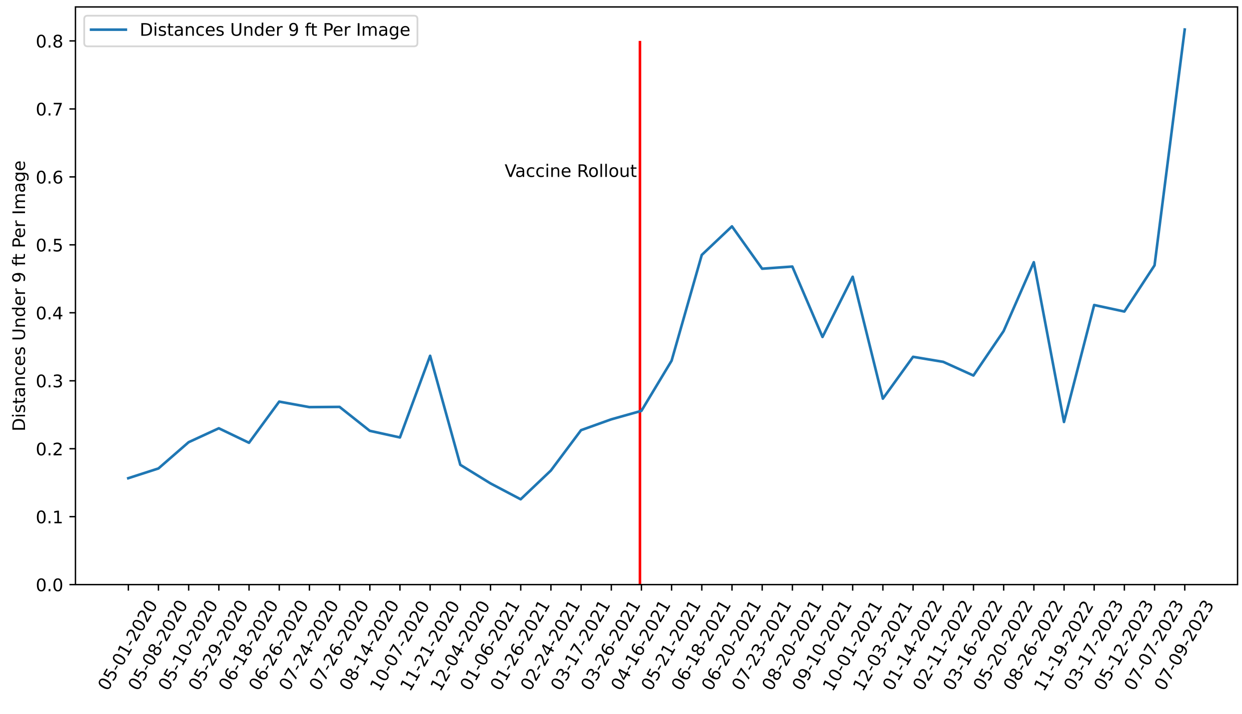


Figure 4.2: Summary data from the outdoor social distancing analysis. The blue line represents the number of distances under 9 ft per image. For example, 1.0 means that there is one pair of pedestrians on average per image, whose distance is under 9 ft.

theorized to be closely tied to community resilience [131] and were used in the design of the data collection survey route. The capitals examined in this study are parks, schools, faith-based organizations, museums, hospitals, medical clinics, and transit stops. Data for these capitals are publicly available through King County GIS Data Hub [105], Washington State Department of Health [220], the City of Seattle [42], and the Association of Religion Data Archives [180]. These locations were chosen to provide coverage over natural, cultural, and built capitals [63], and to provide a view into patterns at public health infrastructure.

We further processed the data by calculating distances between detected bounding boxes. We used the technique of Salazar [184] to obtain estimates of distances between bounding boxes, even in a 2D image. This technique is based on geometric properties and utilizes the

Pythagorean Theorem to estimate distances. Salazar achieved a Root Mean Square Error (RMSE) of 1.13 ft using this method. Applying this method to our data resulted in almost 5 million distance calculations. Lastly, we subset the distances to look at only those estimated by the algorithm to be under 9 ft in length. We decided on this subset as a conservative estimate, using the six feet apart guideline provided by the US CDC and the 1.13 ft RMSE of the distance calculation algorithm [184]. With this RMSE, we can be confident that anything measured as 9 ft or greater by the algorithm is at least 6 ft apart in reality.

After we calculated the distances, we paired the distances with various capitals based on geolocation. A distance is assumed to be at a given capital if it is within 200-ft of the shapefile footprint of that capital. The one exception to this is transit stops, which we only allowed for a 15-ft distance as pedestrians usually wait very near to, or inside of transit stops, and these structures (particularly transit stops) have much smaller footprints when compared to the other analyzed capitals. The last step in data preparation was to normalize the distances counts by the number of images captured in a given census tract. While the survey route was the same from run to run, the exact number of images captured at a given location may not be, so normalizing the number of distances per image for each survey run can help alleviate this inconsistency.

4.2.2 Exploratory Data Analysis and Initial Hypotheses

Before generating hypotheses about social distancing behaviors, we conducted an exploratory data analysis to better understand trends in the data set and formulate hypotheses about possible predictors of social distancing metrics. Fig 4.2 was a key input into our initial hypotheses. First, we noticed that the number of distances under 9 ft increases sharply in April 2021 and stays high over the rest of the surveys. This corresponds with when the initial COVID-19 vaccine became publicly available for all over the age of 16 in the state of Washington. While it is not strongly visible, we also expected there likely will be some seasonality present in the data, particularly an increase in traffic during the summer months. Lastly, while not easily visible in the figure, we expected there will be a relationship between the day of the week and community mobility. Beyond what we could glean from the graph, there are known inequities in how the COVID-19 pandemic affected different racial groups and people of different income levels [189]. To understand how these inequities might manifest themselves in the outdoor social distancing setting, we formulated the below hypotheses:

1. The number of distances under 9 ft per image will increase after the vaccine became publicly available.

2. The number of distances under 9 ft per image will be higher in the summer months.

3. The number of distances under 9 ft per image will be higher on the weekends.

4. Higher income areas will have a lower number of distances under 9 ft per image.

5. Areas with a larger population of people identifying as white will have a lower number of distances under 9 ft per image.

4.2.3 Regression Analyses

Based on the above hypotheses, we developed the following regression model to identify which factors are statistically significant ($\alpha = .05$):

$$\begin{aligned} Y = \beta_0 + \beta_1 \times I_{vaccine} + \beta_2 \times C_{season} + \beta_3 \times I_{weekend} \\ + \beta_4 \times C_{incomelevel} + \beta_5 \times I_{demographicindicator} + \epsilon, \end{aligned} \tag{4.1}$$

where Y is the number of distances under 9 ft per image for each date/census tract combination; $I_{vaccine}$ is an indicator for if the vaccine was available on that date; C_{season} is an indicator variable for if it is summer or not; $I_{weekend}$ is an indicator for if it is the weekend or not; $C_{incomelevel}$ is an indicator variable for if a given census tract has a median income above \$80,820; $I_{demographicindicator}$ is an indicator variable for if the population is 55.5% white or more. The \$80,820 and 55.5% breakpoints were chosen based on Jenk's natural breaks optimization. β_0 is the baseline number of distances per image on a weekday, not in the summer, with the vaccine unavailable, in a census tract with income below \$80,820 and a population less than 55.5% white. β_1 represents the change in the number of distances per image from the vaccine becoming available, and β_2 represents the change for it being the summer. β_3 represents the change from a weekday to the weekend, and β_4 represents the change to the higher income bracket. Lastly, β_5 represents the change from an area that is less than 55.5% white to an area that is more.

We utilized this model on the entire data set, as well as data subsets located at the various community capitals of interest. We hypothesized that different community capitals would have different significant predictors. As an example, traffic around schools is likely to decrease in the summers rather than increase like we expected to see in the data set at large. We also developed a similar model for the proportion of the distances under 9 ft. While this model does violate the linearity assumption of a linear regression model, the predicted outputs are all well within the 0–1 bounds of a proportion. We use a simple form of regression modeling because we are interested in inference rather than prediction. The only other change from the above model is that the regression coefficients are interpreted as changes in predicted proportion, rather than number of distances per image.

In addition to the above regression models, we examined the baseline trends in social distancing patterns at different capital locations. We did this by first calculating the overall proportion of distances under 9 ft for the entire data set and at each community capital. Then we compared the proportions at each capital to the overall proportion using a difference in proportions test. This allowed us to have a baseline understanding of the differences between the community capitals.

4.3 Results

4.3.1 Distances Under 9 ft Per Image

The regression analysis results for the number of distances under 9 ft per image across the entire data set are displayed in Table 4.1. Vaccine availability, whether it was the weekend or not, it being summer or not, and proportion of the population that identified as white were all significant, positive predictors. Income level was a significant, negative predictor. These results align with our hypotheses about the vaccine, summer, weekend, and income effects. The effect of having a higher proportion of people identifying as white was the opposite of our hypothesis. Results for individual capitals are summarized in Table 4.2, with full regression outputs in the supplemental material.

There were a few notable results from the regression analyses at community capitals. One immediately noticeable result from Table 4.2 was that vaccine availability was significant in all models, except at hospitals. In fact, hospitals did not have any significant predictors at all. Museums were the only capital to have a significant, positive income effect. Schools had a significant, negative effect from it being summer, and faith-based organizations had comparably sized vaccine and weekend effects. Lastly, when compared to the other capitals, museums had substantially larger coefficients for vaccine availability, the weekend effect, and the income effect.

4.3.2 Proportion of Distances under 9 ft

To obtain the proportion of distances less than 9 ft, we simply calculated the ratio between the number of distances under 9 ft, and the total number of distances. For the regression analysis, we subset the data at the survey and census tract levels first, then calculated the proportions. The proportion of distances less than 9 ft across the entire data set is 0.264. We then performed the same procedure for only distances at each specific capital and compared them to the total using a difference in proportions test. The results are displayed in Table 4.3

Dep. Variable:	Distances Under 9 ft Per Image	**R-squared:**	0.036
Model:	OLS	**Adj. R-squared:**	0.034
Method:	Least Squares	**F-statistic:**	23.53
No. Observations:	3171	**Prob (F-statistic):**	2.69e-23
Df Residuals:	3165	**Log-Likelihood:**	-3306.3
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.1753	0.032	5.405	0.000	0.112	0.239
Summer	0.0540	0.027	1.968	0.049	0.000	0.108
Vaccine Available	0.1686	0.025	6.824	0.000	0.120	0.217
Weekend	0.1000	0.036	2.807	0.005	0.030	0.170
Income Above $80,820	-0.1969	0.029	-6.797	0.000	-0.254	-0.140
More than 55.5% White	0.1925	0.032	5.937	0.000	0.129	0.256

Table 4.1: OLS Regression Results - Distances under 9 ft per Image. Please note that as there is some overlap that occurs between images, the coefficients here can only be interpreted relative to each other. Stating that the vaccine effect is 3 times as great as the summer effect in the same direction is useful, but the absolute interpretation of these coefficients' sizes is not. Full documentation for the Python package used to make this tabular output is available from the developers [188].

Capital	Intercept	Summer	Vaccine Available	Weekend	Income Above $80,820	More than 55% White
Transit Stops	0.0313***	0.0113	0.0262***	0.0045	-0.0261**	0.0077
Parks	0.0216	0.0007	0.0478***	0.0248	0.0043	0.0430**
Schools	0.0024	-0.0149*	0.0214**	-0.0126	0.0152	0.0093
Hospitals	-0.0002	0.0063	-0.0014	-0.0045	0.0048	0.0160
Medical Clinics	0.0940***	0.0029	0.0606**	0.0101	-0.1342***	0.0606**
Faith-Based Organizations	0.0270**	0.0189*	0.0286***	0.0257**	-0.0276***	0.0188*
Museums	-0.0249	0.0253	0.1094***	0.1075**	0.1915***	-0.0331

Table 4.2: OLS Regression Results - Distances under 9 ft per Image - Summary of Capitals Analysis. The numbers reported are the regression coefficients. Significance level: 0.05*, 0.01**, or <0.001***.

Capital	Proportion	p-value
Transit Stops	0.271	<.00001
Parks	0.256	<.00001
Schools	0.228	<.00001
Hospitals	0.279	<.00001
Medical Clinics	0.264	0.90242
Faith-Based Organizations	0.285	<.00001
Museums	0.261	.014509

Table 4.3: Difference-in-proportions test results for the proportion of distances under 9 ft *at each capital*, when compared to 0.264, the proportion of distances under 9 ft across the *entire* data set.

Dep. Variable:	Proportion of Distances Under 9 ft	**R-squared:**	0.047
Model:	OLS	**Adj. R-squared:**	0.046
Method:	Least Squares	**F-statistic:**	31.24
No. Observations:	3171	**Prob (F-statistic):**	3.81e-31
Df Residuals:	3165	**Log-Likelihood:**	848.19
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.3168	0.009	36.197	0.000	0.300	0.334
Summer	-0.0287	0.007	-3.882	0.000	-0.043	-0.014
Vaccine Available	0.0388	0.007	5.816	0.000	0.026	0.052
Weekend	0.0258	0.010	2.684	0.007	0.007	0.045
Income Above $80,820	0.0786	0.008	10.057	0.000	0.063	0.094
More than 55.5% White	-0.0120	0.009	-1.368	0.171	-0.029	0.005

Table 4.4: OLS Regression Results - Proportion of Distances Under 9 ft. The coefficients here can be directly interpreted as the relative change in the proportion of distances under 9 ft when a given predictor changes from 0 to 1.

The regression analysis results for the proportion of distances under 9 ft across the entire data set are displayed in Table 4.4. Vaccine availability, whether it was the weekend or not, and income level were all significant, positive predictors. It being summer was a significant, negative predictor. Results for individual capitals are summarized in Table 4.5, with full regression outputs in the supplemental material.

A notable trend from the community capitals regressions is how the weekend effect being significant in 4 different models, with a larger regression coefficient than the vaccine effect.

Capital	Intercept	Summer	Vaccine Available	Weekend	Income Above $80,820	More than 55% White
Transit Stops	0.2890***	-0.0355	0.0495**	0.0291	0.0449*	0.0199
Parks	0.3452***	-0.0105	0.0280*	0.0412*	0.0327*	-0.0309
Schools	0.4857***	-0.0312	-0.0144	0.0873*	0.0467	-0.0329
Hospitals	0.3317***	-0.0435	0.0713*	0.1386**	0.0313	-0.0927
Medical Clinics	0.2834***	-0.0358	0.0202	0.0636*	0.0343	0.0202
Faith-Based Organizations	0.3258***	-0.0171	0.0311*	0.0290	0.0436**	-0.0006
Museums	0.3155***	-0.0115	0.0528	0.0199	-0.0187	0.0197

Table 4.5: OLS Regression Results - Proportion of Distances under 9 ft - Summary of Capitals Analysis. The numbers reported are the regression coefficients. Significance level: 0.05*, 0.01**, or <0.001***.

Demographic variables played less of an effect when compared to the regression for number of distances under 9 ft per image. The income effect was only significant in 3 models, and the proportion of the population identifying as white was not significant in any of the models. The vaccine effect was only significant in 4 models, compared to 6 in the number of distances per image models. In all cases, the effect was still positive when significant.

4.4 Discussion

4.4.1 Implications

Our results represent important knowledge for researchers and policymakers. First, our approach represents a novel way of understanding outdoor social distancing behaviors. In particular, there is no other study to our knowledge that examines outdoor social distancing over time at this scale and level of detail. Our regression analysis shows that across the entire city of Seattle, vaccine availability drove pedestrian behavior. Across community capitals, there was an increase in pedestrians being within 9 ft of each other following the public availability of the vaccine in April 2021. Additionally, the overall proportion of distances that were 9 ft or less also increased. This suggests that, even outdoors, people were more likely to accept the risks of being near one another after the vaccine became available. These results also confirm findings from other studies that show that people were more likely to be willing to take risky behaviors such as visiting crowded places and being near others after being vaccinated [93, 194]. This is not the only additional risk people in the Seattle area are taking as the pandemic has continued, people have been more willing to embrace infection risk by not getting updated vaccines [104]. Fig 4.3 shows a stark contrast in the number of people who got the original COVID-19 vaccine series compared to the more recent bivalent boosters. In general, as time has gone on it appears that people are less concerned about COVID-19, possibly due to pandemic fatigue [185, 226].

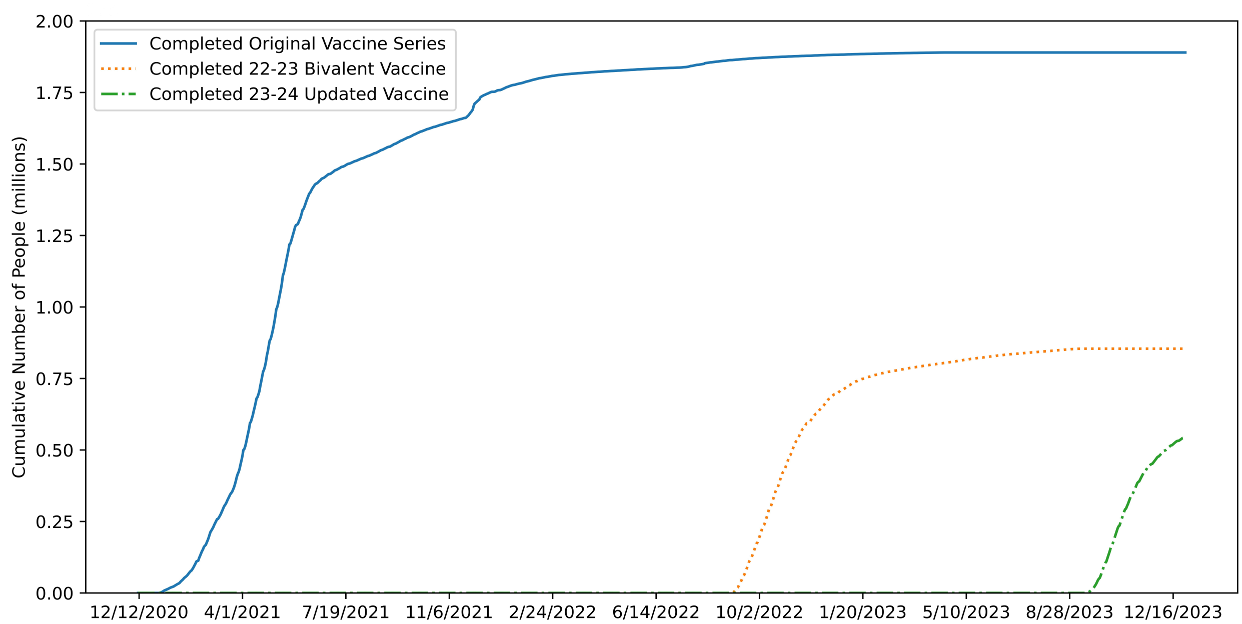


Figure 4.3: Cumulative vaccination totals over time in King County, Washington, where Seattle is the largest city, show diminishing willingness to get vaccines. Data collected from the King County Department of Public Health [104]

There are known inequities in how COVID-19 affected different communities, including higher mortality rates and infection rates for historically disadvantaged groups [222]. Across income groups, driving factors include the inability to work from home, challenges securing housing at all, and challenges securing quality healthcare for lower income groups [78]. Our analysis shows that lower income areas had more pedestrians traveling within 9 ft of each other, and more pedestrian traffic overall, leading to greater risk of exposure. In contrast, higher income areas did have a higher proportion of distances less than 9 ft.

While racial demographics were not shown to significantly predict the proportion of distances under 9 ft, census tracts that had a larger population identifying as white had more distances under 9 ft per image across the entire data set, and at parks, medical clinics, and faith-based organizations. This is a somewhat surprising result, as is the higher proportion of distances under 9 ft in higher income areas, given the known racial disparity in COVID-19 health outcomes in King County (see Fig 4.4) and racial and socioeconomic disparities in the US as a whole [189]. However, it is also known that communities of color are more likely to wear a mask, and that the drivers behind higher infection rates and death are systemic inequities in wealth, income, underlying health conditions, social capital, and lower quality health care [87, 196, 204]. In this light, it is less surprising that rates of social distancing adherence across areas with different racial demographics and income levels did not directly correlate with the divergent health outcomes in those groups.

Lastly, the findings from this chapter show some expected results with regard to weekend effects and seasonality. Across the data set, weekend surveys had higher numbers of distances under 9 ft, and the proportion of distances under 9 ft was also higher. These results held true at many, but not all, the capitals of interest. In the summer, the effect was mixed, with the total number of distances under 9 ft increasing across the entire data set, even though the proportion of distances under 9 ft decreased. This suggests that increased foot traffic and decreased social distancing adherence is not a given relationship when outdoors. In general, it is easier to maintain physical distance when outdoors when compared to being indoors, even if foot traffic is somewhat higher.

When comparing the social distancing results at different community capitals, two results stand out. First is the substantially lower proportion of distances less than 9 ft at schools.

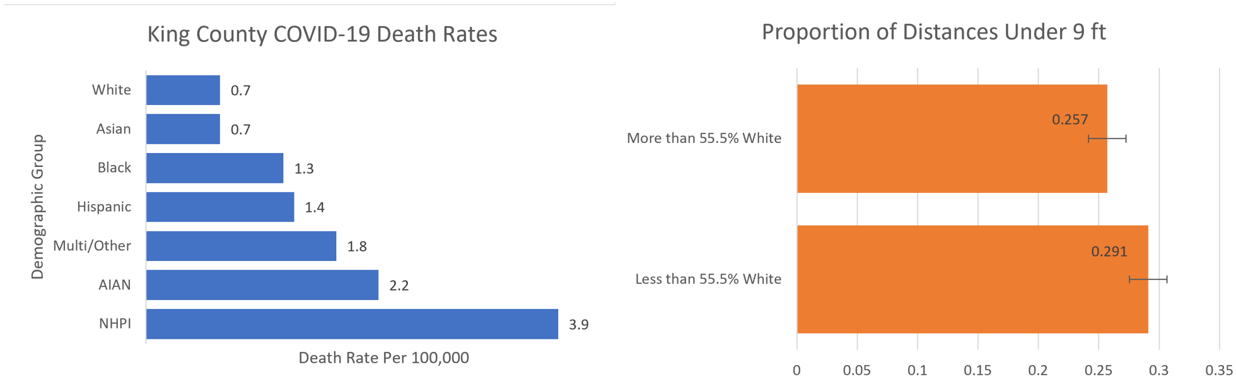


Figure 4.4: The left side graph shows the death rates per 100,000 citizens for different racial groups in King County, WA [103]. (NHPI, Native Hawaiian or Pacific Islander; AIAN, American Indian or Alaska Native). The right side graph shows the expected proportion of distances under 9 ft per image across our entire data set during the summer, in a lower-income census tract, on a weekday, before the vaccine became publicly available, based on our regression results. It is shown here that the more white areas did not have a statistically significant difference from the less white areas.

Schools have authority to enforce strict social distancing rules over their students. While indoors this could prove problematic for a variety of reasons [213], but outdoors, in our study area, these rules are much easier to enforce. Schools also experienced expected results in the regression analyses, with summer being a driver for a decrease in distances under 9 ft. In contrast, faith-based organizations were the location with the highest proportion of defections within 9 ft of each other. The relationship between religious groups and COVID-19 related restrictions in the U.S. can be described as tenuous at best [141], with frequent legal battles over restrictions. Additionally, there is a known link between highly religious Americans (particularly evangelicals) and less concern and support for policy recommendations from public health officials [187]. In this light, it is somewhat expected that social distancing recommendations would be less strictly followed at faith-based organizations compared to other community capitals. However, the regression results at faith-based organizations showed that both the number of distances under 9 ft and the proportion under 9 ft increased after the vaccine became available. Thus, there still was some level of risk avoiding behavior at faith-based organizations prior to vaccine availability. Faith-based organizations experienced an expected result of an increase in distances on weekends.

There were other notable results related to different community capitals from the regression analyses. First, hospitals did not have any significant predictors for the number of distances under 9 ft. Hospital traffic is largely driven by demand rather than other factors, which would lead to none of the seasonality effects being significant. In terms of socioeconomic predictors, hospitals are not limited to serving people who live in the immediate area, so the demographics of the census tract the institution is located would not be expected to impact foot traffic.The COVID-19 infection rate may be a better predictor of foot traffic at hospitals, although more research is needed to confirm this. In contrast, medical clinics had a vaccine effect, income effect, and racial demographic effect that were similar to the overall data set. This is likely due to the outdoor nature of our data. Medical clinics are much more common and cover more of the city than hospitals do, leading them to trend more with the overall data set. This is another capital where more research could be done to better understand social distancing behaviors. Lastly, the large income effect at museums is driven by the fact that the majority of museums in Seattle are located in wealthy areas, leading to a natural bias.

Our findings and methods are of interest to public health researchers and practitioners. In the short term, our results confirm some known findings related to COVID-19 inequities, and demonstrate that there were differences in pedestrian social distancing behaviors at specific locations. This provides a greater understanding of the impact of the COVID-19 pandemic, and can help inform the response for future pandemic events. Using this information would allow creation of targeted interventions for specific locations or groups of people. Additionally, future data collected during a pandemic event could be analyzed quickly, allowing for fast response to the rapidly changing landscape of a pandemic event.

4.4.2 Limitations

One challenge with utilizing computer vision is that the data product created, in this case the number of estimated distances, cannot be interpreted the actual number of distances at a given location. There is overlap in the image data in Martell et al. [127]. Pedestrians that appear in the foreground of one image may end up in the background of another. However, our results in this study are enough to demonstrate that while the raw number of distances may not be perfect, the relative change over time is meaningful. Additionally, this type of data collection only captures people near drive-able roads. As some locations of interest such as hospitals and parks may have large footprints away from roads, some data on pedestrian traffic near these areas will be lost that would potentially be captured by other methods such as using cell phone data. Our methods serve as a complement to data of this type, not a replacement.

The distance estimation algorithm is also not without problems [184]. The authors utilized an experiment to determine the optimal values for the algorithm's parameters through a grid search. In their tests, the ground truth pedestrians stood no more than 12 ft apart. The RMSE of 1.13 ft mentioned earlier in the paper is from those conditions. However, in our data set, pedestrians are frequently further apart than 12 ft, and further away from the vehicle than 20 ft. Thus, under those conditions, the algorithm's validity decreases, sometimes to the point where it overestimates distances by multiple orders of magnitude. However, given the nature of our study, this overestimation is acceptable. As we only care about pedestrians 9ft apart or closer, we are within the bounds of what the algorithm was trained on. When the overestimation occurs, the estimated distance obtains the same classification it would have otherwise, 9 ft or greater.

Another challenge with this method is the inability to differentiate between people from the same social bubbles, and strangers. During the height of the pandemic, households would form 'bubbles' of small non-overlapping groups that would be able to come into contact with each other while still maintaining social distancing benefits [22]. While our methods cannot account for these social bubbles, Danon et al. [22] shows that there are still substantial risks of transmission when this tactic is employed. Given this insight, we argue that this limitation does not substantially harm our findings.

Lastly, while the survey route was designed to provide a good representation of the city of Seattle, it is not without its flaws [65]. There are some key locations in the city that were missed, such as "The Ave" in the University District, that would have been valuable areas to study. Additionally, the route completely omitted West Seattle due to driving time constraints. These omissions do not directly harm the validity of the results in this paper, but they do somewhat limit their scope.

4.4.3 Extensions

These methods have the potential to be applied to future pandemic events to understand the impacts on social distancing behaviors and community mobility as a whole. If possible, conducting an occasional baseline survey would allow for pandemic-era data to be compared to pre-pandemic levels, something that was not done for this study that would allow for better understanding of a return to pre-disaster levels of distancing. Additionally, this type of analysis can be easily extended to other image data sets, whether it be indoors or outdoors, to understand changes in social distancing behaviors over time at key locations.

There are also still more insights that can be gleaned from this data set. For example, some capitals of interest that were not included in this study, such as cycling infrastructure [106,169], could be topics of further study. Additionally, with a larger data set, more predictor variables could be analyzed, such as more specific racial demographics than just white vs nonwhite, English proficiency, age and including interaction terms between predictors. A larger data set could also allow for block group-level analysis instead of tract-level analysis

Additionally, large-scale longitudinal street view imagery has a host of applications outside public health such as studying the built environment, and urban analytics as discussed in chapter 3 [111,183]. Beyond this, distance estimation algorithm in particular has many applications beyond pandemic events. The distance estimation algorithm we utilize here [184] is computationally efficient, and only requires the focal length of the camera, the vertical location of the horizon line in the image, the height of the camera, and estimation of the width of the object(s) of interest. Thus, as long as our objects of interest have fairly consistent real-world sizes (e.g. cars, bicycles, traffic signs), it is possible to analyze distances between them at a large scale. This allows analysis of topics of interest such as the bike-friendliness of a city [218] by understanding the average distance between vehicles and cyclists. In disaster recovery research, distance estimation could be used to understand recovery from catastrophic disaster events by observing less frequent, grouped pedestrian traffic (e.g. disaster response teams and construction work), change into more frequent, spread out pedestrian traffic (e.g. people going about their normal days) [88].

Finally, improvements in generalizable pedestrian detection algorithms or distance estimation in 2-dimensional images would allow for a higher accuracy in model outputs. Improvements in this area could allow for real-world interpretable outputs if pedestrian counts were more accurate. It would also allow for analysis of pedestrians that are further apart from each other, and more confidence in the model outputs for these cases. Similarly, it would become possible to study social distancing patterns for extremely short distances (e.g. less than 3ft) which is currently not possible given the RMSE of the distance estimation algorithm we used.

4.4.4 Conclusions

This chapter represents a first of its kind effort to track outdoor social distancing behaviors. We used a longitudinal street-view image survey, computer vision, and distance estimation techniques to generate over four million estimated distances across a 3-year survey period. We show that vaccine availability was a key driver in outdoor social distancing in the city of Seattle, with an increase in the number of distances under 9 ft after the vaccine became publicly available. Our results also highlight some of the systemic inequities that exist within the city that match broader trends in the US. This included that lower income areas experienced higher levels of pedestrians in proximity to each other, and that whiter areas had higher numbers of pedestrians in proximity to each other, in spite of white people still having better COVID-19 related health outcomes. Lastly, we were able to quantify the differences between community capitals with regard to social distancing behaviors. We observed that faith-based organizations had the lowest levels of social distancing adherence, while schools had the highest adherence levels. Overall, our results show that it is possible to understand one aspect of recovery from a pandemic event, social distancing, using SVI data.

Chapter 5

CONCLUSION

This dissertation advances the understanding of data usage in disaster recovery research. Given the increasing frequency of disaster events, it is critical to understand all the resources available to disaster researchers. With the ever-increasing amount of data being created, and new methods generating new types of data, it is more important now than ever to have a clear picture of how that data can be used to make our communities more resilient. The following chapters address these needs by focusing on a subset of disaster research that historically has been less understood, disaster recovery, building an understanding of past data practices, and demonstrating the potential of new data sources to advance the field.

Chapter 2 reviews the data usage history in lifeline infrastructure restoration modeling. We show that there are interrelationships between modeling approach, hazard type, lifeline system(s) modeled, and data usage. We also showed that data sets are frequently re-used as new methods become available. In spite of this, data sharing and publication are not common practices, as the one of the most common data sources in the literature was from direct collaboration with utilities. From our findings, we suggested directions for future research, including modeling interdependent systems and engagement through benchmarking testbeds as thematic directions, and using alternative data sources and improving validation and testing methods as methodological directions. We also provided a methodology for data management to improve reproducibility in disaster recovery modeling.

Chapter 3 builds on the findings of Chapter 2 by creating a method for collecting longitudinal street-view imagery. Over the course of a 3-year period, we collected over 4 million street-view images across 37 data collection surveys in the city of Seattle. Our methods included a data pipeline for processing the images and counting the pedestrians in them.

As a proof of concept for the method's value in disaster recovery research, we conducted a regression analysis on the pedestrian counts to understand community mobility in the city of Seattle during the COVID-19 pandemic. Our results confirmed the viability of our method for providing unique insights into pedestrian traffic patterns when compared to mobile phone data. We also discussed the potential of the method for many other uses than counting people.

Chapter 4 extends Chapter 3 by using the generated data to understand community recovery from the COVID-19 pandemic. We applied a distance estimation algorithm to the detection data to create an unprecedented longitudinal study of outdoor physical distancing behaviors. We demonstrated that outdoor physical distancing decreased after the public availability of the COVID-19 vaccine. This was shown both by the number and proportion of distances under 9 ft increasing after the vaccine became available. We also showed differences in physical distancing across census tract demographics, and locations of interest such as faith-based organizations and schools. Higher income areas had higher levels of physical distancing adherence, and there was not a significant difference in adherence between racial groups, highlighting the inequities of COVID-19 health outcomes.

We also identified multiple directions for future research to continue to improve our understanding of data in disaster recovery research. Chapter 2 identified promising new data sources including natural language processing of large scale text data, expert elicitation, and synthetic benchmarking testbeds. It also identified the importance of modeling interdependent restoration as a key thematic direction to improve model fidelity. Chapter 3 discussed the myriad of applications for longitudinal street-view data, including evaluation of the built environment, surveying plant species, and recovery from other types of disaster events. Additionally, methodological improvements could be made, including baseline data collection surveys between disaster events, improvements to route design, and improvements to computer vision algorithms. Chapter 4 future directions included the methodological improvements of Chapter 3, the extension of the provided analysis to other types of image data or locations of interest, applications in public policy, and large-scale distance estimation between other objects of interest.

Appendix A

APPENDIX

A.1 Additional Regression Results for Data Pipeline

Dep. Variable:	Two_Or_More_Peds_per_Image	**R-squared:**	0.114
Model:	OLS	**Adj. R-squared:**	0.112
Method:	Least Squares	**F-statistic:**	40.82
No. Observations:	3171	**Prob (F-statistic):**	1.97e-76
Df Residuals:	3160	**Log-Likelihood:**	438.90
Df Model:	10		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.2590	0.016	16.091	0.000	0.227	0.291
Spring	0.0122	0.011	1.073	0.283	-0.010	0.034
Summer	0.0619	0.012	5.278	0.000	0.039	0.085
Winter	-0.0044	0.012	-0.352	0.725	-0.029	0.020
Vaccine Available	-0.0143	0.008	-1.883	0.060	-0.029	0.001
Weekend	-0.0264	0.011	-2.366	0.018	-0.048	-0.005
Income Bracket 2	-0.0981	0.017	-5.675	0.000	-0.132	-0.064
Income Bracket 3	-0.2080	0.017	-12.377	0.000	-0.241	-0.175
Income Bracket 4	-0.2346	0.018	-12.758	0.000	-0.271	-0.199
Income Bracket 5	-0.3230	0.025	-13.081	0.000	-0.371	-0.275
More than 55.5% White	0.1564	0.012	13.494	0.000	0.134	0.179

Table A.1: OLS Regression Results for Detections per Image for the detections subset sharing an image with at least 1 other.

Dep. Variable:	Three_Or_More_Peds_per_Image	**R-squared:**	0.098
Model:	OLS	**Adj. R-squared:**	0.095
Method:	Least Squares	**F-statistic:**	34.23
No. Observations:	3171	**Prob (F-statistic):**	6.49e-64
Df Residuals:	3160	**Log-Likelihood:**	1864.6
Df Model:	10		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.1292	0.010	12.581	0.000	0.109	0.149
Spring	0.0026	0.007	0.356	0.722	-0.012	0.017
Summer	0.0309	0.007	4.128	0.000	0.016	0.046
Winter	-0.0053	0.008	-0.671	0.502	-0.021	0.010
Vaccine Available	0.0047	0.005	0.975	0.330	-0.005	0.014
Weekend	-0.0110	0.007	-1.544	0.123	-0.025	0.003
Income Bracket 2	-0.0657	0.011	-5.961	0.000	-0.087	-0.044
Income Bracket 3	-0.1289	0.011	-12.025	0.000	-0.150	-0.108
Income Bracket 4	-0.1472	0.012	-12.544	0.000	-0.170	-0.124
Income Bracket 5	-0.1926	0.016	-12.225	0.000	-0.223	-0.162
More than 55.5% White	0.0910	0.007	12.308	0.000	0.077	0.106

Table A.2: OLS Regression Results for Detections per Image for the detections subset sharing an image with at least 2 others.

Dep. Variable:	Four_Or_More_Peds_per_Image	**R-squared:**	0.077
Model:	OLS	**Adj. R-squared:**	0.074
Method:	Least Squares	**F-statistic:**	26.38
No. Observations:	3171	**Prob (F-statistic):**	9.11e-49
Df Residuals:	3160	**Log-Likelihood:**	3100.6
Df Model:	10		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0709	0.007	10.198	0.000	0.057	0.085
Spring	0.0009	0.005	0.186	0.853	-0.009	0.011
Summer	0.0197	0.005	3.878	0.000	0.010	0.030
Winter	-0.0052	0.005	-0.972	0.331	-0.016	0.005
Vaccine Available	0.0073	0.003	2.224	0.026	0.001	0.014
Weekend	-0.0041	0.005	-0.842	0.400	-0.014	0.005
Income Bracket 2	-0.0432	0.007	-5.787	0.000	-0.058	-0.029
Income Bracket 3	-0.0781	0.007	-10.764	0.000	-0.092	-0.064
Income Bracket 4	-0.0889	0.008	-11.186	0.000	-0.104	-0.073
Income Bracket 5	-0.1144	0.011	-10.721	0.000	-0.135	-0.093
More than 55.5% White	0.0527	0.005	10.522	0.000	0.043	0.063

Table A.3: OLS Regression Results for Detections per Image for the detections subset sharing an image with at least 3 others.

A.2 Additional Sample Image Outputs for Social Distancing Algorithm

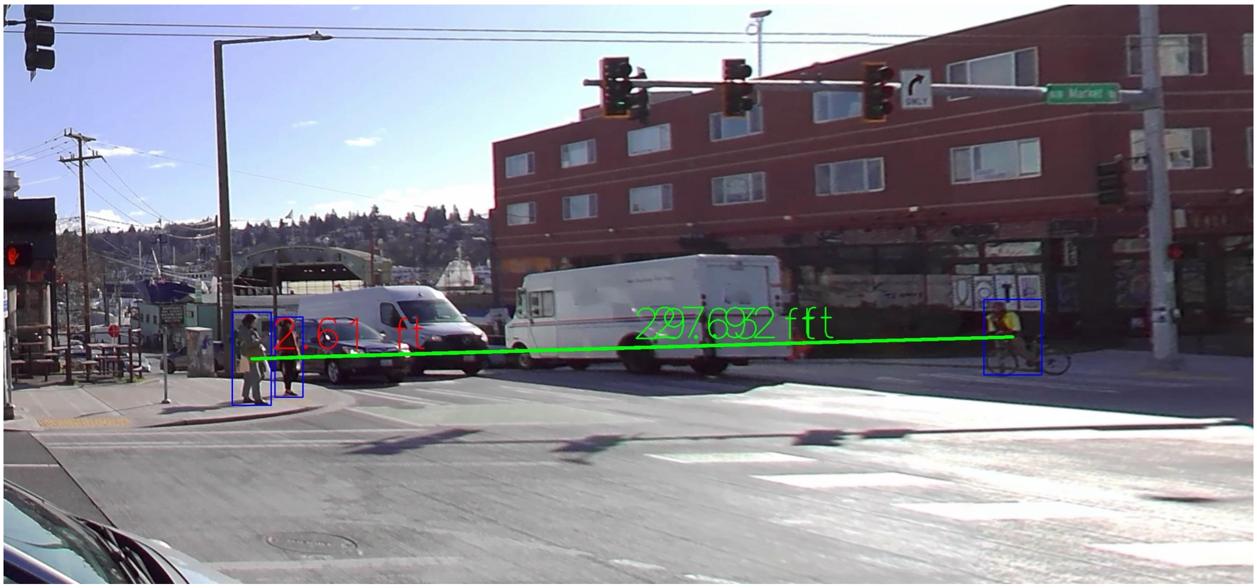


Figure A.1: Sample output of the pedestrian detection and physical distancing algorithm. This image contains two pedestrians near each other on one side of the street, about 30 feet away from a pedestrian across the street.

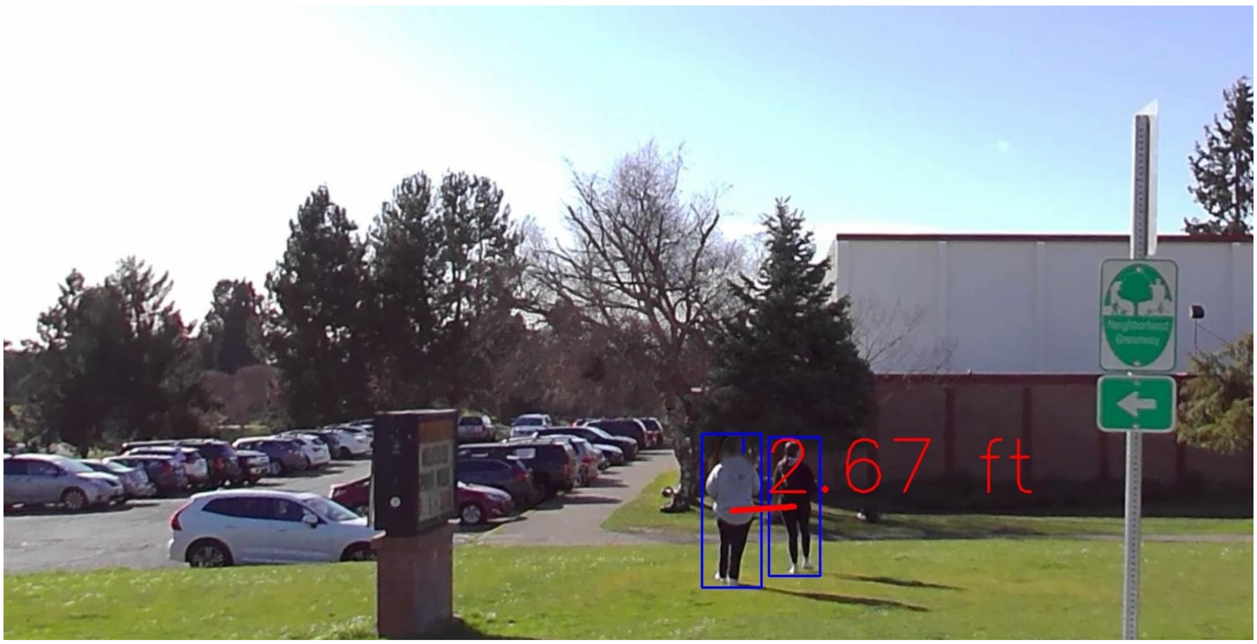


Figure A.2: Sample output of the pedestrian detection and physical distancing algorithm. This image contains two pedestrians walking nearby each other with their backs to the camera.

Figure A.3: Sample output of the pedestrian detection and physical distancing algorithm. This image contains two pedestrian walking nearby each other walking parallel with the vehicle.

A.3 Capitals Regression Results - Number of Distances Under 9 ft per Image

Dep. Variable:	Short_per_Image	**R-squared:**	0.021
Model:	OLS	**Adj. R-squared:**	0.017
Method:	Least Squares	**F-statistic:**	5.661
No. Observations:	1348	**Prob (F-statistic):**	3.57e-05
Df Residuals:	1342	**Log-Likelihood:**	858.81
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0313	0.009	3.677	0.000	0.015	0.048
Summer	0.0113	0.008	1.473	0.141	-0.004	0.026
Vaccine Available	0.0262	0.007	3.692	0.000	0.012	0.040
Weekend	0.0045	0.011	0.429	0.668	-0.016	0.025
Income Above $80,820	-0.0261	0.008	-3.248	0.001	-0.042	-0.010
More than 55.5% White	0.0077	0.009	0.860	0.390	-0.010	0.025

Table A.4: OLS Regression Results - Transit Stops - Number of Distances Under 9 ft

Dep. Variable:	Short_per_Image	**R-squared:**	0.031
Model:	OLS	**Adj. R-squared:**	0.027
Method:	Least Squares	**F-statistic:**	9.634
No. Observations:	1532	**Prob (F-statistic):**	4.55e-09
Df Residuals:	1526	**Log-Likelihood:**	953.40
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0270	0.008	3.350	0.001	0.011	0.043
Summer	0.0189	0.007	2.533	0.011	0.004	0.034
Vaccine Available	0.0286	0.007	4.234	0.000	0.015	0.042
Weekend	0.0257	0.010	2.625	0.009	0.007	0.045
Income Above $80,820	-0.0276	0.007	-3.730	0.000	-0.042	-0.013
More than 55.5% White	0.0188	0.008	2.399	0.017	0.003	0.034

Table A.5: OLS Regression Results - Faith-Based Organizations - Number of Distances Under 9 ft

Dep. Variable:	Short_per_Image	**R-squared:**	0.028
Model:	OLS	**Adj. R-squared:**	0.010
Method:	Least Squares	**F-statistic:**	1.545
No. Observations:	275	**Prob (F-statistic):**	0.176
Df Residuals:	269	**Log-Likelihood:**	562.77
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0002	0.009	-0.019	0.985	-0.019	0.019
Summer	0.0063	0.004	1.492	0.137	-0.002	0.015
Vaccine Available	-0.0014	0.004	-0.360	0.719	-0.009	0.006
Weekend	-0.0045	0.006	-0.792	0.429	-0.016	0.007
Income Above $80,820	0.0048	0.004	1.223	0.223	-0.003	0.013
More than 55.5% White	0.0160	0.010	1.661	0.098	-0.003	0.035

Table A.6: OLS Regression Results - Hospitals - Number of Distances Under 9 ft

Dep. Variable:	Short_per_Image	**R-squared:**	0.072
Model:	OLS	**Adj. R-squared:**	0.066
Method:	Least Squares	**F-statistic:**	12.42
No. Observations:	809	**Prob (F-statistic):**	1.25e-11
Df Residuals:	803	**Log-Likelihood:**	-19.400
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0940	0.020	4.612	0.000	0.054	0.134
Summer	0.0029	0.020	0.149	0.882	-0.035	0.041
Vaccine Available	0.0606	0.018	3.431	0.001	0.026	0.095
Weekend	0.0101	0.025	0.395	0.693	-0.040	0.060
Income Above $80,820	-0.1342	0.019	-7.050	0.000	-0.172	-0.097
More than 55.5% White	0.0606	0.021	2.845	0.005	0.019	0.102

Table A.7: OLS Regression Results - Medical Clinics - Number of Distances Under 9 ft

Dep. Variable:	Short_per_Image	**R-squared:**	0.190
Model:	OLS	**Adj. R-squared:**	0.176
Method:	Least Squares	**F-statistic:**	13.52
No. Observations:	294	**Prob (F-statistic):**	7.46e-12
Df Residuals:	288	**Log-Likelihood:**	32.332
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0249	0.025	-0.983	0.326	-0.075	0.025
Summer	0.0253	0.029	0.880	0.380	-0.031	0.082
Vaccine Available	0.1094	0.026	4.231	0.000	0.058	0.160
Weekend	0.1075	0.038	2.817	0.005	0.032	0.183
Income Above $80,820	0.1915	0.035	5.448	0.000	0.122	0.261
More than 55.5% White	-0.0331	0.034	-0.967	0.334	-0.100	0.034

Table A.8: OLS Regression Results - Museums - Number of Distances Under 9 ft

Dep. Variable:	Short_per_Image	**R-squared:**	0.026
Model:	OLS	**Adj. R-squared:**	0.023
Method:	Least Squares	**F-statistic:**	9.512
No. Observations:	1783	**Prob (F-statistic):**	5.75e-09
Df Residuals:	1777	**Log-Likelihood:**	434.16
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.0216	0.011	1.915	0.056	-0.001	0.044
Summer	0.0007	0.010	0.065	0.948	-0.019	0.021
Vaccine Available	0.0478	0.009	5.238	0.000	0.030	0.066
Weekend	0.0248	0.013	1.873	0.061	-0.001	0.051
Income Above $80,820	0.0043	0.011	0.376	0.707	-0.018	0.026
More than 55.5% White	0.0430	0.013	3.415	0.001	0.018	0.068

Table A.9: OLS Regression Results - Parks - Number of Distances Under 9 ft

Dep. Variable:	Short_per_Image	**R-squared:**	0.031
Model:	OLS	**Adj. R-squared:**	0.026
Method:	Least Squares	**F-statistic:**	5.611
No. Observations:	871	**Prob (F-statistic):**	4.24e-05
Df Residuals:	865	**Log-Likelihood:**	855.59
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0024	0.009	0.281	0.779	-0.014	0.019
Summer	-0.0149	0.007	-2.200	0.028	-0.028	-0.002
Vaccine Available	0.0214	0.006	3.428	0.001	0.009	0.034
Weekend	-0.0126	0.009	-1.343	0.180	-0.031	0.006
Income Above $80,820	0.0152	0.010	1.450	0.148	-0.005	0.036
More than 55.5% White	0.0093	0.010	0.923	0.356	-0.010	0.029

Table A.10: OLS Regression Results - Schools - Number of Distances Under 9 ft

A.4 Capitals Regression Results - Proportion of Distances Under 9 ft

Dep. Variable:	Prop_Short	**R-squared:**	0.014
Model:	OLS	**Adj. R-squared:**	0.010
Method:	Least Squares	**F-statistic:**	3.682
No. Observations:	1348	**Prob (F-statistic):**	0.00259
Df Residuals:	1342	**Log-Likelihood:**	-427.14
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.2890	0.022	13.081	0.000	0.246	0.332
Summer	-0.0355	0.020	-1.776	0.076	-0.075	0.004
Vaccine Available	0.0495	0.018	2.688	0.007	0.013	0.086
Weekend	0.0291	0.027	1.064	0.288	-0.025	0.083
Income Above $80,820	0.0449	0.021	2.150	0.032	0.004	0.086
More than 55.5% White	0.0199	0.023	0.860	0.390	-0.025	0.065

Table A.11: OLS Regression Results - Transit Stops - Proportion of Distances Under 9 ft

Dep. Variable:	Prop_Short	**R-squared:**	0.009
Model:	OLS	**Adj. R-squared:**	0.006
Method:	Least Squares	**F-statistic:**	2.854
No. Observations:	1532	**Prob (F-statistic):**	0.0143
Df Residuals:	1526	**Log-Likelihood:**	-235.43
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.3258	0.018	18.615	0.000	0.291	0.360
Summer	-0.0171	0.016	-1.057	0.291	-0.049	0.015
Vaccine Available	0.0311	0.015	2.117	0.034	0.002	0.060
Weekend	0.0290	0.021	1.359	0.174	-0.013	0.071
Income Above $80,820	0.0436	0.016	2.708	0.007	0.012	0.075
More than 55.5% White	-0.0006	0.017	-0.034	0.973	-0.034	0.033

Table A.12: OLS Regression Results - Faith-Based Organizations - Proportion of Distances Under 9 ft

Dep. Variable:	Prop_Short	**R-squared:**	0.052
Model:	OLS	**Adj. R-squared:**	0.034
Method:	Least Squares	**F-statistic:**	2.928
No. Observations:	275	**Prob (F-statistic):**	0.0136
Df Residuals:	269	**Log-Likelihood:**	-21.084
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.3317	0.079	4.183	0.000	0.176	0.488
Summer	-0.0435	0.035	-1.228	0.221	-0.113	0.026
Vaccine Available	0.0713	0.032	2.214	0.028	0.008	0.135
Weekend	0.1386	0.048	2.915	0.004	0.045	0.232
Income Above $80,820	0.0313	0.033	0.948	0.344	-0.034	0.096
More than 55.5% White	-0.0927	0.080	-1.153	0.250	-0.251	0.066

Table A.13: OLS Regression Results - Hospitals - Proportion of Distances Under 9 ft

Dep. Variable:	Prop_Short	**R-squared:**	0.017
Model:	OLS	**Adj. R-squared:**	0.011
Method:	Least Squares	**F-statistic:**	2.811
No. Observations:	809	**Prob (F-statistic):**	0.0159
Df Residuals:	803	**Log-Likelihood:**	-76.898
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.2834	0.022	12.953	0.000	0.240	0.326
Summer	-0.0358	0.021	-1.709	0.088	-0.077	0.005
Vaccine Available	0.0202	0.019	1.064	0.288	-0.017	0.057
Weekend	0.0636	0.027	2.326	0.020	0.010	0.117
Income Above $80,820	0.0343	0.020	1.678	0.094	-0.006	0.074
More than 55.5% White	0.0202	0.023	0.883	0.377	-0.025	0.065

Table A.14: OLS Regression Results - Medical Clinics - Proportion of Distances Under 9 ft

Dep. Variable:	Prop_Short	**R-squared:**	0.012
Model:	OLS	**Adj. R-squared:**	-0.005
Method:	Least Squares	**F-statistic:**	0.6991
No. Observations:	294	**Prob (F-statistic):**	0.625
Df Residuals:	288	**Log-Likelihood:**	-10.138
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.3155	0.029	10.770	0.000	0.258	0.373
Summer	-0.0115	0.033	-0.347	0.729	-0.077	0.054
Vaccine Available	0.0528	0.030	1.769	0.078	-0.006	0.112
Weekend	0.0199	0.044	0.452	0.651	-0.067	0.107
Income Above $80,820	-0.0187	0.041	-0.461	0.645	-0.099	0.061
More than 55.5% White	0.0197	0.040	0.498	0.619	-0.058	0.097

Table A.15: OLS Regression Results - Museums - Proportion of Distances Under 9 ft

Dep. Variable:	Prop_Short	**R-squared:**	0.008
Model:	OLS	**Adj. R-squared:**	0.005
Method:	Least Squares	**F-statistic:**	2.878
No. Observations:	1786	**Prob (F-statistic):**	0.0136
Df Residuals:	1780	**Log-Likelihood:**	-123.92
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.3454	0.015	22.391	0.000	0.315	0.376
Summer	-0.0115	0.014	-0.830	0.407	-0.039	0.016
Vaccine Available	0.0285	0.012	2.289	0.022	0.004	0.053
Weekend	0.0408	0.018	2.255	0.024	0.005	0.076
Income Above $80,820	0.0321	0.015	2.077	0.038	0.002	0.062
More than 55.5% White	-0.0312	0.017	-1.809	0.071	-0.065	0.003

Table A.16: OLS Regression Results - Parks - Proportion of Distances Under 9 ft

Dep. Variable:	Prop_Short	**R-squared:**	0.010
Model:	OLS	**Adj. R-squared:**	0.004
Method:	Least Squares	**F-statistic:**	1.735
No. Observations:	874	**Prob (F-statistic):**	0.124
Df Residuals:	868	**Log-Likelihood:**	-327.78
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.4866	0.033	14.557	0.000	0.421	0.552
Summer	-0.0331	0.026	-1.259	0.209	-0.085	0.019
Vaccine Available	-0.0138	0.024	-0.572	0.568	-0.061	0.034
Weekend	0.0845	0.036	2.325	0.020	0.013	0.156
Income Above $80,820	0.0459	0.041	1.125	0.261	-0.034	0.126
More than 55.5% White	-0.0341	0.039	-0.876	0.381	-0.111	0.042

Table A.17: OLS Regression Results - Schools - Proportion of Distances Under 9 ft

A.5 Regression Results - All Distances Per Image

Dep. Variable:	Distances_per_Image	**R-squared:**	0.027
Model:	OLS	**Adj. R-squared:**	0.025
Method:	Least Squares	**F-statistic:**	17.39
No. Observations:	3171	**Prob (F-statistic):**	5.01e-17
Df Residuals:	3162	**Log-Likelihood:**	-7893.7
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.6767	0.138	4.910	0.000	0.406	0.947
Summer	0.2963	0.117	2.541	0.011	0.068	0.525
Vaccine Available	0.5625	0.105	5.360	0.000	0.357	0.768
Weekend	0.3271	0.151	2.161	0.031	0.030	0.624
Income Above $80,820	-0.7484	0.123	-6.081	0.000	-0.990	-0.507
More than 55.5% White	0.6949	0.138	5.044	0.000	0.425	0.965

Table A.18: OLS Regression Results - Distances Per Image. This table contains regression results for a regression containing ALL distances, not just those under 9 ft.

BIBLIOGRAPHY

[1] Apple covid-19 mobility trends reports, 2020.

[2] Google covid-19 community mobility reports, 2020.

[3] Seattle's activist-occupied zone is just the latest in a long history of movements and protests; the community-controlled capitol hill organized protest, or 'chop' has drawn nationwide attention, but the city's radical history goes back as far as 1919. *The Guardian (London)*, 2020.

[4] Alberto Aleta, David Martín-Corral, Michiel A. Bakker, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E. Dean, M. Elizabeth Halloran, Ira M. Longini, Alex Pentland, Alessandro Vespignani, Yamir Moreno, and Esteban Moro. Quantifying the importance and location of sars-cov-2 transmission events in large metropolitan areas. *Proceedings of the National Academy of Sciences*, 119(26):e2112182119, 2022.

[5] Applied Technology Council. *Critical Assessment of Lifeline System Performance: Understanding Societal Needs in Disaster Recovery.* National Institute of Standards and Technology (NIST), 2016.

[6] Hannah M. Badland, Simon Opit, Karen Witten, Robin A. Kearns, and Suzanne Mavoa. Can Virtual Streetscape Audits Reliably Replace Physical Streetscape Audits? *Journal of Urban Health*, 87(6):1007–1016, December 2010.

[7] Homa Bahmani and Wei Zhang. Why Do Communities Recover Differently after Socio-Natural Disasters? Pathways to Comprehensive Success of Recovery Projects Based on Bam's (Iran) Neighborhoods' Perspective. *International journal of environmental research and public health*, 19(2):678, 2022. Place: Switzerland Publisher: MDPI AG.

[8] Patricia S Baker, Eric V Bodner, and Richard M Allman. Measuring life-space mobility in community-dwelling older adults. *Journal of the American Geriatrics Society*, 51(11):1610–1614, 2003.

[9] A Barabadi and Y.Z Ayele. Post-disaster infrastructure recovery: Prediction of recovery rate using historical data. *Reliability Engineering and System Safety*, 169:209–223, 2018.

[10] Kash Barker and Hiba Baroud. Proportional hazards models of infrastructure system recovery. *Reliability Engineering and System Safety*, 124:201–206, 2014.

[11] Kash Barker and Yacov Y Haimes. Assessing uncertainty in extreme events: Applications to risk-based decision making in interdependent infrastructure sectors. *Reliability Engineering and System Safety*, 94(4):819–829, 2009.

[12] Vaughn Barry, Sharoda Dasgupta, Daniel L Weller, Jennifer L Kriss, Betsy L Cadwell, Charles Rose, Cassandra Pingali, Trieste Musial, J. Danielle Sharpe, Stephen A Flores, Kurt J Greenlund, Anita Patel, Andrea Stewart, Judith R Qualters, LaTreace Harris, Kamil E Barbour, and Carla L Black. Patterns in covid-19 vaccination coverage, by social vulnerability and urbanicity — united states, december 14, 2020–may 1, 2021, 2021.

[13] Sean Bechhofer, David De Roure, Matthew Gamble, Carole Goble, and Iain Buchan. Research objects: Towards exchange and reuse of digital knowledge. *Nature precedings*, 2010.

[14] Cecilia Jakobsson Bergstad, Amelie Gamble, Tommy Gärling, Olle Hagman, Merritt Polk, Dick Ettema, Margareta Friman, and Lars E Olsson. Subjective well-being related to satisfaction with daily travel. *Transportation*, 38(1):1–15, 2011.

[15] Michel Bessani, Rodrigo Zempulski Fanucchi, Jorge Alberto Achcar, and Carlos Dias Maciel. A statistical analysis and modeling of repair data from a Brazilian power distribution system. In *2016 17th International Conference on Harmonics and Quality of Power (ICHQP)*, volume 2016-, pages 473–477. IEEE, 2016.

[16] Filip Biljecki and Koichi Ito. Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, 215:104217, November 2021.

[17] Adam B Birchfield, Kathleen M Gegner, Ti Xu, Komal S Shetye, and Thomas J Overbye. Statistical considerations in the creation of realistic synthetic power grids for geomagnetic disturbance studies. *IEEE Transactions on Power Systems*, 32(2):1502–1510, 2017.

[18] Adam B Birchfield, Ti Xu, Kathleen M Gegner, Komal S Shetye, and Thomas J Overbye. Grid structural characteristics as validation criteria for synthetic networks. *IEEE Transactions on Power Systems*, 32(4):3258–3265, 2017.

[19] Amit Birenboim and Noam Shoval. Mobility research in the age of the smartphone. *Annals of the American Association of Geographers*, 106(2):283–291, 2016.

[20] S. A. Brink, R. A. Davidson, and Taronne Tabucchi. Estimated durations of post-earthquake water service interruptions in Los Angeles. In *TCLEE 2009: Lifeline Earthquake Engineering in a Multihazard Environment*, volume 357, pages 1–12. American Society of Civil Engineers, 2009.

[21] Susan A. Brink, R. A. Davidson, and Taronne H.P Tabucchi. Strategies to reduce durations of post-earthquake water service interruptions in Los Angeles. *Structure and Infrastructure Engineering*, 8(2):199–210, 2012.

[22] Ellen Brooks-Pollock, Leon Danon, Thibaut Jombart, and Lorenzo Pellis. Modelling that shaped the early COVID-19 pandemic response in the UK. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1829):20210001, July 2021.

[23] R.E Brown, S Gupta, R.D Christie, S.S Venkata, and R Fletcher. Distribution system reliability assessment: momentary interruptions and storms. *IEEE Transactions on Power Delivery*, 12(4):1569–1575, 1997.

[24] Richard E Brown and Rick Pinkerton. Distribution reliability optimization using synthetic feeders. *Energies*, 12(18):3510, 2019.

[25] Tommaso Celeste Bulfone, Mohsen Malekinejad, George W Rutherford, and Nooshin Razani. Outdoor Transmission of SARS-CoV-2 and Other Respiratory Viruses: A Systematic Review. *The Journal of infectious diseases*, 223(4):550–561, 2021. Place: US Publisher: Oxford University Press.

[26] United States Census Bureau. Understanding Geographic Identifiers (GEOIDs), 2021. Section: Government.

[27] Christopher Burton, Jerry T. Mitchell, and Susan L. Cutter. Evaluating post-Katrina recovery in Mississippi using repeat photography. *Disasters*, 35(3):488–509, July 2011.

[28] Z. Çağnan and R. A. Davidson. Discrete event simulation of the post-earthquake restoration process for electric power systems. *Int. J. of Risk Assessment and Management*, 7(8), 2007.

[29] Z Çağnan, R. A. Davidson, and S. D. Guikema. Post-earthquake restoration planning for Los Angeles electric power. *Earthquake Spectra*, 22(3):589–608, 2006.

[30] Burak Çavdaroğlu, Erik Hammel, John Mitchell, Thomas Sharkey, and William Wallace. Integrating restoration and scheduling decisions for disrupted interdependent infrastructure systems. *Annals of Operations Research*, 203(1):279–294, 2013.

[31] Pew Research Center. Mobile fact sheet, 2021.

[32] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589(7840):82–87, January 2021.

[33] Stephanie E. Chang. Urban disaster recovery: a measurement framework and its application to the 1995 kobe earthquake. *Disasters*, 34(2):303–327, 2010.

[34] Stephanie E. Chang, Timothy Mcdaniels, Jana Fox, Rajan Dhariwal, and Holly Longstaff. Toward disaster-resilient cities: Characterizing resilience of infrastructure systems with expert judgments. *Risk Analysis*, 34(3):416–434, 2014.

[35] Stephanie E. Chang, Timothy L. McDaniels, Joey Mikawoz, and Krista Peterson. Infrastructure failure interdependencies in extreme events: power outage consequences in the 1998 ice storm. *Natural Hazards*, 41(2):337–358, 2007.

[36] Stephanie E. Chang, Hope A. Seligson, and Ronald T. Eguchi. Estimation of the economic impact of multiple lifeline disruption: Memphis light, gas and water division case study. Technical report, NCEER, 1996.

[37] Stephanie E Chang, Walter D Svekla, and Masanobu Shinozuka. Linking infrastructure and urban economy: Simulation of water-disruption impacts in earthquakes. *Environment and Planning B: Planning and Design*, 29(2):281–301, 2002.

[38] Kyle Chard, Ian Foster, and Steven Tuecke. Globus: Research data management as service and platform. PEARC17, New York, NY, USA, 2017. Association for Computing Machinery.

[39] Jeongwook Choi, Do Guen Yoo, and Doosun Kang. Post-earthquake restoration simulation model for water supply networks. *Sustainability*, 10(10):3618, 2018.

[40] Gian Paolo Cimellaro, Daniele Solari, and Michel Bruneau. Physical infrastructure interdependency and regional resilience index after the 2011 Tōhoku earthquake in Japan. *Earthquake Engineering & Structural Dynamics*, 43(12):1763–1784, 2014.

[41] Jonathan Cinnamon and Lindi Jahiu. Panoramic Street-Level Imagery in Data-Driven Urban Research: A Comprehensive Global Review of Applications, Techniques, and Practical Considerations. *ISPRS International Journal of Geo-Information*, 10(7):471, July 2021.

[42] City of Seattle. Seattle Open Data, 1995.

[43] Luca Clissa, Mario Lassnig, and Lorenzo Rinaldi. How big is Big Data? A comprehensive survey of data production, storage, and streaming in science and industry. *Frontiers in Big Data*, 6:1271639, October 2023.

[44] Colorado State University. Centerville community resilience testbed, 2018. Accessed: 2020-01-25.

[45] Roger M. Cooke. *Experts in uncertainty : opinion and subjective probability in science*. Environmental ethics and science policy. Oxford University Press, New York, 1991.

[46] Lee Cooper, Noel N. Schulz, and Terry Nielsen. Computer program for the estimation of restoration times during storms. volume 2, pages 777–780. Illinois Inst of Technology, 1998.

[47] Andrew Curtis and Jacqueline W. Mills. Spatial video data collection in a post-disaster landscape: The Tuscaloosa Tornado of April 27th 2011. *Applied Geography*, 32(2):393–400, March 2012.

[48] S Ćurčić, C.S Özveren, L Crowe, and P.K.L Lo. Electric power distribution network restoration: a survey of papers and a review of the restoration problem. *Electric Power Systems Research*, 35(2):73–86, 1995.

[49] Dhaval Dave, Drew McNichols, and Joseph J. Sabia. The contagion externality of a superspreading event: The Sturgis Motorcycle Rally and COVID-19. *Southern economic journal*, 87(3):769–807, 2021. Place: Hoboken, USA Publisher: John Wiley & Sons, Inc.

[50] R. A. Davidson, Haibin Liu, and Tatiyana V. Apanasovich. *Estimation of Post-Storm Restoration Times for Electric Power Distribution Systems*, chapter 7, pages 251–284. John Wiley & Sons, Ltd, 2017.

[51] R. A. Davidson, Haibin Liu, Isaac K Sarpong, Peter Sparks, and David V Rosowsky. Electric power distribution system performance in carolina hurricanes. *Natural Hazards Review*, 4(1):36–45, 2003.

[52] Leen Depauw, Haben Blondeel, Emiel De Lombaerde, Karen De Pauw, Dries Landuyt, Eline Lorer, Pieter Vangansbeke, Thomas Vanneste, Kris Verheyen, and Pieter De Frenne. The use of photos to investigate ecological change. *Journal of Ecology*, 110(6):1220–1236, June 2022.

[53] Alejandro Díaz-Delgado Bragado. Downtime estimation of lifelines after an earthquake. Master's thesis, University of California, Berkeley, United States, 2016.

[54] Shaohua Dong, Wujie Zhou, Caie Xu, and Weiqing Yan. Egfnet: Edge-aware guidance fusion network for rgb–thermal urban scene parsing. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–13, 2023.

[55] Annie Doubleday, Youngjun Choe, Tania Busch Isaksen, Scott Miles, and Nicole A Errett. How did outdoor biking and walking change during COVID-19?: A case study of three U.S. cities. *PloS one*, 16(1):e0245514, 2021. Place: United States Publisher: Public Library of Science.

[56] Annie Doubleday, Youngjun Choe, Scott Miles, and Nicole Errett. Daily bicycle and pedestrian activity as an indicator of disaster recovery: A Hurricane Harvey case study. *International Journal of Environmental Research and Public Health*, 16(16), 2019.

[57] Dublin Core Metadata Initiative. Dublin core metadata initiative metadata terms, 2008.

[58] Leonardo Dueñas-Osorio and Alexis Kwasinski. Quantification of lifeline system interdependencies after the 27 February 2010 M sub(w) 8.8 Offshore Maule, Chile, earthquake. *Earthquake Spectra*, 28(S1):S581–S603, 2012.

[59] R. B Duffey and Taesung Ha. The probability and timing of power system restoration. *IEEE Transactions on Power Systems*, 28(1):3–9, 2013.

[60] Romney B Duffey. Power restoration prediction following extreme events and disasters. *International Journal of Disaster Risk Science*, 10(1):134–148, 2019.

[61] Erica Eassom, Domenico Giacco, Aysegul Dirik, and Stefan Priebe. Implementing family involvement in the treatment of patients with psychosis: a systematic review of facilitating and hindering factors. *BMJ open*, 4(10):e006108, 2014.

[62] Justin Elarde, Joon-Seok Kim, Hamdi Kavak, Andreas Züfle, and Taylor Anderson. Change of human mobility during COVID-19: A United States case study. *PLOS ONE*, 16(11):e0259031, November 2021. Publisher: Public Library of Science.

[63] M Emery and C Flora. Spiraling-up: mapping community transformation with community capitals framework. *Community development (Columbus, Ohio)*, 37(1):19–35, 2006. Place: Columbus Publisher: Taylor & Francis Group.

[64] Joshua M Epstein. Why model? *Journal of Artificial Societies and Social Simulation*, 11(4):12, 2008.

[65] Nicole A. Errett, Joseph Wartman, Scott B. Miles, Ben Silver, Matthew Martell, and Youngjun Choe. Street view data collection design for disaster reconnaissance. *arXiv preprint arXiv:2308.06284*, 2023.

[66] Xiaodan Fan, Xuelin Zhang, A.U. Weerasuriya, Jian Hang, Liyue Zeng, Qiqi Luo, Cruz Y. Li, and Zhenshun Chen. Numerical investigation of the effects of environmental conditions, droplet size, and social distancing on droplet transmission in a street canyon. *Building and Environment*, 221:109261, August 2022.

[67] FEMA. Hazus: FEMA's methodology for estimating potential losses from disasters. `https://www.fema.gov/hazus`, 2011. Accessed: 2020-01-25.

[68] Francesco Finazzi. Replacing discontinued big tech mobility reports: a penetration-based analysis. *Scientific reports*, 13(1):935–935, 2023.

[69] A Stewart Fotheringham and David WS Wong. The modifiable areal unit problem in multivariate statistical analysis. *Environment and planning A*, 23(7):1025–1044, 1991.

[70] Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences - PNAS*, 114(50):13108–13113, 2017.

[71] K. M. Gegner, A. B. Birchfield, Ti Xu, K. S. Shetye, and T. J. Overbye. A methodology for the creation of geographically realistic synthetic power flow models. In *2016 IEEE Power and Energy Conference at Illinois (PECI)*, pages 1–6, 2016.

[72] Robert Gentleman and Duncan Temple Lang. Statistical analyses and reproducible research. *Journal of Computational and Graphical Statistics*, 16(1):1–23, 2007.

[73] Laurel P Gibson, Renee E Magnan, Emily B Kramer, and Angela D Bryan. Theory of Planned Behavior Analysis of Social Distancing During the COVID-19 Pandemic: Focusing on the Intention–Behavior Gap. *Annals of behavioral medicine*, 55(8):805–812, 2021. Place: US Publisher: Oxford University Press.

[74] Cheryl Gilge. Google Street View and the Image as Experience. *GeoHumanities*, 2(2):469–484, 2016. Publisher: Routledge _eprint: https://doi.org/10.1080/2373566X.2016.1217741.

[75] Andrés D. González, Airlie Chapman, Leonardo Dueñas-Osorio, Mehran Mesbahi, and Raissa M. D' Souza. Efficient infrastructure restoration strategies using the recovery operator. *Computer-Aided Civil and Infrastructure Engineering*, 32(12):991–1006, 2017.

[76] Andrés D. González, Leonardo Dueñas-Osorio, Mauricio Sánchez-Silva, and Andrés L. Medaglia. The interdependent network design problem for optimal infrastructure system restoration. *Computer-Aided Civil and Infrastructure Engineering*, 31(5):334–350, 2016.

[77] Jesús González-Barahona and Gregorio Robles. On the reproducibility of empirical software engineering studies based on data retrieved from development repositories. *Empirical Software Engineering*, 17(1-2):75–89, 2012.

[78] Heidi Green, Ritin Fernandez, and Catherine MacPhail. The social determinants of health and health outcomes among adults during the COVID-19 pandemic: A systematic review. *Public Health Nursing*, 38(6):942–952, November 2021.

[79] Seth D. Guikema, Steven M. Quiring, and Seung-Ryong Han. Prestorm estimation of hurricane damage to electric power distribution systems. *Risk Analysis*, 30(12):1744–1752, 2010.

[80] Pedro Gullon, Dustin Fry, Jesse J. Plascak, Stephen J. Mooney, and Gina S. Lovasi. Measuring changes in neighborhood disorder using Google Street View longitudinal imagery: a feasibility study. *Cities & Health*, 7(5):823–829, September 2023.

[81] Peter Haddawy, Saranath Lawpoolsri, Chaitawat Sa-ngamuang, Myat Su Yin, Thomas Barkowsky, Anuwat Wiratsudakul, Jaranit Kaewkungwal, Amnat Khamsiriwatchara, Patiwat Sa-angchai, Jetsumon Sattabongkot, and Liwang Cui. Effects of COVID-19 government travel restrictions on mobility in a rural border area of Northern Thailand: A mobile phone tracking study. *PLOS ONE*, 16(2):e0245842, February 2021.

[82] Jian Hang, Xia Yang, Cui-Yun Ou, Zhi-Wen Luo, Xiao-Dan Fan, Xue-Lin Zhang, Zhong-Li Gu, and Xian-Xiang Li. Assessment of exhaled pathogenic droplet dispersion and indoor-outdoor exposure risk in urban street with naturally-ventilated buildings. *Building and Environment*, 234:110122, April 2023.

[83] Xin Hao, Heng Lyu, Ze Wang, Shengnan Fu, and Chi Zhang. Estimating the spatial-temporal distribution of urban street ponding levels from surveillance videos based on computer vision. *Water Resources Management*, 36(6):1799–1812, April 2022.

[84] Irtiza Hasan, Shengcai Liao, Jinpeng Li, Saad Ullah Akram, and Ling Shao. Generalizable pedestrian detection: The elephant in the room. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11328–11337, 2021.

[85] Irtiza Hasan, Shengcai Liao, Jinpeng Li, Saad Ullah Akram, and Ling Shao. Pedestrian detection: Domain generalization, cnns, transformers and beyond. *arXiv preprint arXiv:2201.03176*, 2022.

[86] Xian He and Eun Jeong Cha. Modeling the damage and recovery of interdependent critical infrastructure systems from natural hazards. *Reliability Engineering and System Safety*, 177:162–175, 2018.

[87] Brittany N. Hearne and Michael D. Niño. Understanding How Race, Ethnicity, and Gender Shape Mask-Wearing Adherence During the COVID-19 Pandemic: Evidence from the COVID Impact Survey. *Journal of Racial and Ethnic Health Disparities*, 9(1):176–183, February 2022.

[88] Marccus D. Hendricks and Michelle Annette Meyer. Modeling Long-Term Housing Recovery after Technological Disaster Using a Virtual Audit with Repeated Photography. *Journal of planning education and research*, page 739456, 2021. Publisher: SAGE Publishing.

[89] Isaac Hernandez-Fajardo and Leonardo Dueñas Osorio. Probabilistic study of cascading failures in complex interdependent lifeline systems. *Reliability Engineering and System Safety*, 111:260–272, 2013.

[90] Michelle L. Holshue, Chas DeBolt, Scott Lindquist, Kathy H. Lofy, John Wiesman, Hollianne Bruce, Christopher Spitters, Keith Ericson, Sara Wilkerson, Ahmet Tural, George Diaz, Amanda Cohn, LeAnne Fox, Anita Patel, Susan I. Gerber, Lindsay Kim, Suxiang Tong, Xiaoyan Lu, Steve Lindstrom, Mark A. Pallansch, William C. Weldon, Holly M. Biggs, Timothy M. Uyeki, and Satish K. Pillai. First Case of 2019 Novel Coronavirus in the United States. *New England Journal of Medicine*, 382(10):929–936, March 2020.

[91] Boyeong Hong, Bartosz J. Bonczak, Arpit Gupta, and Constantine E. Kontokosta. Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nature Communications*, 12(1):1870, March 2021.

[92] Boyeong Hong, Bartosz J Bonczak, Arpit Gupta, and Constantine E Kontokosta. Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nature communications*, 12(1):1–9, 2021.

[93] Md. Emran Hossain, Md. Sayemul Islam, Md. Jaber Rana, Md. Ruhul Amin, Mohammed Rokonuzzaman, Sudipto Chakrobortty, and Sourav Mohan Saha. Scaling the changes in lifestyle, attitude, and behavioral patterns among COVID-19 vaccinated people: insights from Bangladesh. *Human vaccines & immunotherapeutics*, 18(1):2022920–2022920, 2022. Place: United States Publisher: Taylor & Francis.

[94] Xiao Hou, Song Gao, Qin Li, Yuhao Kang, Nan Chen, Kaiping Chen, Jinmeng Rao, Jordan S. Ellenberg, and Jonathan A. Patz. Intracounty modeling of covid-19 infection

with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proceedings of the National Academy of Sciences*, 118(24):e2020524118, 2021.

[95] Suzan Iloglu and Laura A Albert. An integrated network design and scheduling problem for network recovery and emergency response. *Operations Research Perspectives*, 5:218–231, 2018.

[96] Ryoji Isoyama and Tsuneo Katayama. Practical performance evaluation of water supply networks during seismic disaster. In *Lifeline Earthquake Engineering: The Current State of Knowledge, 1981 ; Proceedings of the Second Specialty Conference of the Technical Council on Lifeline Earthquake Engineering, Oakland Hyatt House, Oakland, California, August 20-21, 1981*, pages 111–126. ASCE, 1981.

[97] Masanori Isumi, Noriaki Nomura, and Takao Shibuya. Simulation of post-earthquake restoration for lifeline systems. *International Journal of Mass Emergencies and Disasters*, 3(1):87–105, 1985.

[98] Grant D. Jacobsen and Kathryn H. Jacobsen. Statewide COVID-19 Stay-at-Home Orders and Population Mobility in the United States. *World Medical & Health Policy*, 12(4):347–356, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wmh3.350.

[99] Jonas Klingwort and Rainer Schnell. Critical Limitations of Digital Epidemiology. *Survey research methods*, 14(2), 2020. Publisher: European Survey Research Association.

[100] Omar Kammouh, Gian Paolo Cimellaro, and Stephen A Mahin. Downtime estimation and analysis of lifelines after an earthquake. *Engineering Structures*, 173:393–403, 2018.

[101] Robert Kates and David Pijawka. From rubble to monument: The pace of reconstruction. In J. E. Haas, R. W. Kates, M. J. Bowden, editor, *Reconstruction following disaster*, pages 1–23. MIT Press, Cambridge, Mass., 1977.

[102] Alyson Kenward and Urooj Raja. Blackout: Extreme Weather, Climate Change and Power Outages, 2014. Accessed: 2020-06-30.

[103] King County Department of Public Health. COVID-19 race and ethnicity data.

[104] King County Department of Public Health. Summary of COVID-19 vaccination among King County Residents.

[105] King County GIS Center. King County GIS Data Hub, 2017.

[106] Sebastian Kraus and Nicolas Koch. Provisional COVID-19 infrastructure induces large, rapid increases in cycling. *Proceedings of the National Academy of Sciences*, 118(15):e2024399118, April 2021.

[107] Vaidyanathan Krishnamurthy, Alexis Kwasinski, and Leonardo Dueñas Osorio. Comparison of power and telecommunications dependencies and interdependencies in the 2011 Tōhoku and 2010 Maule Earthquakes. *Journal of Infrastructure Systems*, 22(3), 2016.

[108] Yury Kryvasheyeu, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. Rapid assessment of disaster damage using social media activity. *Science Advances*, 2(3):e1500779, March 2016.

[109] E.E Lee, J.E Mitchell, and W.A Wallace. Restoration of services in interdependent infrastructure systems: A network flows approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6):1303–1317, 2007.

[110] Jori Lewis. Exposures in the city: Looking for socioeconomic patterns for the urban exposome. *Environmental Health Perspectives*, 127(4):044003, 2019.

[111] Yongchang Li, Li Peng, Chengwei Wu, and Jiazhen Zhang. Street View Imagery (SVI) in the Built Environment: A Theoretical and Systematic Review. *Buildings*, 12(8):1167, August 2022.

[112] Lucy H Lin, Scott B Miles, and Noah A Smith. Natural language processing for analyzing disaster recovery trends expressed in large text corpora. In *2018 IEEE Global Humanitarian Technology Conference (GHTC)*, pages 1–8. IEEE, 2018.

[113] Michael K. Lindell. Recovery and reconstruction after disaster. In Peter T. Bobrowsky, editor, *Encyclopedia of Natural Hazards*, pages 812–824. Springer Netherlands, Dordrecht, 2013.

[114] Haibin Liu, R. A. Davidson, David V Rosowsky, and Jery R Stedinger. Negative binomial regression of electric power outages in hurricanes. *Journal of Infrastructure Systems*, 11(4):258–267, 2005.

[115] Haibin Liu, R.A Davidson, and T Apanasovich. Statistical forecasting of electric power restoration times in hurricanes and ice storms. *IEEE Transactions on Power Systems*, 22(4):2270–2279, 2007.

[116] Yutian Liu, Rui Fan, and Vladimir Terzija. Power system restoration: a literature review from 2006 to 2016. *Journal of Modern Power Systems and Clean Energy*, 4(3):332–341, 2016.

[117] Zhanlin Liu, Pariyakorn Maneekul, Claire Pendergrast, Annie Doubleday, Scott B. Miles, Nicole A. Errett, and Youngjun Choe. Physical activity monitoring data following disasters. *Sustainable Cities and Society*, 81:103814, 2022.

[118] Gil Loewenthal, Shiran Abadi, Oren Avram, Keren Halabi, Noa Ecker, Natan Nagar, Itay Mayrose, and Tal Pupko. COVID-19 pandemic-related lockdown: response time is more important than its strictness. *EMBO Molecular Medicine*, 12(11):e13171, 2020. _eprint: https://www.embopress.org/doi/pdf/10.15252/emmm.202013171.

[119] Ryan Loggins, Richard G Little, John Mitchell, Thomas Sharkey, and William A Wallace. CRISIS: modeling the restoration of interdependent civil and social infrastructure systems following an extreme event. *Natural Hazards Review*, 20(3), 2019.

[120] Ronaldo Luna, Nandini Balakrishnan, and Cihan H Dagli. Postearthquake recovery of a water distribution system: Discrete event simulation using colored petri nets. *Journal of Infrastructure Systems*, 17(1):25–34, 2011.

[121] Jian Ma, Yongman Guo, Jing Gao, Hanxing Tang, Keqiang Xu, Qiyong Liu, and Lei Xu. Climate Change Drives the Transmission and Spread of Vector-Borne Diseases: An Ecological Perspective. *Biology*, 11(11):1628, November 2022.

[122] Cameron A MacKenzie and Kash Barker. Empirical data and regression analysis for estimation of infrastructure resilience with application to electric power outages. *Journal of Infrastructure Systems*, 19(1):25–35, 2013.

[123] Ron Mahabir, Ross Schuchard, Andrew Crooks, Arie Croitoru, and Anthony Stefanidis. Crowdsourcing Street View Imagery: A Comparison of Mapillary and OpenStreetCam. *ISPRS International Journal of Geo-Information*, 9(6), 2020.

[124] Mapillary. Mapillary, 2013.

[125] Leandro Tolomeu Marques, Alexandre Claudio B Delbem, and Joao Bosco Augusto London. Service restoration with prioritization of customers and switches and determination of switching sequence. *IEEE Transactions on Smart Grid*, 9(3):2359–2370, 2018.

[126] Matthew Martell, Scott B Miles, and Youngjun Choe. Review of empirical quantitative data use in lifeline infrastructure restoration modeling. *Natural hazards review*, 22(4), 2021.

[127] Matthew Martell, Nick Terry, Ribhu Sengupta, Chris Salazar, Nicole A. Errett, Scott B. Miles, Joseph Wartman, and Youngjun Choe. Open-source data pipeline for street-view images: a case study on community mobility during covid-19 pandemic, 2024.

[128] Timothy Mcdaniels, Stephanie Chang, Krista Peterson, Joey Mikawoz, and Dorothy Reed. Empirical framework for characterizing infrastructure failure interdependencies. *Journal of Infrastructure Systems*, 13(3):175–184, 2007.

[129] Silvia Mendolia, Olena Stavrunova, and Oleg Yerokhin. Determinants of the community mobility during the covid-19 epidemic: The role of government regulations and information. *Journal of economic behavior & organization*, 184:199–231, 2021.

[130] Michelle Annette Meyer and Marccus D. Hendricks. Using Photography to Assess Housing Damage and Rebuilding Progress for Disaster Recovery Planning. *Journal of the American Planning Association*, 84(2):127–144, 2018. Place: ABINGDON Publisher: Taylor & Francis.

[131] Scott B Miles. Foundations of community disaster resilience: Well-being, identity, services, and capitals. *Environmental Hazards*, 14(2):103–121, 2015.

[132] Scott B Miles, Henry V Burton, and Hua Kang. Community of practice for modeling disaster recovery. *Natural Hazards Review*, 20(1):4018023, 2019.

[133] Scott B. Miles and Stephanie E. Chang. Modeling community recovery from earthquakes. *Earthquake Spectra*, 22(2):439–458, 2006.

[134] Scott B Miles and Stephanie E Chang. Resilus: A community based disaster resilience model. *Cartography and Geographic Information Science*, 38(1):36–51, 2011.

[135] Scott B Miles, Hannah Gallagher, and Charles J Huxford. Restoration and impacts from the September 8, 2011, San Diego Power Outage. *Journal of Infrastructure Systems*, 20(2), 2014.

[136] Sveta Milusheva, Daniel Bjorkegren, and Leonardo Viotti. Assessing bias in smartphone mobility estimates in low income countries. In *ACM SIGCAS Conference on Computing and Sustainable Societies*, pages 364–378, 2021.

[137] D Mitsova, Am Esnard, A Sapat, and BS Lai. Socioeconomic vulnerability and electric power restoration timelines in Florida: the case of Hurricane Irma. *Natural Hazards*, 94(2):689–709, 2018.

[138] Mohammad Mojtahedi, Sidney Newton, and Jason Meding. Predicting the resilience of transport infrastructure to a natural disaster using Cox's proportional hazards regression model. *Natural Hazards*, 85(2):1119–1133, 2017.

[139] Mauricio Monsalve and Juan Carlos de La Llera. Data-driven estimation of interdependencies and restoration of infrastructure systems. *Reliability Engineering and System Safety*, 181:167–180, 2019.

[140] Hiroyoshi Morita, Shinichiro Nakamura, and Yoshitsugu Hayashi. Changes of Urban Activities and Behaviors Due to COVID-19 in Japan. *SSRN Electronic Journal*, 2020.

[141] Gareth Morley. Rights and religion: Canadian and American courts face challenges over COVID restrictions on religious gatherings. *Inroads (Ottawa)*, (49):36, 2021. Publisher: Inroads, Inc.

[142] Rebecca E. Morss, Olga V. Wilhelmi, Gerald A. Meehl, and Lisa Dilling. Improving Societal Outcomes of Extreme Weather in a Changing Climate: An Integrated Perspective. *Annual Review of Environment and Resources*, 36(1):1–25, November 2011.

[143] Sayanti Mukherjee, Roshanak Nateghi, and Makarand Hastak. Data on major power outage events in the continental U.S. *Data in Brief*, 19:2079 – 2083, 2018.

[144] Sayanti Mukherjee, Roshanak Nateghi, and Makarand Hastak. A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S. *Reliability Engineering and System Safety*, 175:283–305, 2018.

[145] Claire Mulry, Giselle Rivera, Kristopher Musini, Alex Rankin, and Marissa Astorini. The relationship between community mobility, health status, and quality of life (qol) in older adults. *The American Journal of Occupational Therapy*, 73(4_Supplement_1):7311505078p1–7311505078p1, 2019.

[146] Multidisciplinary Center for Earthquake Engineering Research. MCEER, 1986.

[147] Nitish S Mutha. How to map equirectangular projection to rectilinear projection, 2017.

[148] Roshanak Nateghi, Seth Guikema, and Steven Quiring. Forecasting hurricane-induced power outage durations. *Natural Hazards*, 74(3):1795–1811, 2014.

[149] Roshanak Nateghi, Seth D. Guikema, and Steven M. Quiring. Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes. *Risk Analysis*, 31(12):1897–1906, 2011.

[150] National Institute of Standards and Technology (NIST). *Community Resilience Planning Guide for Buildings and Infrastructure*. 2015.

[151] Ali Nejat and Souparno Ghosh. LASSO model of postdisaster housing recovery: Case study of Hurricane Sandy. *Natural Hazards Review*, 17(3), 2016.

[152] N. Nojima and H. Kato. Validation of an assessment model of post-earthquake lifeline serviceability - based on the Great East Japan Earthquake Disaster -. *Journal of Social Safety Science*, 18:229–239, 2012.

[153] N. Nojima and H. Kato. Modification and validation of an assessment model of post-earthquake lifeline serviceability based on the Great East Japan Earthquake disaster. *Journal of Disaster Research*, 9(2):108–120, 2014.

[154] N. Nojima and M. Sugito. Empirical estimation of outage and duration of lifeline disruption due to earthquake disaster. *Proc. of the Fourth China-Japan-U.S. Trilateral Symposium on Lifeline Earthquake Engineering*, pages 349–356, 2002.

[155] N. Nojima and M. Sugito. Probabilistic assessment model for post earthquake serviceability of utility lifelines and its practical application. *Proc. of the 9th International Conference on Structural Safety and Reliability*, pages 279–287, 2005.

[156] Nobuoto Nojima and Yoshihisa Maruyama. Comparison of functional damage and restoration processes of utility lifelines in the 2016 Kumamoto earthquake, japan with two great earthquake disasters in 1995 and 2011. *JSCE Journal of Disaster FactSheets*, 2016.

[157] Tessio Novack, Leonard Vorbeck, Heinrich Lorei, and Alexander Zipf. Towards Detecting Building Facades with Graffiti Artwork Based on Street View Images. *ISPRS International Journal of Geo-Information*, 9(2), 2020.

[158] Sarah G Nurre, Burak Çavdaroğlu, John E Mitchell, Thomas C Sharkey, and William A Wallace. Restoring infrastructure systems: An integrated network design and scheduling (INDS) problem. *European Journal of Operational Research*, 223(3):794–806, 2012.

[159] Sarah G. Nurre and Thomas C. Sharkey. Integrated network design and scheduling problems with parallel identical machines: Complexity results and dispatching rules. *Networks*, 63(4):306–326, 2014.

[160] Oregon Seismic Safety Policy Advisory Commission (OSSPAC). *The Oregon Resilience Plan: Reducing Risk and Improving Recovery for the Next Cascadia Earthquake and Tsunami*. 2013.

[161] T. D. O'Rourke and Thomas Briggs. Critical infrastructure, interdependencies, and resilience. *The Bridge*, 37(1):22–29, 2007.

[162] Ashley Ossimetha, Angelina Ossimetha, Cyrus M Kosar, and Momotazur Rahman. Socioeconomic disparities in community mobility reduction and covid-19 growth. In *Mayo Clinic Proceedings*, volume 96, pages 78–85. Elsevier, 2021.

[163] Min Ouyang. Review on modeling and simulation of interdependent critical infrastructure systems. *Reliability Engineering and System Safety*, 121:43–60, 2014.

[164] Min Ouyang and Leonardo Dueñas Osorio. Multi-dimensional hurricane resilience assessment of electric power systems. *Structural Safety*, 48:15–24, 2014.

[165] Min Ouyang, Leonardo Dueñas Osorio, and Xing Min. A three-stage resilience analysis framework for urban infrastructure systems. *Structural Safety*, 36-37:23–31, 2012.

[166] Min Ouyang and Zhenghua Wang. Resilience assessment of interdependent infrastructure systems: With a focus on joint restoration modeling and analysis. *Reliability Engineering and System Safety*, 141:74, 2015.

[167] Antonio Paez. Using google community mobility reports to investigate the incidence of covid-19 in the united states. *Findings*, 5 2020.

[168] J. Park, N. Nojima, and D.A. Reed. Nisqually earthquake electric utility analysis. *Earthquake Spectra*, 22(2):491–509, 2006.

[169] Sunhee Park, Beomsoo Kim, and Jaeil Lee. Social Distancing and Outdoor Physical Activity During the COVID-19 Outbreak in South Korea: Implications for Physical Distancing Strategies. *Asia-Pacific journal of public health*, 32(6-7):360–362, 2020. Place: Los Angeles, CA Publisher: SAGE Publications.

[170] Johanna Petersen, Daniel Austin, Nora Mattek, and Jeffrey Kaye. Time out-of-home and cognitive, physical, and emotional wellbeing of older adults: a longitudinal mixed effects model. *PloS one*, 10(10):e0139643, 2015.

[171] Hannele Polku, Tuija M Mikkola, Erja Portegijs, Merja Rantakokko, Katja Kokko, Markku Kauppinen, Taina Rantanen, and Anne Viljanen. Life-space mobility and dimensions of depressive symptoms among community-dwelling older adults. *Aging & Mental Health*, 19(9):781–789, 2015.

[172] Li Qi, Wenge Tang, Ju Wang, Yu Xiong, Yi Yuan, Baisong Li, Lin Yang, Tingting Li, Lianjian Yang, Xiaoyuan Su, Qin Li, and Lijie Zhang. An Outbreak of SARS-CoV-2 Omicron Subvariant BA.2.76 in an Outdoor Park - Chongqing Municipality, China, August 2022. *China CDC Weekly*, 4(46):1039–1042, 2022. Place: China Publisher: Editorial Office of CCDCW, Chinese Center for Disease Control and Prevention.

[173] J. Rexiline Ragini, P.M. Rubesh Anand, and Vidhyacharan Bhaskar. Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management*, 42:13–24, 2018.

[174] V Ramachandran, S. K Long, T Shoberg, S Corns, and H. J Carlo. Framework for modeling urban restoration resilience time in the aftermath of an extreme event. *Natural Hazards Review*, 16(4), 2015.

[175] Ellen M Rathje, Clint Dawson, Jamie E Padgett, Jean-Paul Pinelli, Dan Stanzione, Ashley Adair, Pedro Arduino, Scott J Brandenberg, Tim Cockerill, Charlie Dey, Maria Esteva, Fred L Haan, Matthew Hanlon, Ahsan Kareem, Laura Lowes, Stephen Mock, and Gilberto Mosqueda. Designsafe: New cyberinfrastructure for natural hazards engineering. *Natural Hazards Review*, 18(3):06017001, 2017.

[176] Ellen M Rathje, Clint Dawson, Jamie E Padgett, Jean-Paul Pinelli, Dan Stanzione, Ashley Adair, Pedro Arduino, Scott J Brandenberg, Tim Cockerill, Charlie Dey, Maria Esteva, Fred L Haan, Matthew Hanlon, Ahsan Kareem, Laura Lowes, Stephen Mock, and Gilberto Mosqueda. Designsafe: New cyberinfrastructure for natural hazards engineering. *Natural hazards review*, 18(3), 2017.

[177] Dorothy A Reed. Electric utility distribution analysis for extreme winds. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(1):123–140, 2008.

[178] S.M Rinaldi, J.P Peerenboom, and T.K Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems*, 21(6):11–25, 2001.

[179] John Ringland, Martha Bohm, So-Ra Baek, and Matthew Eichhorn. Automated survey of selected common plant species in Thai homegardens using Google Street View imagery and a deep neural network. *Earth Science Informatics*, 14(1):179–191, March 2021.

[180] Roger Finke, Christopher Bader, and Andrew Whitehead. The Association of Religion Data Archives, 1997.

[181] Kenneth A Rose, Shaye Sable, Donald L Deangelis, Simeon Yurek, Joel C Trexler, William Graf, and Denise J Reed. Proposed best modeling practices for assessing the effects of ecosystem restoration on fish. *Ecological Modelling*, 300:12–29, 2015.

[182] Avipsa Roy, Trisalyn A. Nelson, A. Stewart Fotheringham, and Meghan Winters. Correcting bias in crowdsourced data to map bicycle ridership of all bicyclists. *Urban Science*, 3(2), 2019.

[183] Amanda Rzotkiewicz, Amber L. Pearson, Benjamin V. Dougherty, Ashton Shortridge, and Nick Wilson. Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. *Health & Place*, 52:240–246, July 2018.

[184] Christopher Salazar. Estimating distance between pedestrians from street view images using geometric properties, 2021.

[185] Evan W. Sandlin and Daniel J. Simmons. Polarized perceptions: how time and vaccination status modify Republican and Democratic COVID-19 risk perceptions. *Journal of Elections, Public Opinion and Parties*, pages 1–19, May 2023.

[186] Frank Schlosser, Benjamin F. Maier, Olivia Jack, David Hinrichs, Adrian Zachariae, and Dirk Brockmann. Covid-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences*, 117(52):32883–32890, 2020.

[187] Landon Schnabel and Scott Schieman. Religion Protected Mental Health but Constrained Crisis Response During Crucial Early Days of the COVID-19 Pandemic. *Journal for the scientific study of religion*, 61(2):530–543, 2022. Place: Hoboken Publisher: Wiley Subscription Services, Inc.

[188] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[189] Edmund Seto, Esther Min, Carolyn Ingram, Bj Cummings, and Stephanie A. Farquhar. Community-Level Factors Associated with COVID-19 Cases and Testing Equity in King County, Washington. *International Journal of Environmental Research and Public Health*, 17(24):9516, December 2020.

[190] Shuai Shao, Zijian Zhao, Boxun Li, Tete Xiao, Gang Yu, Xiangyu Zhang, and Jian Sun. Crowdhuman: A benchmark for detecting human in a crowd. *arXiv preprint arXiv:1805.00123*, 2018.

[191] Thomas C Sharkey, Burak Çavdaroğlu, Huy Nguyen, Jonathan Holman, John E Mitchell, and William A Wallace. Interdependent network restoration: On the value of information-sharing. *European Journal of Operational Research*, 244(1):309–321, 2015.

[192] Thomas C Sharkey, Sarah G Nurre, Huy Nguyen, Joe H Chow, John E Mitchell, and William A Wallace. Identification and classification of restoration interdependencies in the wake of Hurricane Sandy. *Journal of Infrastructure Systems*, 22(1), 2016.

[193] M Shinozuka, A Rose, and R.T Eguchi. Engineering and socioeconomic impacts of earthquakes: An analysis of electricity lifeline disruptions in the new madrid area, 1998.

[194] Ruishi Si, Yumeng Yao, Xueqian Zhang, Qian Lu, and Noshaba Aziz. Investigating the Links Between Vaccination Against COVID-19 and Public Attitudes Toward Protective Countermeasures: Implications for Public Health. *Frontiers in public health*, 9:702699–702699, 2021. Publisher: Frontiers Media S.A.

[195] Surya Singh, Mujaheed Shaikh, Katharina Hauck, and Marisa Miraldo. Impacts of introducing and lifting nonpharmaceutical interventions on COVID-19 daily growth rate and compliance in the United States. *Proceedings of the National Academy of Sciences*, 118(12):e2021359118, March 2021.

[196] Brian D Smedley, Adrienne Y Stith, Committee on Understanding Care, Eliminating Racial, and Ethnic Disparities in Health. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. National Academies Press, Washington, 2002. Backup Publisher: Board on Health Sciences Policy Publication Title: Unequal Treatment.

[197] Adam B Smith. U.s. billion-dollar weather and climate disasters, 1980-prsent.

[198] Cara M. Smith, Joel D. Kaufman, and Stephen J. Mooney. Google street view image availability in the Bronx and San Diego, 2007–2020: Understanding potential biases in virtual audits of urban built environments. *Health & Place*, 72:102701, November 2021.

[199] Cara M. Smith, Joel D. Kaufman, and Stephen J. Mooney. Google street view image availability in the bronx and san diego, 2007–2020: Understanding potential biases in virtual audits of urban built environments. *Health & Place*, 72:102701, 2021.

[200] Gavin P. Smith and Dennis Wenger. *Sustainable Disaster Recovery: Operationalizing An Existing Agenda*, pages 234–257. Springer New York, New York, NY, 2007.

[201] Dan Stanzione, John West, R. Todd Evans, Tommy Minyard, Omar Ghattas, and Dhabaleswar K. Panda. *Frontera: The Evolution of Leadership Computing at the National Science Foundation*, page 106–111. Association for Computing Machinery, New York, NY, USA, 2020.

[202] Nathaniel Stockham, Peter Washingon, Brianna Chrisman, Kelley Paskov, Jae-Yoon Jung, and Dennis P Wall. Causal Modeling to Mitigate Selection Bias and Unmeasured Confounding in Internet-Based Epidemiology of COVID-19: Model Development and

Validation. *JMIR public health and surveillance*, 8(7):e31306–e31306, 2022. Place: Canada Publisher: JMIR Publications.

[203] Esra Suel, Samir Bhatt, Michael Brauer, Seth Flaxman, and Majid Ezzati. Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. *Remote Sensing of Environment*, 257:112339, 2021.

[204] Laura Sullivan and Tatjana Meschede. How Measurement of Inequalities in Wealth by Race/Ethnicity Impacts Narrative and Policy: Investigating the Full Distribution. *Race and social problems*, 10(1):19–29, 2018. Place: New York Publisher: Springer US.

[205] L. Sun, M. Didier, E. Delé, and B. Stojadinović. Probabilistic demand and supply resilience model for electric power supply system under seismic hazard. In *12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP 2015*. University of British Columbia, 2015.

[206] S Szabo, S Walker, T Edgar, J Kretnzel, K Johnston, and B Stone. EPH7 A Survey of the Occurrence and Transmission of COVID-19 in Outdoor Education Settings in Canada. *ISPOR Europe 2022 Abstracts*, 25(12, Supplement):S192, December 2022.

[207] Kristine Sørensen, Orkan Okan, Barbara Kondilis, and Diane Levin-Zamir. Rebranding social distancing to physical distancing: calling for a change in the health promotion vocabulary to enhance clear communication during a pandemic. *Global health promotion*, 28(1):5–14, 2021. Place: London, England Publisher: SAGE Publications.

[208] Taronne Tabucchi, R. A. Davidson, and Susan A. Brink. Simulation of post-earthquake water supply system restoration. *Civil Engineering and Environmental Systems*, 27(4):263–279, 2010.

[209] T.H.P. Tabucchi and R.A. Davidson. *Post-Earthquake Restoration of the Los Angeles Water Supply System*. Multidisciplinary Center for Earthquake Engineering Research (MCEER), 2008.

[210] The White House. *Presidential Policy Directive 21– Critical Infrastructure Security and Resilience*. The White House, 2013.

[211] Tatjana Thimm and Ralf Seepold. Past, present and future of tourist tracking. *Journal of Tourism Futures*, 2(1):43–55, 2016.

[212] Dilek Tuzun Aksu and Linet Ozdamar. A mathematical model for post-disaster road restoration: Enabling accessibility and evacuation. *Transportation Research Part E*, 61:56–67, 2014.

[213] Lori Uscher-Pines, Heather L Schwartz, Faruque Ahmed, Yenlik Zheteyeva, Jennifer Tamargo Leschitz, Francesca Pillemer, Laura Faherty, and Amra Uzicanin. Feasibility of Social Distancing Practices in US Schools to Reduce Influenza Transmission During a Pandemic. *Journal of public health management and practice*, 26(4):357–370, 2020. Place: United States Publisher: Ovid Technologies (Wolters Kluwer Health).

[214] Julie Vallée, Emmanuelle Cadot, Christelle Roustit, Isabelle Parizot, and Pierre Chauvin. The role of daily mobility in mental health inequalities: the interactive influence of activity space and neighbourhood of residence on depression. *Social science & medicine*, 73(8):1133–1144, 2011.

[215] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.

[216] Sergio A Velastin, Rodrigo Fernández, Jorge E Espinosa, and Alessandro Bay. Detecting, Tracking and Counting People Getting On/Off a Metropolitan Train Using a Standard Video Camera. *Sensors (Basel, Switzerland)*, 20(21):6251, 2020. Place: Switzerland Publisher: MDPI AG.

[217] Alexander Veremyev, Alexey Sorokin, Vladimir Boginski, and Eduardo L Pasiliao. Minimum vertex cover problem for coupled interdependent networks with cascading failures. *European Journal of Operational Research*, 232(3):499–511, 2014.

[218] Lan Wang, Kaichen Zhou, Surong Zhang, Anne Vernez Moudon, Jinfeng Wang, Yong-Guan Zhu, Wenyao Sun, Jianfeng Lin, Chao Tian, and Miao Liu. Designing bike-friendly cities: Interactive effects of built environment factors on bike-sharing. *Transportation research. Part D, Transport and environment*, 117:103670–, 2023. Publisher: Elsevier Ltd.

[219] Ruoyu Wang, Yuan Yuan, Ye Liu, Jinbao Zhang, Penghua Liu, Yi Lu, and Yao Yao. Using street view data and machine learning to assess how perception of neighborhood safety influences urban residents' mental health. *Health & Place*, 59:102186, September 2019.

[220] Washington State Department of Health. Data & Statistical Reports, 2016.

[221] Washington State Seismic Safety Committee (WASSC). *Resilient Washington State: A Framework For Minimizing Loss And Improving Statewide Recovery After An Earthquake*. Washington State Emergency Management Council, 2012.

[222] Joakim A. Weill, Matthieu Stigler, Olivier Deschenes, and Michael R. Springborn. Social distancing responses to COVID-19 emergency declarations strongly differentiated

by income. *Proceedings of the National Academy of Sciences*, 117(33):19658–19660, August 2020. Publisher: Proceedings of the National Academy of Sciences.

[223] Milton C. Weinstein, Bernie O' Brien, John Hornberger, Joseph Jackson, Magnus Johannesson, Chris Mccabe, and Bryan R. Luce. Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ispor task force on good research practices—modeling studies. *Value in Health*, 6(1):9–17, 2003.

[224] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Merc/'e Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G Gray, Paul Groth, Carole Goble, Jeffrey S Grethe, Jaap Heringa, Peter A. C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J Lusher, Maryann E Martone, Albert Mons, Abel L Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3(1):160018, 2016.

[225] Baichao Wu, Aiping Tang, and Jie Wu. Modeling cascading failures in interdependent infrastructures under terrorist attacks. *Reliability Engineering and System Safety*, 147:1–8, 2016.

[226] Xueying Wu, Yi Lu, and Bin Jiang. Built environment factors moderate pandemic fatigue in social distance during the COVID-19 pandemic: A nationwide longitudinal study in the United States. *Landscape and Urban Planning*, 233:104690, May 2023.

[227] Ningxiong Xu, Seth D. Guikema, R. A. Davidson, Linda K. Nozick, Zehra Çağnan, and Kabeh Vaziri. Optimizing scheduling of post-earthquake electric power restoration tasks. *Earthquake Engineering & Structural Dynamics*, 36(2):265–284, 2007.

[228] Shangyao Yan and Yu-Lin Shih. Optimal scheduling of emergency roadway repair and subsequent relief distribution. *Computers and Operations Research*, 36(6):2049–2065, 2009.

[229] Wenyan Yang, Yanlin Qian, Joni-Kristian Kamarainen, Francesco Cricri, and Lixin Fan. Object Detection in Equirectangular Panorama. In *ICPR*, pages 2190–2195. IEEE, 2018. Journal Abbreviation: ICPR.

[230] Li Yin, Qimin Cheng, Zhenxin Wang, and Zhenfeng Shao. ‘Big data’ for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts. *Applied Geography*, 63:337–345, September 2015.

[231] Jin-Zhu Yu and Hiba Baroud. Quantifying community resilience using hierarchical bayesian kernel methods: A case study on recovery from power outages. *Risk Analysis*, 39(9):1930–1948, 2019.

[232] Hadi Zanganeh Kia, Yunsoo Choi, Delaney Nelson, Jincheol Park, and Arman Pouyaei. Large eddy simulation of sneeze plumes and particles in a poorly ventilated outdoor air condition: A case study of the University of Houston main campus. *The Science of the total environment*, 891:164694–164694, 2023. Publisher: Elsevier B.V.

[233] Wei Zhai, Mengyang Liu, and Zhong-Ren Peng. Social distancing and inequality in the united states amid covid-19 outbreak. *Environment and Planning A: Economy and Space*, 53(1):3–5, 2021.

[234] Jing Zhu and Yingling Fan. Daily travel behavior and emotional well-being: Effects of trip mode, duration, purpose, and companionship. *Transportation Research Part A: Policy and Practice*, 118:360–373, 2018.

[235] Conrad R Zorn and Asaad Y Shamseldin. Post-disaster infrastructure restoration: A comparison of events for future planning. *International Journal of Disaster Risk Reduction*, 13:158–166, 2015.

[236] Lei Zou, Nina S. N Lam, Heng Cai, and Yi Qiang. Mining twitter data for improved understanding of disaster resilience. *Annals of the American Association of Geographers*, 108(5):1422–1441, 2018.