Towards New Measures of Resilience: Leveraging Location Based Services Data for Evaluating Hazard-Induced Changes in Access to Essential Services and Community Recovery

by

Tessa Swanson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2023

Doctoral Committee:

Professor Seth Guikema, Chair
Professor James Bagian
Professor Amy Cohn
Professor Robert Goodspeed

Tessa Swanson
tlswan@umich.edu
ORCID iD: 0000-0002-4814-3436

DEDICATION

To my family and friends, who exhibit each in their own ways how to imagine and show up for a better future.

"Inside the word 'emergency' is 'emerge'; from an emergency new things come forth. The old certainties are crumbling fast, but danger and possibility are sisters."

-Rebecca Solnit, *Hope in the Dark*

ACKNOWLEDGEMENTS

To start, I must acknowledge my funding sources that made this work possible. This work was supported by the National Science Foundation Graduate Research Fellowship, grant number DGE 1841052. I was also supported by the University of Engineering's Skunkworks Initiative and the University of Michigan's Rackham Predoctoral Fellowship.

UM's IOE community is the primary reason I came to Michigan, and for all involved I am truly grateful. First and foremost, to Tom Logan, who's work and attitude continue to inspire me from the other side of the world. You have not only mentored me in my academic pursuits, but in how to be a better friend, advocate, and steward. Your hope in spite of it all ignites much of mine. I am so grateful to Adam VanDeusen and Anna White, who kept me grounded in times of overwhelm and reminded me through their example that our work could make a difference. Your commitment to service within your careers and beyond is what all engineers should aspire to. To Colonel Julia Oh Coxen, I've leaned on your encouraging words long after you moved on from Michigan. I deeply admire your simultaneous commitment to your family, friends, and personal goals. Thank you to the rest of the brilliant and creative researchers in the Guikema group: Chengwei Zhai, Elnaz Kabir, Thomas Chen, Tim Williams, Valerie Washington, Zaira Pagan Cajigas, Christopher Doehring, and Nolan Feeny. You are all full of fascinating ideas and amaze me with the breadth and depth of your skills. Another thank you to Valerie for providing the neatest, most well-commented code I've ever seen– your time and attention to detail made my life easier and will serve as an example for my future collaborations. Thank you so much to my cohort, who made IOE home, especially during the unique challenges of the early years. Seokhyun Chung, Luke DeRoos, Rohan Ghuge, Huiwen Jia, Yifan Li, Daniel Felipe Otero-Leon, Erkin Otles, Karanvir Panesar, Sajjad Seyedsalehi, Kam Tabattanon, Haoming Shen, Xubo Yue– I learned so much from you and I am very proud to be in your company.

To my advisor, Professor Seth Guikema, thank you for believing in the questions I was asking and encouraging me to acquire new skills to answer them. The research family you've developed is really special and is a testament to your vision for how risk scientists can make the world a better place. Thank you to my committee members for accepting the task of reviewing this dissertation and much of my other work over the years: to Professor Jim

Bagian, thank you for your guidance in our COVID-19 modeling work and for helping me understand how to better communicate ideas about risk for more impact; to Professor Amy Cohn, you've change the world not only with your own work, but in the work you've inspired through all of your students; to Professor Robert Goodspeed, thank you for embracing my urban planning interests as well as teaching me, and a new generation of planners, how to more inclusively, collaboratively, and sustainably adapt technology to serve our communities. Thank you to Professors Marina Epelman and Brian Denton for your efforts to prioritize students and make our program run smoothly, especially during those unprecedented times. I'd also like to thank Professor Karen Smilowitz, my undergraduate research advisor and mentor at Northwestern University, who was the first to tell me I could do this.

To my family, thank you for going along with this journey, even when you had no idea what it was. Thank you to my mom and my dad for letting me know it was OK when things felt hard– it meant more than you could know. To my sibling Jamie– I like to think some of this work could serve communities some day, but I know you're actually out there doing it. It's a joy to be privy to your creative energy and compassionate spirit.

To all my friends, I am eternally grateful for and impressed by how you show up. Alex, Matti, Max, Patsy, and Sammy– thanks for the endless adventures and support. You all made Ann Arbor home. To my oldest friends Nancy, Anne Burke, Allison, and Claiborne, thank you for always being there, no matter how much time has passed. To John Badir, for being an easy phone call and for bringing me back to my true passion: trains. Thank you to the Ni family: Cynthia, Sam, Mickey, and Nika for bringing me into your community and for making graduate school seem like so much fun. Thank you to Emma and Regan, your friendship has truly helped me navigate adulthood and I am so happy every moment I spend with you, from eating microwaved leftovers on the floor to peaking at the Era's Tour. To Alex Koh, David Tyson, and Michael and Mary Feldergraf– you have helped me grow into a better friend and a more enthusiastic dancer, in large part because of how free to be myself I feel around you; for that I am so grateful. To Neil, thank for all of the above and for believing in me as a researcher and as a friend. Thanks to Joanna, you are a role model for how to show up for people and have been such a positive force since I've known you, even while complaining about the woes of being women in our 20s. Your friendship energizes me. Jeremiah, I am so grateful for your companionship over the many slow mornings and late nights, through the stresses of a PhD, a global pandemic, and a hundred-pound puppy all at the same time. I can never repay you for all the coffee and comfort you've brought me– if it's any consolation, you made these years an adventure...you made these years rich. And to Liz, thank you for holding me, for validating me, for seeing me, for the confidence I won't have to go through whatever's next alone.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

ABSTRACT

Access to essential services determines individuals' ability to meet health, safety, and social needs that enable them to thrive in their daily lives. But such access is not evenly distributed across populations and disparities in access can be exacerbated when a community is faced with a disruption, like a natural hazard. How such access varies over populations, space, time, and in response to events is valuable for evaluating equity.

Opportunistically collected location-based services (LBS) data available from cell phones offers new opportunities to evaluate access to essential services. LBS data reveals regular mobility patterns of cell phone users over time. These mobility patterns can reveal regular visits to home, workplaces, and essential services facilities like supermarkets and schools. Deviations from those patterns may indicate a disruption, and how long individuals experience that disruption is impacted by how long it takes to meet essential needs. Data science, risk analysis, and urban planning offer tools to quantify those deviations and evaluate the factors contributing to recovery.

In this dissertation, I utilize LBS data with methods spanning data science, risk analysis, and urban planning to quantify the relationship between access and resilience in Southwestern Florida in the period surrounding Hurricane Irma in September 2017. In Chapter 2, I present a large-scale data-driven method to identify when facilities experience a change in visit patterns that may indicate change in access following a disruptive event, reflecting the actual closure of the facility or other barriers to access such as lack of supply or disruptions to the connecting transportation network. I demonstrate my approach by analyzing loss of access to supermarkets, schools, health care facilities, and home improvement stores in Southwest Florida both visually and using machine learning anomaly detection. Next, in Chapter 3 I present a novel Bayesian network based method for estimating recovery periods in LBS user's home and work appearances to show differences in household and workplace recovery over space and time. Results show the proportion of users experiencing an anomalous period and the average length of recovery consistent with the storm's path, with more users experiencing a workplace disruption but with home disruptions lasting longer on average. I validate these results against available survey results on evacuation. Finally, in Chapter 4 I statistically model the relationship between these estimated recovery periods and travel time

to available essential services facilities, along with variables representing storm parameters, utility outages, and socioeconomic variables. Through highly accurate random forest models estimating the state of a user's recovery on a given day, I show the importance of measures of access to essential services before and after an event for household and workplace recovery. Power, cell service, and school outages all rank highly in importance, followed by measures of the change in travel time to essential service facilities following Irma's landfall. Importantly, I demonstrate in these models that measures of access rank as more important for estimating recovery status than Social Vulnerability Index measures, which are currently incorporated into measures of community resilience. This may be attributed to access measures capturing both social and infrastructure drivers of recovery, as compared to the static values represented in index measures. Together, the methods described in Chapters 2-4 of this dissertation form a framework for assessing access to essential services before, during, and after a disruptive event that can inform interventions for facilitating recovery and building more resilient communities.

CHAPTER 1

Introduction

1.1 Motivation

Disasters are recurring in the United States and across the globe. Globalization, climate change, aging infrastructure, and economic activity clustered towards the coasts increase the occurrence of such disasters and intensify their impacts [1, 2, 3]. These disasters do not discriminate, but the paths to recovery do based on the geographic, social, environmental, and economic characteristics of a household or community [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. Vulnerabilities in these characteristics exacerbate the risks associated with natural and man-made hazards, with the most vulnerable communities presenting greater challenges throughout all phases of the disaster cycle: mitigation, preparation, response, and recovery [3, 9, 12, 13, 14, 15, 16]. Equity is the consideration of the extent to which costs, risks, and benefits are distributed fairly (or unfairly) across groups [17, 18]. Equitable outcomes may require different treatments and interventions across groups to compensate for pre-existing inequities within a system [11, 19]. Equity is inherently tied to risk, as the "weakest link" in a system defines the mitigative capacity of that system [12]. For a community facing a hazard, inequities carried over from times of stability may compound with those introduced by the hazard itself as well as recovery efforts that may reinforce previous inequities or introduce new ones through maladaptation [11, 12, 14, 20, 21, 22].

Resiliency planning offers the opportunity to transform this paradigm. Resilience represents the capacity of a system to absorb changes and return to a previous state of progress or adapt to a new normal [17, 20, 22, 23, 24, 25, 26, 27]. Community resilience includes infrastructure, people, cultural facilities, and anything else that enables a system and all within it to thrive [12, 16, 22, 27, 28]. Achieving community resilience must integrate technical and social approaches to fully evaluate and mitigate risk across the distribution community members, infrastructure, and institutions [10, 11]. The qualities that define a community are inherently specific and hazards are unpredictably unique which complicates risk analysis

to guide decision making in preparation for or in response to an event. Moreover, cities in particular face challenges associated with urbanization, climate migration, and housing shortages while public utilities and infrastructure must serve larger populations over smaller geographies; these factors in turn contribute to more people affected by failures occurring within a smaller space. But urban spaces can also offer increased access to critical resources such as food, healthcare, and economic opportunities in the form of density and transportation options besides private car ownership. Such access could be critical at all stages of disaster planning–preparation, response, and recovery– and thus can reflect and contribute to the resilience of a community [28]. Incorporating access into resiliency planning can also promote transformation, or a systemic change of an existing system, through mitigative and adaptive measures to avoid resiliency traps that often reinstitute inequities when communities strive to recover to an already un-resilient baseline [2, 20, 24, 28]. Quantifying the relationship between access and community resilience could contribute to motivating improvement in access at every stage of the hazard cycle for the households and communities who need it most.

At the same time, data representing human interactions with the built environment are more available than ever, including smartphone location data, open data portals from government agencies, and internet of things sensors [29, 30, 31, 32, 33, 34]. Advances in complex systems planning and machine learning methods enable intelligent and efficient use of such data, but require a transdisciplinary approach to operationalize for effective application. Through the methods described in this dissertation, I integrate tools spanning data science, operations research, urban planning, risk analysis, and civil engineering to utilize Location Based Services (LBS) data from cell phones to evaluate impacts of hazards on access to essential services and community recovery. With this research, I present methods that I hope can contribute to the design of more accessible and resilient communities.

1.2 Background

Location data generated by cell phones have revolutionized understanding of human mobility [35]. Historically, mobility analysis relied on expensive and labor-intensive surveys to provide snapshots of travel patterns [36, 37]. Using this method to understand changes in travel patterns in response to an event requires administration of surveys prior to and following the event, or for participants to adequately report their pre-event patterns in the post-event survey. In contrast, widespread ownership of mobile phones enables passive and ongoing collection of location data to supplement intensive household travel surveys with real-time data and larger sample sizes at relatively low cost. At this point, the utility of cell phone

traces versus travel surveys is well understood and cell phone location data are considered to be well suited for supplementing existing survey and mobility analysis methods [38, 39]. Now, LBS data are regularly used for commercial and research purposes in the development of travel demand models [40, 41].

In this dissertation, I consider resilience as the capacity of a community to absorb and recover from a disruption, or return to normal, including any adaptations or transformations that result in a new normal [17, 20, 23, 24, 25, 26]. When transportation networks are disrupted, LBS data can show changes in behaviors prior to and following the disruption that reveal adaptation and recovery previously only quantifiable by surveys. Changes in access impact the recovery, adaptation, and resilience of the affected community [28, 42]. Literature states that disruptions such as those from transit strikes, bridge closures, special events, and earthquakes result in behavioral changes in route choice, departure time, and travel mode [43]. This appraisal is consistent with analysis of major road closures due to the Northridge Earthquake in 1994 which also revealed that transportation network redundancy and the ability for travelers to choose between a variety of adjustments led to more rapid recovery [44]. In the New York City metropolitan area following superstorm Sandy, the probability of altering normal behavior showed the differences in adaptability to the transportation network disruption was based on transit dependence, number of children, and ability to telecommute [45]. These studies all suggest that access facilitated by transportation networks following a disruptive event can result in quicker adaptation and recovery.

Access to essential services such as power, food, childcare, and medical care is necessary for individuals to recover from disruptive events to adapt and eventually return to stability [27, 46, 47]. But these essential services are not available at the same standard across all members of a community. High mobility and access are linked to socioeconomic development including higher income and education rates as well as lower deprivation and unemployment rates [48]. Access extends beyond spatial proximity, but is also a function of attributes of service provision such as availability, affordability, accommodation, acceptability, and awareness [49, 50]. The ability to identify home, work, and other stay locations from LBS data offers new opportunities to evaluate access on the individual level by evaluating household proximity and travel times to essential services.

Previous assessment of access relies on metrics or distributions from a single point, like a centroid, to represent census tracts or counties [47]. This prohibits the ability to distinguish differences in access within those geographies, which may strengthen or weaken the observed relationships between socio-economic factors contributing to vulnerability. When individual-level access data are coupled with data representing the changing availability and affordability of those services, such as closures of public schools or of grocery stores ac-

cepting food assistance, these nuances of accessibility can be better quantitatively captured. Social vulnerability is also influenced by features of the built environment, so understanding accessibility more precisely allows simulation and scenario analysis of how interventions in that built environment might affect accessibility when exposed to various hazards [27, 46]. This can contribute to better quantification of resilience and the development of indices representing vulnerability as a function of exposure, sensitivity, and adaptive capacity [47].

1.3 LBS Data

Veraset LLC, a company specializing in gathering and consolidating cellular phone location data from mobile applications, generously provided the LBS data for this research effort [51]. Work with this data was reviewed by This study was reviewed an approved by the University of Michigan Institutional Review Board as study number HUM00143302 and was determined to be exempt and not regulated.

Each data point represents any interaction of a phone with a cellular network or Wi-Fi, including calls and data activities. Each point includes a de-identified user identification code, latitude, longitude, timestamp, accuracy measure, and the device type. Starting with 250 GB of user data spanning August 1, 2017 through October 3, 2017 across Florida, Georgia, Alabama, and South Carolina, I identified the home locations of over 1.5 million users, with 500,000 of those users also having identifiable workplaces. I applied additional filtering to identify households with both home and work places in Florida that were home at least 80% of dates in the study period and appeared at home within the week prior to Hurricane Irma's landfall, resulting in four weeks of observations of LBS data for 123,445 households across Florida.

Importantly, LBS data are collected opportunistically, so users may next ping seconds or days following their previous data point. So, LBS data does not capture a complete mobility profile of a user, but can reveal patterns when aggregated over time and geography. Surveys are still valuable for capturing daily activity diaries as well as risk perception, socioeconomic information, and other factors influencing decision making, but the scale of LBS data offers additional analysis beyond the scope of post-event household surveys. Following Hurricane Irma, even the largest scale surveys received fewer than 1000 responses to try to characterize state-wide behavior [52, 53], less than 1% of the number of users whose behaviors can be observed through this LBS data. Further, while surveys may be targeted toward underrepresented groups, long timelines and declining response rates among younger and difficult-to-reach populations means that surveys must trade-off between breadth and inclusion. The ubiquity of cell phones and availability of LBS data may free up resources for

surveys to capture more nuanced experiences for critical analysis of local disaster response. Other technologies, like social media posts and drone coverage, offer additional value for evaluating behavioral responses to disasters, but the geographic and temporal scale and specificity of LBS data offers unique opportunities for large-scale data-driven approaches for evaluating disaster response and recovery.

It is critical to also consider when LBS data are not useful. LBS data provides access to large volumes of user data over large geographies, but is not comparable to collecting highly localized data that describes the human experience of a specific community. Such local data remains critical to account for local context to evaluate and address specific risks in the preparation and response stages. Furthermore, LBS data's observational frequency varies vastly by user, so portions of mobility patterns may be missing at highly different rates across users. LBS data are not attached to any demographic information, and do not necessarily capture fully representative data, particularly over small geographic scales for older populations and minority groups [54]. Using LBS data to support decision making for policy development and resource allocation should proceed with caution to avoid disproportionately harming these populations. Finally, the use of LBS data necessitates a conversation about data privacy and mass surveillance. While the provided LBS are fully anonymized, it may be possible to determine sociodemographic and political groups of users, or even their identities [32, 55, 56]. Institutions and researchers are responsible for enacting and upholding policies and practices that protect privacy of individual users, including de-identification that prohibits re-identification, safe data sharing protocols, and only publishing results in aggregate [57, 58]. For this reason, I report and visualize only data aggregated over geographic space or time. The use of LBS data for research also contributes to the normalization of mass surveillance and should acknowledge the potential of such data to be used as means of control [59, 60, 61]. I understand surveillance as existing on a continuum between care and control [61, 62] and I hope that my contributions to data-driven planning within and beyond this dissertation always fall on the side of care. With these ethical dimensions in mind, I present these LBS-based methods only as tools to be used in collaboration with many other approaches for understanding how hazards impact communities.

1.4 Research Objectives

1.4.1 Using LBS Data to Evaluate In-access of Essential Services Following Hazards

In Chapter 2, I present a novel data-driven method for identifying facility closures following a natural hazard or other disruptive event. I describe location-based data and their utility for human mobility analysis, particularly for recognizing behavioral changes following disruptive events. A functional closure refers to a change in access to a facility, due to actual facility closure, power outage, transportation network disruption, or other barrier rendering a facility unavailable. This definition is inclusive of all reasons a facility might become inaccessible, such transportation barriers, perceived safety, and lack of supplies, thus capturing multiple dimensions of access.

I present an approach for segmenting LBS data based on desirable facility location footprints, effective analytical time periods, and users of interest to determine when a facility is functionally unavailable. I demonstrate the effectiveness of this method on location data from Southwest Florida in the weeks surrounding the landfall of Hurricane Irma in September, 2017 using samples of supermarket, primary school, secondary school, urgent care, and home improvement facilities in Collier County and Key West, Florida. I then demonstrate how anomaly detection methods may be used for automating the detection of functional closures through LBS data. Periods of disruption are visible in the resulting time series of individual facilities, but machine learning can more precisely determine periods of functional closures at across a much larger number of facilities (e.g., all grocery stores in a state). I also present the use anomaly detection methods to automatically detecting functional closure periods.

This chapter describes a method for detailed, facility-level closure detection that can easily scale to identifying facility closures over large regions for large-scale analysis of communities' responses to disruptions. Each step of identifying facilities and aggregating data over users and time periods can be easily adapted for any type of facility and user group of interest, for any type of disruption. Because LBS data is collected continuously, this method can also be replicated for comparative analysis between two communities facing the same disaster, or for comparing the same community's response to multiple disruptive events. More broadly, this chapter presents and promotes the use of location-based services data for hazards and resilience researchers for understanding relationships between human behavior and disruptive events.

1.4.2 Using LBS Data for Quantifying Large-scale Household-level Disaster Recovery

LBS data has the capacity to reveal frequently visited user locations, including households and workplaces. In Chapter 3, I developed a place-finding algorithm for identifying home and work based on standards consistent with relevant literature [40, 63]. Commuter behavior in particular follows regular patterns so that disruption is detectable and reflects economic and social disturbances. Return-to-productivity based on changes in home and work appearances can be used as a metric for assessing a community's recovery. Return-to-productivity is typically understood only after a disruption via surveys that capture how commuter behavior changed [45]. Using machine learning on LBS data, I detect behavioral changes from before, during, and after a disruption that can quantify return-to-productivity periods at household levels.

I design a Bayesian-network based anomaly detection method for both home and work appearances to detect disruptions in home and work appearance patterns in Florida following Hurricane Irma in 2017. This approach goes beyond merely tracking the return to homes; I quantify recovery by considering the restoration of normal visit behaviors both at home and work, thus expanding the definition of recovery to be more inclusive of the variety of human experiences after a disruptive event.

Results of this method include maps showing recovery patterns across all Florida counties as well as across census tracts for the two most impacted counties, Monroe and Collier. I also present recovery curves for each county and the empirical cumulative distribution functions for total estimated recovery duration for home and work. Finally, I validated results against a survey of evacuation rates during to Irma. This data-driven method can be utilized with LBS data spanning other disruptive events and geographies to identify spatial and temporal trends in household and workplace recovery, beyond the scope of typical household post-event surveys. Broader representation of the diverse experiences of recovery could reveal patterns by community or hazard type to deepen understanding of the social and infrastructure drivers of recovery.

1.4.3 Quantifying the Relationship between Access to Essential Services and Household Disaster Recovery

From Chapter 3, the length of anomalies serves as a proxy for household- level resilience based the amount of time it takes users to resume their normal patterns or return-to-productivity. I use this period as a dependent variable for a statistical model evaluating the influence of various access measures on resilience. I capture proximity by using routing tools such as

OpenTripPlanner to evaluate travel times by various modes between an individual and their workplace as well as the nearest supermarkets, healthcare facilities, schools, gas stations, and libraries. I incorporate availability using the facility closure detection method described in Chapter 2. I also add data on power and cell network outages, public school closures, and storm variables, including wind and storm surge. I also integrate census-tract-level data from the Social Vulnerability Index (SoVI), including household income, car ownership rates, and percentage of population insured. From this dataset, I construct random forest models to estimate recovery status of home and workplaces for each day in September 2017 to capture the month surrounding Hurricane Irma's landfall on September 10, 2017. The results of these models reveal quantified and interpretable importance rankings of the included variables, revealing power, cell service, and school outages all rank highly in importance, followed by measures of the change in travel time to essential service facilities following Irma's landfall. Measures of access rank as more important for estimating recovery status than Social Vulnerability Index measures, which are currently incorporated in many hazard recovery models. This work demonstrates the importance of including measures of access when estimating recovery and evaluating community resilience.

CHAPTER 2

Using LBS Data to Evaluate In-access of Essential Services Following Hazards

Natural hazards bring about changes in the access to essential services such as grocery stores, health care, schools, and day care because of facility closures, transportation system disruption, evacuation orders, power outages, and other barriers to access. Understanding changes in access to essential services following a disruption is critical to ensure equitable recovery and more resilient communities. However, past approaches to understanding facility closures and inaccessibility such as surveys and interviews are labor-intensive and of limited geographic scope. In this chapter I develop an approach to understanding facility-level inaccessibility across a broad geographic area based on location-based services (LBS) data collected from cell phones. This approach supplements current approaches and helps both researchers and emergency response planners better understand which communities lose access to essential services and for how long. I demonstrate my approach by analyzing loss of access to supermarkets, schools, health care facilities, and home improvement stores in Southwest Florida leading up to and following the landfall of Hurricane Irma in 2017.

Keywords: access; community resilience; location-based data; natural hazards

Note: The research presented in this chapter has been published in *Risk Analysis*. Citation: Swanson, T, & Guikema, S (2023). Using mobile phone data to evaluate access to essential services following natural hazards. Risk Analysis, 00, 1– 24. https://doi.org/10.1111/risa.14201

2.1 Introduction

Access to essential services such as food, healthcare, day care, and materials needed to restore damage to homes determines individuals' ability to meet health, safety, and social

needs that enable them to thrive in their daily lives and after a natural hazard [28, 64]. However, access relies on buildings and infrastructure networks that are vulnerable to significant damage following natural and human-caused hazards [65]. For example, damage from Hurricane Sandy reduced subway and bus service that particularly affected those reliant on public transportation [45]. Similarly, the COVID-19 pandemic induced reduced access to schools and childcare that led to potential loss of productivity and income for those with young children [66]. Access to essential services inherently relates to whether households and organizations can adequately prepare for a hazard before it strikes and then recover from them afterward. Hence, understanding impacts of the change in access and inequities in access is critical for hazard and resilience management and should be a significant consideration for households' and communities' capacity to prepare for, mitigate, and recovery from disruptive events. Quantifying how such access varies over populations, space, time, and in response to events is necessary for evaluating and improving equity of community resilience.

Such quantification is especially difficult in the periods surrounding hazardous events as essential service providers are hard to reach at scale and often do not necessarily record their recovery status for future evaluation. While knowing when essential services facilities were closed or inaccessible (due to transportation barriers, for example) is critical for planning emergency response and recovery activities, gathering this information has been problematic, especially for commercially-provided services such as healthcare, food, child care, and home restoration supplies. The closure status of these facilities is not typically available in an aggregated manner and is perishable data. Unless collected in real time (through, for example, calling individual facilities or scraping facility web pages) information on which facilities were closed and for how long is not available for future research or emergency response planning. In addition, facility closure data does not capture broader issues with accessibility of the facility that may be caused by disruptions to other services or infrastructure, such as transportation network closures. While accessibility data for these types of facilities may be collected through surveys or manually calling each facility, these approaches do not scale well to capture both neighborhood-level and state-level analysis and, are subject to recall bias when conducted after an event. More robust and scalable methods are needed for assessing when facilities are functional and accessible to support both hazard research and improved community resilience planning.

Opportunistically collected location-based services (LBS) data available from cell phones offer new opportunities to evaluate accessibility of essential services. LBS data consists of cell phone locations over time that are captured opportunistically from apps on smartphones for which the user has opted-in to having their location recorded. These location "pings" are collected and anonymized by data aggregators who then provide the anonymized locations

or aggregated access to specific locations as a commercial product, thought was provided for this research in its unaggregated but anonymized form. It is important to note that this data are (1) available only in anonymized form and (2) is not available commercially in real-time. A typical use case for this data is to help a company understand the volume and temporal patterns of foot traffic at a location at which they are considering opening a facility such as a restaurant. In this chapter I develop a way to use this data for a different purpose: to understand the temporal patterns of access to critical services at a facility level across a broad geographic scale without relying exclusively on manually contacting each facility or post-event surveys.

Over time and at different levels of aggregation, LBS data reveals visit patterns at specific facilities. I present a novel method to evaluate deviations from pre-event visit patterns to detect substantial changes in access to critical facilities during the periods surrounding a hazard event. I term this a functional closure. This could occur because the facility has declared itself to be closed (e.g., due to direct damage, lack of employees, or loss of infrastructure services such as water or power), because it is inaccessible due to disruption to transportation system, or because few are trying to access the facility because the area has been evacuated. My approach allows the accessibility of essential services to be estimated at a temporal and spatial specificity that has not been possible with previous approaches, at least at a broad geographic scope (e.g., county-wide or state-wide). This provides a significant advance for enabling detailed, facility-level estimates of functional closures that can scale to entire states and regions for broader analysis for evaluating disaster response and recovery. LBS-based approaches provide substantially more detailed information on the locations of anonymized individuals over much broader samples of the population than surveys do, but do not allow one to ask other critical questions about motivations or reasons for movements. Surveys and focus groups remain indispensable for understanding the underlying causal reasons for changes in mobility and access. I present this approach to supplement more traditional methods such as surveys.

2.2 Background

Access is the degree of fit between a system and the users of that system [49]. Specifically, access consists of the interdependent dimensions of proximity, availability, acceptability, affordability, adequacy, and more recently included, awareness [49, 50]. This access concept historically features in medical contexts, but access literature also includes examples of evaluating access to nutritious food, green space, and first responders [33, 67, 68]. Access to essential services, including healthcare, childcare, food, and cultural amenities are necessary

for individuals to recover and adapt to disruptive events [28, 64, 69]. COVID-19 stay at home orders, flooding of New York's subway system, and large-scale power outages following winter storms in Texas all exemplify the potential for drastic and unexpected changes in access to essential services that result in economic, physical and mental health, and social consequences [45, 66]. However, the relationship between access and resilience is not easy to quantify and analyze [28]. Identifying and collecting data capturing the dimensions of access across any one service is complex even under every-day, non-hazard circumstances. For just one facility, accounting for multiple dimensions of access includes (but is not limited to) investigating whether the facility is operating, is safe, has the necessary goods and services in stock at an affordable price, is open appropriate hours, and is reachable in a reasonable amount of time through usable transportation modes [49]. Furthermore, a household's ability to recover from a disruptive event typically involves access to a combination of multiple facilities that provide essential services [64], and so conceptualizing community resilience as related to access to services requires evaluating all such facilities within that community, presenting data collection and scalability difficulties.

Location data generated by cell phones present an opportunity to overcome these data collection and scaling difficulties to identify when facilities are and are not accessible. Since becoming available to researchers, LBS data have transformed the understanding of human mobility and travel behavior [35, 70]. LBS data has already been demonstrated to reveal how users interact with their built environment consistent with prior data collection methods. Historically, mobility analysis relied on expensive and labor-intensive surveys to provide snapshots of travel patterns [36, 37, 70]. For understanding disaster response, surveys must be administered prior to and following the event, or for participants to adequately report their pre-event patterns in retrospect. Now, widespread ownership of mobile phones enables passive and ongoing collection of large sample sizes and relatively lower costs of mobility data to supplement intensive household travel surveys [71].

Several studies have compared the utility of cell phone traces versus travel surveys and revealed the promising advancement of cell phone location data for supplementing existing survey and mobility analysis methods [38]. An abundance of commercial and research applications rely on LBS to identify stay locations such as home and work, create origin-destination matrices and even infer trip purposes to feed transportation forecasting models [40, 41, 70, 72]. As smartphone app-based data collection increases in proliferation and improves in accuracy, researchers are increasingly able to access LBS data through licenses with providers for applications from transportation to pandemic epidemiology [73, 74, 75].

Beyond conventional transportation planning, LBS data has contributed to evaluating and simulating evacuation scenarios to understand population-level response to natural hazards

[76, 77]. Understanding responses to hazards is crucial to understanding resilience in terms of recovery and adaptation [23, 25, 26]. Whether a facility is open or closed is critical for measuring the availability component of access, but current hazards research methods rely on awaiting and aggregating FEMA reports, resident surveys, and individual-facility-based investigations [70, 76]. These traditional methods are laborious, time-consuming, and do not always provide detailed, facility-level information with fine temporal resolution for all relevant facilities, rendering them impractical for some research and policy questions seeking to evaluate all individual facilities across a large geographic region.

Other technologies like social media sources or drones are also becoming more utilized for hazard response [30, 31, 78], but are limited in their ability to capture the same relationships between people and their built environment as LBS and so are not sensible for assessing access to essential services. For example, Twitter lacks the same geographic specificity as LBS data and geographically enabled tweets have a much lower permeation rate across populations and so are not representative for smaller geographies necessary to evaluate individual essential services facilities [79]. Aerial footage and other data available from drones is useful for risk assessment and identifying damage, but are limited by coverage areas and endurance for larger scale disasters like hurricanes or earthquakes, do not capture human behavior over time, and present interference issues with other aircraft involved with response [80]. When infrastructure networks and access to essential facilities are disrupted, LBS data can enable a large-scale data-driven approach to evaluate changes in human behaviors prior to and following a disruption beyond just population level evacuation movements without formidable data collection efforts to capture household and facility-level trends across an entire county or state [59, 81]. While LBS data are anonymous, users' frequent locations can be aggregated and reconciled with census and travel survey data [40, 73]. These methods can enable evaluation of differences in disaster response and recovery behaviors by geographies as granular as census tracts. Understanding these differences may reveal systemic gaps in recovery and contribute to understanding of interdependencies between access to physical facilities, infrastructure, and social vulnerability.

Changes in behavior detectable through LBS can indicate if a facility is closed, but also if that facility is rendered inaccessible for some other reason such as a lack of inventory, high prices, or transportation system disruptions. These behavioral changes can indicate changes in some of the dimensions of access without having to query individual facilities about their operational status and service availability over time. For selected communities and facilities, methods for detecting behavioral anomalies can be automated and scaled to observe community-, county-, or even state- level patterns in accessibility and recovery [76]. Anomaly detection methods for time series data are common in many systems to identify unexpected

observations or sequences of observations within time series data [82]. Anomaly detection can identify changes in regular patterns, such as those captured by sensors monitoring systems from an individual's personal health to quality control in manufacturing to the air quality of an entire community [82]. Simple models can identify anomalous behavior when new data falls outside of the range of data seen before. Machine learning models ranging in complexity from linear regression to neural networks can identify anomalous points from univariate or multivariate data when a new observation is too far off from what the model predicted [83, 84, 85]. I show how coupling anomaly detection methods with LBS data can facilitate the large-scale data-driven identification of anomalous periods experienced by essential services facilities during and following a disruptive event.

2.3 An LBS Data Approach to Identifying Change in Access

I present the following method for addressing the challenge of identifying functional closures resulting from a hazardous event. For a given facility in a given time period, a large change in the number of unique user appearances at that facility likely indicates some change in facility policies or accessibility. By counting unique user appearances at a facility, raegular visit patterns emerge. Identifying deviations from those patterns of user appearances enables the detection of changes in facility policies or access.

2.3.1 Data

LBS data for this analysis was provided by Veraset LLC, a company that collects and aggregates cell phone location data from mobile phone applications. Each data point, or ping, represents a phone's interaction with a cellular network or Wi-Fi. These interactions include intentional user activities as well as passively collected data via background processes in smartphone applications. The data for each interaction includes an anonymized user identifier, timestamp, latitude, longitude, the geohash code, the horizontal accuracy (a radius of error around the location), and the device type (Android versus iOS). Table 2.1 illustrates a synthetic version of the LBS data:

User	Latitude	Longitude	Timestamp	Accuracy (m)	Device type
ax295	28.385390220048	81.564435515210	1501596000	50	iOS
ax295	28.123456789878	81.564435515225	1501597200	50	iOS
ax295	27.645848484243	81.564435515228	1501597500	50	iOS
...	...	...	...	...	...
xx874	24.999999999999	82.128485517854	1501596120	60	android
yz101	21.248486542137	84.551521048888	1501597020	15	iOS
yz101	20.918438111351	84.762632159876	1501597035	12	iOS

Table 2.1: Synthetic LBS data

Critical data for this analysis includes only phone identifiers and timestamped locations, consistent with other mobile phone-based data-sets. It is important to note that pings are not necessarily collected at regular time intervals– a single user's inter-ping times could range from seconds to hours depending on the active versus background applications on the individual phone. Further, horizontal accuracy ranges from 2 meters to over 2000 meters (75% of points are under 30 meters of error, 95% are under 110 meters). With these limitations in mind, LBS data does not capture a complete or precise profile of an individual's mobility or all locations visited. For this reason, I consider patterns within the data when aggregated by users, time periods, and geography. How the data are aggregated is the subject of the remainder of this chapter.

2.3.2 General Approach

My approach consists of the key steps shown in Figure 2.1, each of which will be described in more detail with examples from a case study below:

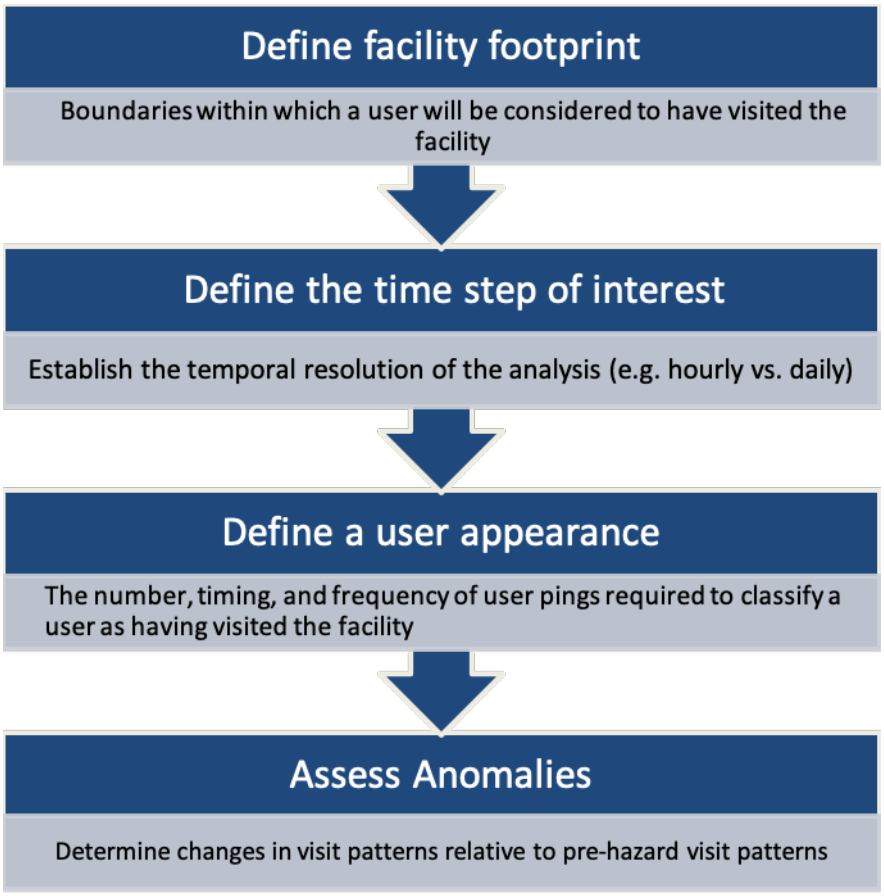


Figure 2.1: Approach for identifying functional closures from LBS data

2.4 Case Study: Hurricane Irma in Southwest Florida

Hurricane Irma made landfall in the Florida Keys as a category 4 hurricane on the morning of Sunday, September 10, 2017. Irma subsequently travelled up the Gulf coast of Florida making landfall on the mainland that evening as a category 3 hurricane and travelling into northern Florida into the morning of Monday, September 11, 2017 [86].

Figure 2.2 below (from [87]) shows the corresponding time of landfall for the Keys early Sunday morning and Collier County by Sunday evening and indicates both regions sustained wind speeds over 110 miles per hour.

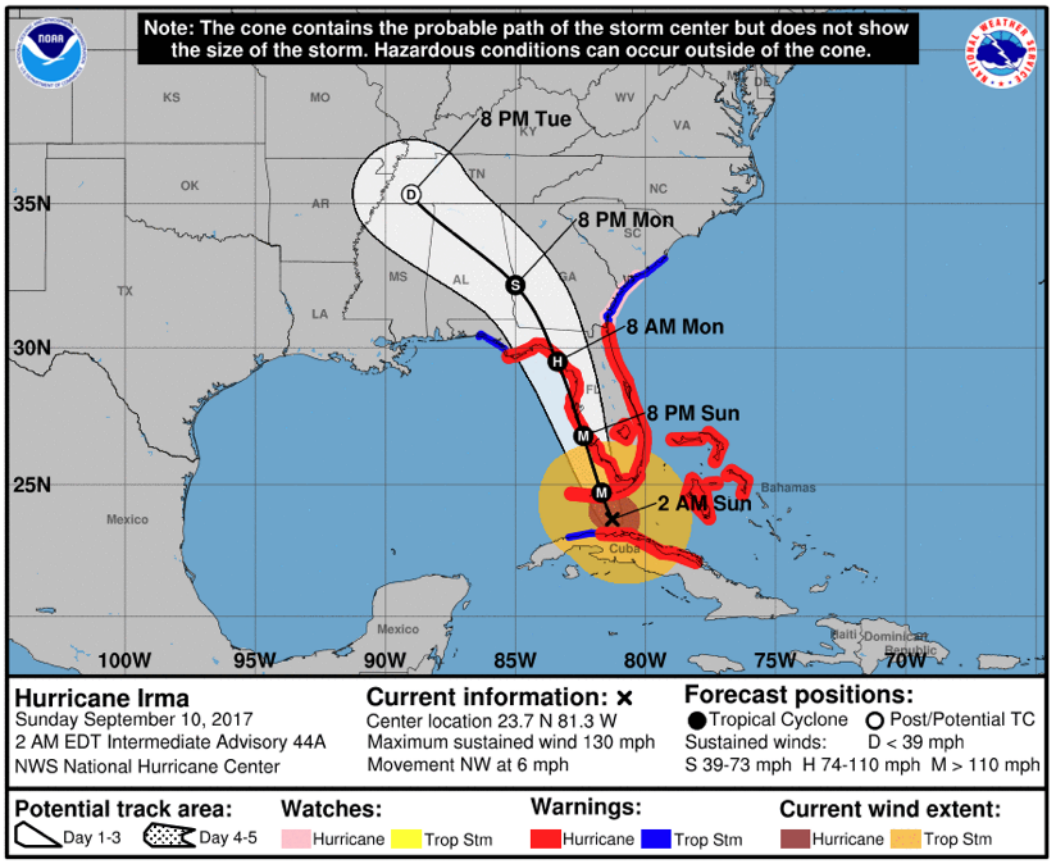


Figure 2.2: 3-day forecast track, initial wind field and watch/warning graphic from NOAA National Hurricane Center IRMA Graphics Archive

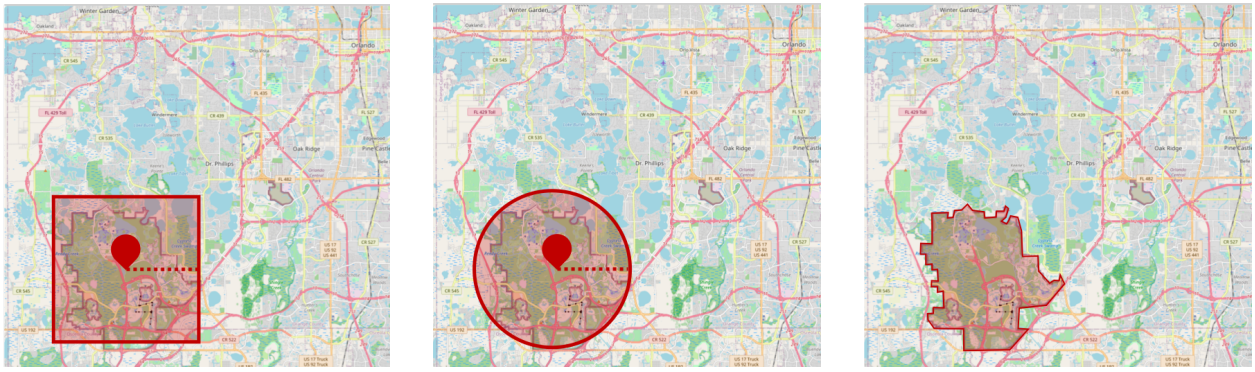
Weather forecasts up to five-days prior to landfall consistently predicted Irma's threat to Southern Florida, allowing time for evacuation planning and stocking up on supplies that ultimately resulted in grocery stores running out of food and water while gas stations ran out of fuel up to two days before landfall [86, 88]. Over 6.5 million people across South Florida were ordered to evacuate, with evacuation orders in place on September 6 for the Florida Keys and September 8 and 9 for Collier County [89, 90].

Hurricane Irma resulted in 7 direct and 77 indirect known casualties in the state of Florida and an estimated 50 billion USD in damages [86]. FEMA estimated about 90% of houses in the Keys sustained damage, with 25% of buildings totally destroyed [86]. In Southeast Florida over 90% of customers lost power for over a week [88]. Collier County Public Schools were closed to students September 7 and reopened September 25 [91] while Key West schools were closed September 6 and reopened September 27 [92]. I use Walt Disney World Resort in Orlando, Florida to highlight my methodology as Hurricane Irma forced the theme parks to close for only the sixth time in its history, from September 10 to 11, 2017 (though the

hotels remained operational) [93].

2.4.1 Define Facility

"Facilities" may range from units such as storefronts or schools to larger geographical entities such as shopping centers or commercial business districts. Following identifying a facility of interest, I next define the geographic footprint of that facility. A simple method for defining a footprint, shown in Figure 2.3a using Disney World Orlando as an example, is to set a degree- based grid around a facility and filter out cell phone pings outside of the latitudinal and longitudinal range of that grid boundary. A constant distance from the centroid of a facility could also define a facility, as in Figure 2.3b, where all pings within a desired distance are included. These methods are simpler, requiring only one latitude and longitude as input (the location of the marker in Figures 2.3a and 2.3b) along with a distance parameter (the length of the dashed line in Figures 2.3a and 2.3b) and so may be suitable where rough estimates of user throughput are enough to show a pattern change. For example, for a big-box store like Walmart or Lowe's with a large parking lot, demand could be captured well-enough by a large box or circle footprint, as large changes in demand are easily discernible even if the footprint leaves out some corners of the store or includes part of a road.



(a) Grid based method (b) Radius method (c) Facility footprint method

Figure 2.3: Facility definition for Walt Disney World Resort, Orlando, FL with OSM Standard base map

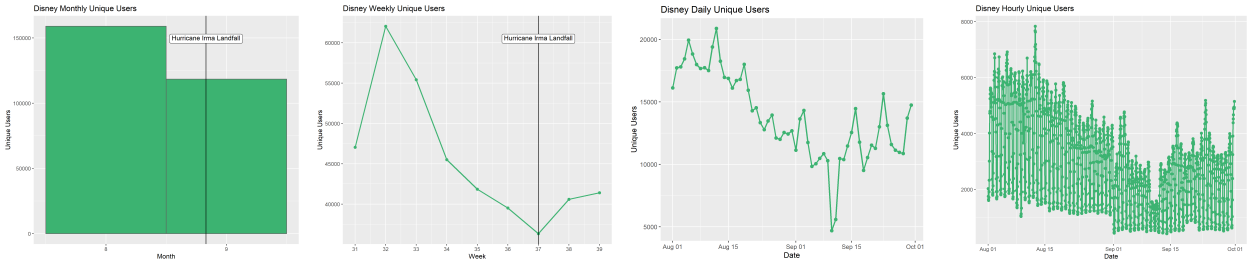
However, the grid-based or constant-distance methods may not suit denser communities or facilities in close proximity to highly-trafficked transportation routes, as properties are often not defined by rectangles or circles and imprecision may risk including or excluding too many points to detect changes in visit patterns to that facility. Consider an elementary

school on a snow day next to a park with a popular sledding hill. If the grid- or radius-based outline designed to capture school visits includes too much of the park, the decline in school visits may be offset by the increase in park visits such that the disruption to school visits is undetectable. When precision is necessary for the intended analysis, cartographic resources such as OpenStreetMap (OSM) [94] or satellite imagery can provide APIs to extract exact facility footprints, exhibited in Figure 2.3c. Generating and storing the precise facility footprint is more computationally expensive and so may not be necessary when rough estimates will do.

2.4.2 Define Time Steps

The next step involves specifying the time step of interest, i.e., the temporal resolution. All activity within a time step aggregates to a single data point. That is, the data will be aggregated to a count of unique visitors within each time step. The time step must be substantial enough to include a sufficient number of user appearances, but small enough to capture time-variant nuances in human behavior. Thus, time periods largely depend on the application of interest. For example, a food bank or transit station may exhibit anomalies at the hourly scale that would not be apparent at a daily scale, whereas students and teachers spend similar hours at a school every day, so investigating behaviors at a daily time step may be sufficient.

To show the impact of this parameter choice, Figure 2.4 shows the sum of all LBS points collected within Disney World Orlando (using a 0.07 degrees-wide grid-based facility footprint) at 2.4a monthly, 2.4b weekly, 2.4c daily, and 2.4d hourly intervals over August through September 2017.



(a) Monthly users (b) Weekly users (c) Daily users (d) Hourly users

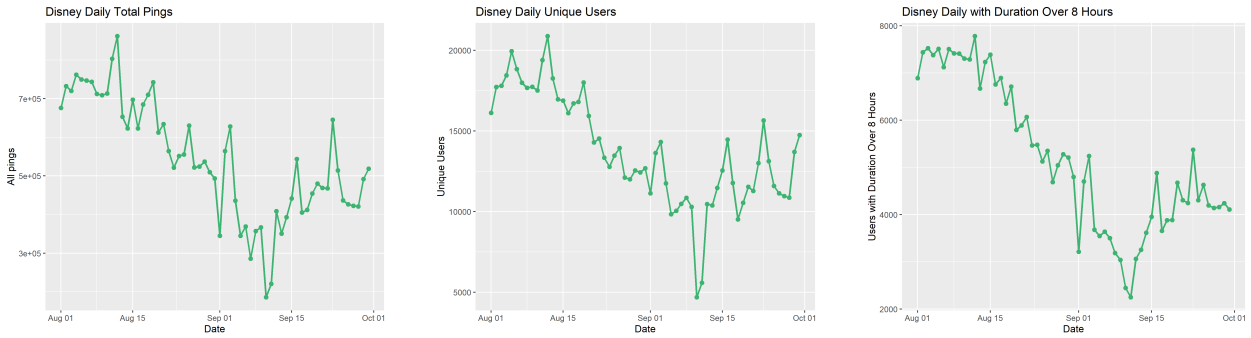
Figure 2.4: Comparing time period aggregation for visits to Walt Disney World Resort, Orlando, FL

Figure 2.4 shows that the LBS data aggregated at the monthly and weekly levels show a decline in appearances between August and September, but the closure period aligning

with landfall of Hurricane Irma is more distinct when the data are aggregated at the daily or hourly level. The hourly data are much noisier with daily patterns showing more pings during daytime operational hours and fewer pings at night as many users leave the park while a fraction stay in on-site hotels. This noise may cloud the appearance of a noticeable closure on September 10 and 11, 2017. These plots together showcase how the results of this analysis may vary with the time- period parameter choice, thus highlight the need to consider the context of the hazard event and the facility use for determining this parameter.

2.4.3 Define User Appearance Criteria

The third step is to define what is required to classify a user as having visited a facility. An appearance may include a unique user entering the boundaries of the facility at least once during a specified time period. Stricter policies may require multiple pings from individual users within in a time period, perhaps to eliminate pedestrians or car traffic that crosses through a facility footprint. Multiple pings from a unique user in a single facility can estimate a length of time a user stays in a facility, so minimum length of stay could also define a unique user's appearance. There is a trade-off between two types of errors here. On the one hand, setting too loose of criteria would tend to include passersby who did not actually spend time at the facility. On the other hand, setting too strict of criteria would exclude users who did visit but whose phones only pinged occasionally while there. Figure 2.5 shows visits to Disney World in August and September 2017, aggregated to include 2.5a all pings, 2.5b only unique users, 2.5c only unique users with a detectable duration over 8 hours.



(a) Daily pings (b) Daily unique users (c) Daily 8+ hours users

Figure 2.5: Comparing user appearance criteria for identifying visits to Walt Disney World Resort, Orlando, FL

For the most part, the time series at these different levels of aggregation show similar patterns, though at different scales. When aggregated to include all points as in 5a, some

users with high ping frequencies are over-represented. Further, during a disaster period, users may actively use their devices or the apps that collect their location data more frequently, resulting in a smaller proportional decrease in all points as seen in 5a compared to the decline visible when looking at unique users visits in 5b. When imposing a stricter constraint of users who spent over 8 hours a day in the facility as in 5c, I capture fewer visitors making day trips to the park while still including all visitors who stayed in resort hotels. As resort hotels remained open during Hurricane Irma, imposing this stricter criterion fails to show the full extent of the disruption from the resort closing its attractions.

2.4.4 Assess Anomalies

The final step is to assess whether or not there has been a substantial deviation in visit patterns. This can be done qualitatively and graphically (e.g., by examining graphs of visits over time) or through formal statistical anomaly or changepoint detection algorithms. In some cases qualitative methods are sufficient, for example just from looking at the daily unique visits to Disney World Orlando in Figure 2.5, one may reasonably identify the functional closure period on September 10 and 11, 2017. However, if there are many facilities (e.g., examining changes in access to all grocery stores in Florida) then automated anomaly detection algorithms can be utilized.

While assessing all anomaly detection methods that could be applied to time series generated from LBS data are out of the scope of this chapter, I walk through the process of selecting an anomaly detection algorithm for the Disney World Orlando daily unique visits time series. Given there are only 64 days in the data-set, it is not suited for a deep learning algorithm like a neural network. Further, I am examining the data retrospectively and as a complete set—that is I am not identifying incoming observations as they stream in. This means I can utilize an "offline" method that utilizes all of the available observations, including those after the anomaly, to identify an anomalous period. Offline methods contrast with "online" methods that follow observations chronologically to identify anomalies, thus only incorporating the observations prior to the anomalous period. Finally, I am only looking at one variable—the number of unique daily visits—and so should choose a univariate anomaly detection method. If additional data was available for the same location at the same temporal scale, for example weather or the previous year's visit totals, one may incorporate that additional data into multivariate methods. To quantify uncertainty around the start and end points of the recovery periods, some anomaly detection methods incorporate uncertainty quantification as a measure of how similar the anomalous data are to that previously seen [95, 96]. Uncertainty about the anomalous period could also be quantified by applying dif-

ferent anomaly detection methods to the same data to identify how consistently dates in the study period are identified as anomalous.

Binary segmentation is a popular approximate method that involves detecting a change in a specified statistic (such as mean or variance) within a time series and splitting the series around that point, repeating for a defined number of points or to maximize an objective with some penalty for each additional changepoint [97]. As this data set is only 64 days, I utilize an exact method called PELT [98] available through the 'changepoint' R package [99] to determine changepoints based on a Poisson distribution, as the unique user appearances represent discrete counts and so would expect changepoints to represent a statistically significant change in both mean and variance. This dynamic programming method optimizes the maximum log-likelihood for Poisson distributions fit to the data separated by each possible changepoint minus a specified penalty value to avoid overfitting (in this case, I used the default setting of the modified Bayes information criterion, or MBIC).

Figure 2.6 shows changes in Disney Resorts (normalized) appearances over the study period along with the results of the PELT changepoint detection method. This closure period is consistent with the theme parks closing September 10 and 11.

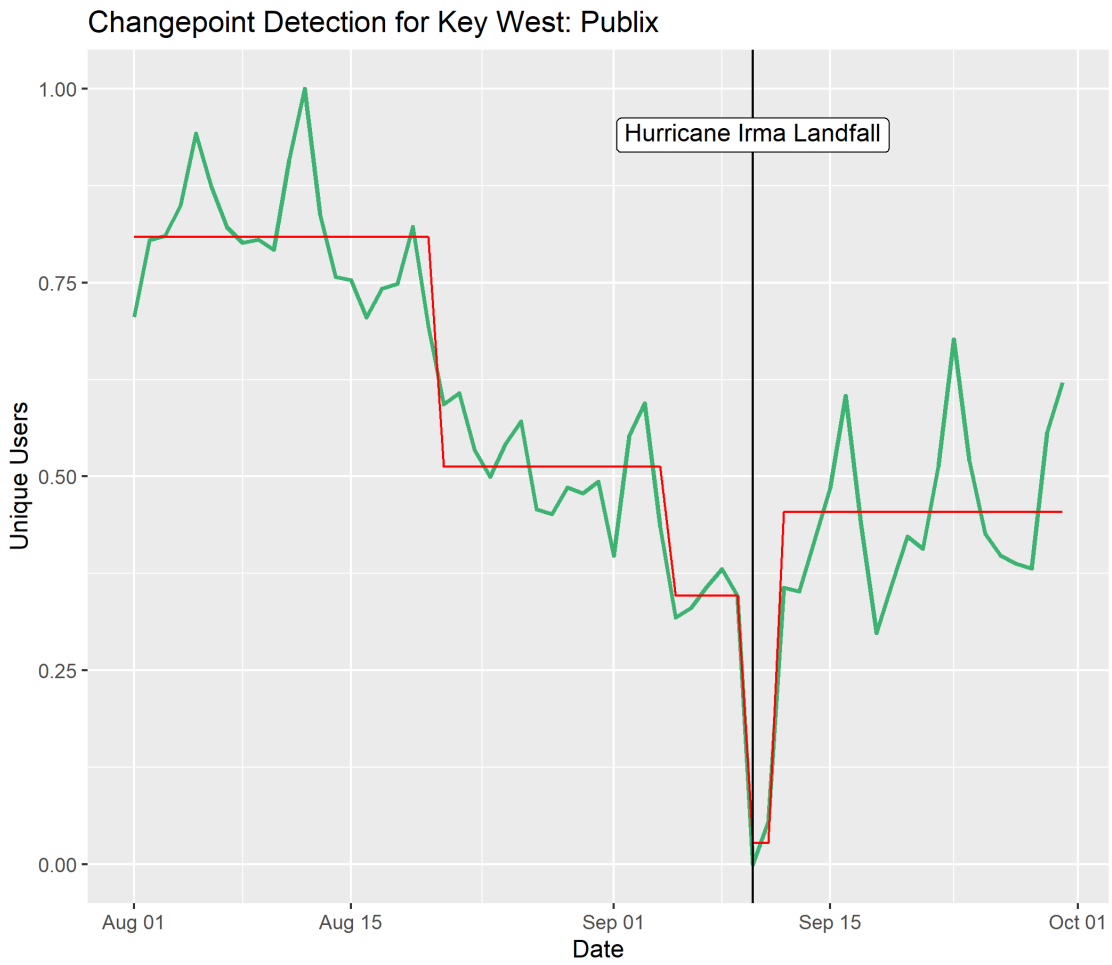


Figure 2.6: Daily unique user appearances for Walt Disney World Resorts in Orlando, Florida showing functional closure impacts of Hurricane Irma August 1, 2017- October 3, 2017

2.5 Functional Closures of Essential Services

I apply this method to a number of other facilities to further demonstrate the approach and its utility for evaluating access to essential services following a hazard. I identified a set of geographically distinct facility locations including supermarkets, primary schools (grades K-5), secondary schools (grades 9-12), urgent care facilities, and home improvement stores around Naples, Marco Island, and Key West. I selected these locations based on their proximity to the path of Hurricane Irma and their vulnerability to prolonged infrastructure damage due to wind and storm surges. Supermarket locations come from the USDA Food and Nutrition Service Supplemental Nutrition Assistance Program (SNAP) list to identify locations meeting the USDA's staple food requirements [100]. I identified primary and secondary schools based on Collier County Public Schools in Naples and Marco Island and

Monroe County Public Schools as well as Charter Schools in Key West. I identified urgent care facilities from a 2014 data-set from the Florida Department of Health, and confirmed they were still in operation as of January 2021 [101]. I included Lowes and Home Depot stores around the selected cities to represent area home improvement facilities as these companies are known to facilitate emergency response and residential restoration following disasters. All selected locations and sources are listed and mapped by category in Appendix A.

As I conducted this analysis for demonstrative purposes, I used the simple grid-based facility specification method (as shown in Figure 2.3a). For each location in Southwest Florida on each day from August 1, 2017 through October 3, 2017, I identify all unique users passing within a .0005 degree-wide (approximately 50 meters) grid-based geographic area around the centroids of the facilities. I define appearances as any unique user pinging at least once in the facility grid and I aggregate the data to generate a daily time series for all facilities other than the schools, for which I generate time series only from weekdays.

2.5.1 Sample Facility Results

The following plots in Figure 2.7 show changes in appearances at select facilities before, during, and after Hurricane Irma's landfall on September 10, 2017.

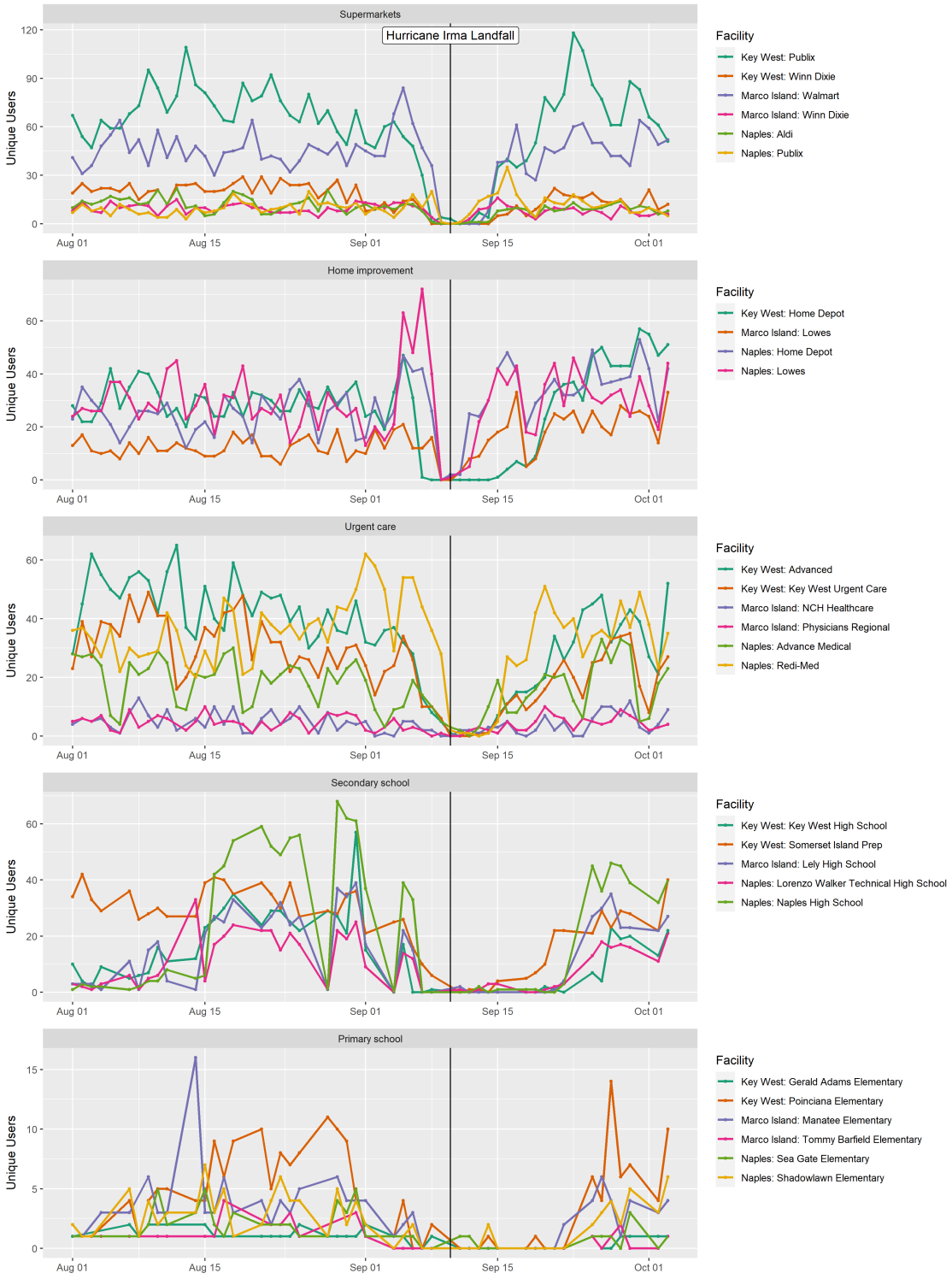


Figure 2.7: Daily unique user appearances at select facilities showing functional closure impacts of Hurricane Irma August 1, 2017- October 3, 2017

The uppermost plot shows supermarket appearances with an approximate functional clo-

sure period coinciding with evacuation periods and hurricane landfall. Interestingly, several of the facilities indicate upticks in appearances immediately prior to and/or immediately following the period of closure. This observation aligns with expectations of increased sales in the days immediately prior to and following a hazard event as households stock up on food and emergency supplies in the days before a hazard threat and/or purchase or re-stock those supplies immediately following landfall [102].

The second plot in Figure 2.7 shows that Lowes and Home Depot stores experienced increases in appearances in the days preceding and immediately following Hurricane Irma landfall, again consistent with expectations representing stock-up and restock behaviors. Based on Figure 2.7, the Key West Home Depot in particular seems to suffer from a longer functional closure period, but also experiences a prolonged increase in appearances through the end of the study period in early October. This closure could reflect physical damage or power outage at the facility, but could also reflect a lack of needed supplies due to disruption to the supply chain. This reflects the prolonged closure and reconstruction period faced by the Florida Keys as the Keys suffered significantly more damage and are more isolated compared to the rest of South Florida [86]. The extended increase in visits to the Key West Home Depot may also reflect visits by the same users over multiple days as merchandise came back in stock, which could be further investigated through this method by defining user appearance criteria (as described in 2.4.3) as users appearing in the facility more than once in the days following landfall. Urgent Care facility appearance activity at the Key West and Naples facilities is also consistent with Hurricane Irma landfall dates, while the Marco Island facilities do not reveal enough demand in the pre-hazard period to identify deviance from the norm following Irma's landfall. The declines in appearances at both Key West facilities relatively earlier than the declines visible in the Naples facilities' appearances align with the earlier evacuation orders in Key West.

The final two plots in Figure 2.7 show appearances at primary and secondary schools on weekdays. The upward trends from the beginning of August align with a ramp up to school starting for teachers on August 9 and for students August 16. A steep drop in appearances on August 28 coincides with Collier County school closures on August 28 due to flooding followed 22 by another low-appearance day coinciding with Labor Day on September 4, 2017. The low sample of appearances in primary schools follows expectations of low cell phone ownership and school-day operation rates for that age group; the unique users who do appear may include pedestrians due to the loose definitions of footprints and user appearances used for this particular analysis. Thus, there is not enough data to properly evaluate primary school facilities. The larger magnitude of high school appearances shows the functional closure periods more reliably than primary schools. Key West High School

enters a functional closure period before other schools and reopens after the other school closures, consistent with the closure dates of the various school districts. Somerset Island Preparatory High School appears to reopen earlier than all schools, indicating a potential private school policy differing from the county school systems.

Interestingly, Lely High School was opened as a shelter for Collier County [90] and Key West High Schools as a shelter for Key West [103]. As shown in Figure 2.8 below, these facilities show anomalously high appearances during the weekend days September 9 and 10. This result highlights the utility of this LBS data-based approach to identifying other anomalous facility appearance behaviors that characterize hazard response beyond facility closures. This example also demonstrates the criticality of thoughtful parameter selection as these appearances only appear anomalous with the context of expected school appearances on weekends versus weekdays.

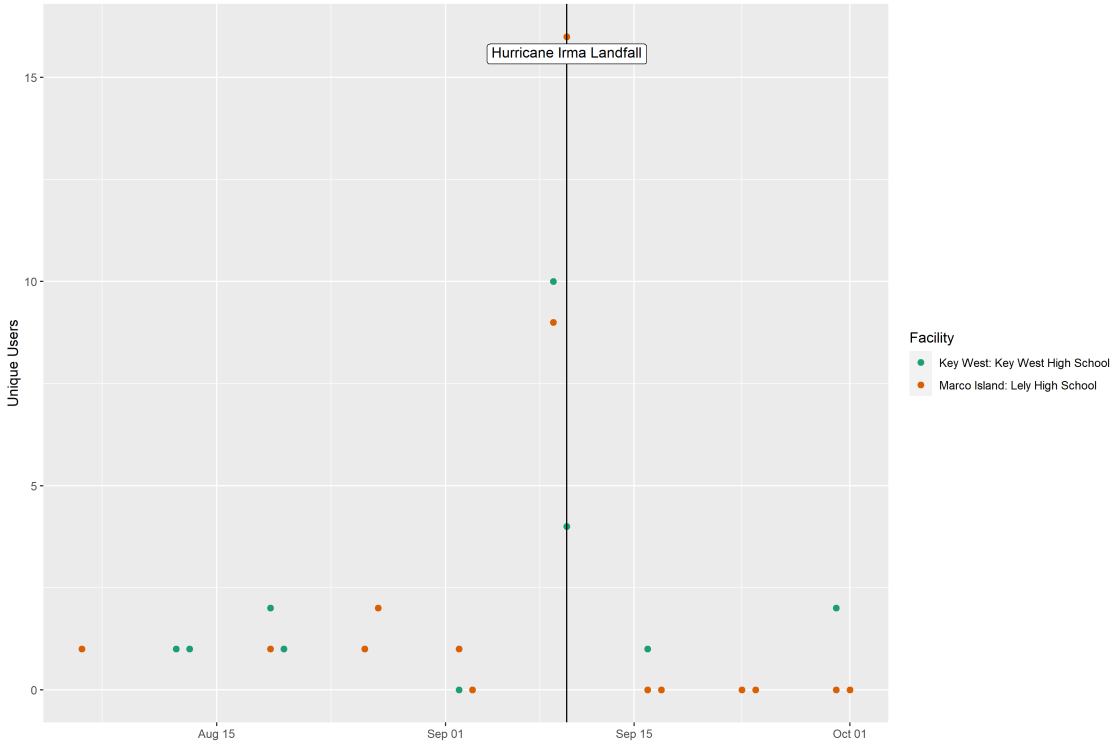


Figure 2.8: Weekend appearances at Key West and Lely high schools showing anomalous use as shelter facilities

2.5.2 Anomaly Detection

While from the above figures I can see the start dates, end dates, and periods of disruption are visible through plots on an individual facility basis, statistical methods can automatically

determine periods of functional closures at a much larger scale. Figure 2.9 shows the PELT changepoint detection method applied to the same data-set for the Walmart near Marco Island and the Publix facility and Key West, with the red lines indicating subseries delineated by each changepoint.

Figure 2.9: Examples of machine learning applied to identify change in access periods

These example results show a number of changepoints in the data that align with changes in user behavior, including a functional closure period. At the Walmart facility, a functional closure is detected on September 7, 2017 with reopening indicated on September 14, 2017, consistent with Irma's landfall on September 10. For Key West's Publix, this method detects a decline in activity on August 28, followed by a functional closure September 6, 2017, a slow return on September 14, and finally a return to normal on September 19. These example results align with the earlier evacuation period of the Keys and the greater damage sustained there [86] that would result in longer facility closures. These methods can easily scale to include all facilities of interest with any number of changepoints to detect the periods of functional closure as well as any periods of ramp up or decline in user appearances that may indicate change in access by assessing the mean value of each segment. I show the closure durations detected by applying the PELT method for each facility in Table 2.2, further summarized by location in Table 2.3 and facility type in Table 2.4. I share the PELT plots and additional closure period details of all selected essentials services facilities in Appendix

A.

Location	Facility	Category	Duration (days)
Key West	Publix	Supermarket	7
Key West	Winn Dixie	Supermarket	7
Marco Island	Walmart	Supermarket	6
Marco Island	Winn Dixie	Supermarket	3
Naples	Aldi	Supermarket	6
Naples	Publix	Supermarket	4
Key West	Key West High School	Secondary School	21
Key West	Somerset Island Prep	Secondary School	14
Marco Island	Lely High School	Secondary School	18
Naples	Naples High School	Secondary School	18
Naples	Lorenzo Walker Technical	Secondary School	18
Key West	Advanced	Urgent Care	8
Key West	Key West Urgent Care Inc.	Urgent Care	8
Marco Island	Physicians Regional	Urgent Care	NA*
Marco Island	NCH Healthcare	Urgent Care	NA*
Naples	Redi-Med	Urgent Care	5
Naples	Advance Medical	Urgent Care	6
Key West	Home Depot	Home Improvement	9
Marco Island	Lowes	Home Improvement	2
Naples	Home Depot	Home Improvement	3
Naples	Lowes	Home Improvement	4

*not enough unique users in sample

Table 2.2: Sample facility recovery duration

Location	Average Duration (days)
Key West	10.6
Marco Island	7.25
Naples	8

Table 2.3: Facility functional closure duration by location

Category	Average Duration (days)
Supermarket	5.5
Secondary school	17.8
Urgent Care	6.75
Home Improvement	4.5

Table 2.4: Facility functional closure duration by category

While this sample of facilities are small and not representative of all facilities in a given location or of a given type, the results in Table 2.3 are consistent with Key West bearing more damage during Hurricane Irma, thus leading to longer recovery. For facility types, Table 2.4 reveals access to these select supermarkets and home improvement stores seems to return sooner, while secondary schools remain closed much longer. Extending this method to include all facilities of interest in a region is necessary to fairly interpret these results, but these results may indicate that access to food and home improvement was prioritized over schools, or that supermarket and home improvement facilities benefit from being part of national private chains, versus public school systems operated by local county administrators.

2.6 Discussion

The results above show clear periods of lack of access to facilities, whether due to facility closure, transportation complications, cellular network outages, power outages, or other possible barriers. I show that in addition to identifying access in terms of facility closures, LBS data can provide even more utility including identifying when a facility is repurposed (as when schools were reopened as shelters) and detecting changes in supply or need as users exhibit behavioral anomalies in the days before and after disruptive events as well.

2.6.1 Validation

I validate these results with known public-school closure dates. Figure 2.10 shows multiple other known open and closure dates for secondary school facilities in the time periods surrounding Hurricane Irma without rolling-average smoothing. This plot shows a stark increase in appearances on the first day of school (August 16), closure due to flooding prior to Hurricane Irma (August 28), and closure on Labor Day (September 4) [91]. Students returned to school on September 25 in Naples and Marco Island and September 27 in Key West [91, 92]. These results validate this approach for other facility types where closure and reopening dates are not explicitly known.

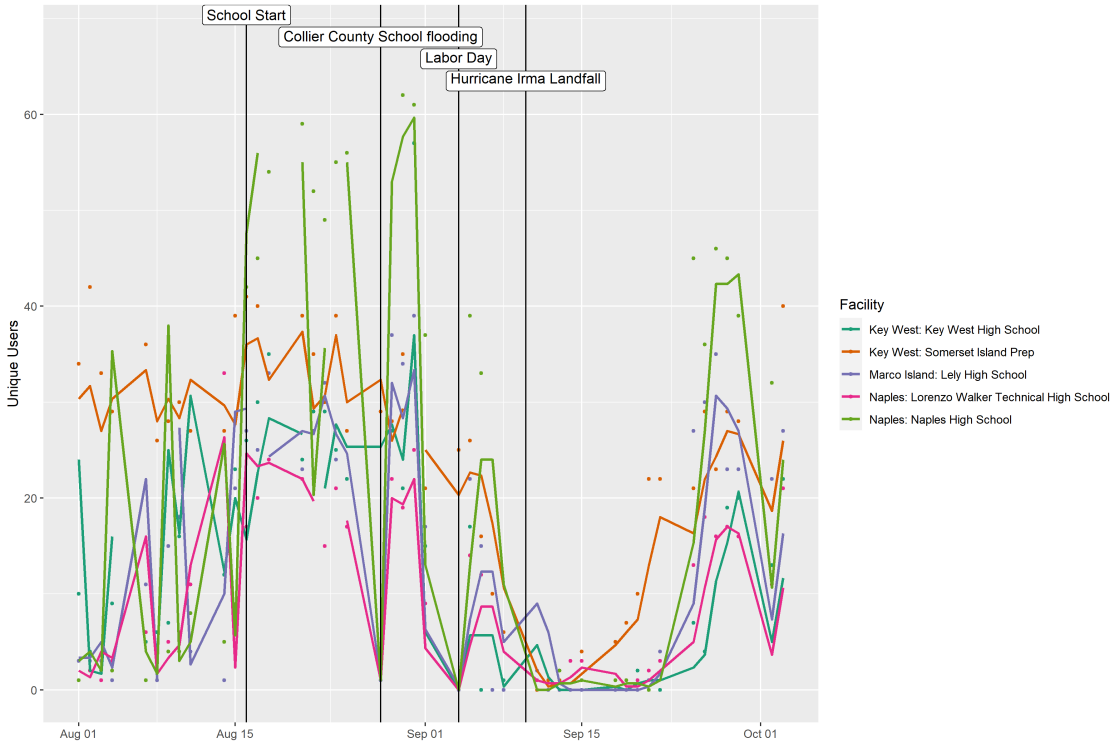


Figure 2.10: Daily unique user appearances for secondary schools showing additional known closure and opening dates for validation

The results for home improvement locations in particular highlight the utility in detecting human behavioral changes through LBS data for identifying deviations such as stock-ups, closures, and restocks, as well as how those deviations may vary between individual facilities or communities based on the differences in hazard experience.

2.6.2 Limitations

Like survey-based hazards response analysis, LBS data comes with its own risks and limitations regarding representation, bias, and privacy. Ensuring representativeness of relevant study areas and populations is a key step in LBS data processing to reduce data-collection bias [59]. Further, while this method is useful for a generalized measure of access like "functional closure", I cannot infer the specific dimension rendering a facility inaccessible. That is, I cannot conclude if reduced appearances indicate a facility is closed, as changes in behavior could also indicate the facility is too expensive, is out of supplies, is unsafe, or is just difficult to get to because of transportation network damage. LBS data may not be useful for childcare, elementary schools, elder care facilities or anywhere else a population would have a low share of smart phones or poor cell phone coverage. These methods may also not

be suitable for facilities in close proximity to other facilities or residences that would render appearances specific to that facility indiscernible, though for some purposes such specificity may not be necessary and anomalies in user appearances across multiple facilities may still be detectable and useful. LBS data include only intermittent timestamped locations that may indicate a change in behavior, but do not include a complete profile of any individual user or contain information on the mechanisms underlying behavioral patterns. Thus, careful attention and diligence is required for understanding apparent behavioral changes from LBS data. Finally, LBS data are at high risk of misemployment that could intentionally or unintentionally abuse sensitive user information. Institutions and researchers are responsible for enacting and upholding policies and practices that protect privacy of individual users, including de-identification that prohibits re-identification, safe data sharing protocols, and only publishing results in aggregate [57, 58].

2.6.3 Implications for Risk Analysis

Risk analysis seeks to understand what the potential scenarios area, the outcomes and consequences or each scenario, and the uncertainties about the scenarios, outcomes, and consequences [23, 28]. In the context of natural hazards risk analysis, there has been significant development of methods for estimating direct damage and losses to building stock, infrastructure, and economies. There remain ongoing challenges in estimating changes in access to essential services after a hazard due to a lack of data as access, damage, and recovery vary greatly between households and hazards may impact vast geographies and populations. However, as [28] point out, this access is essential to recovery and community resilience. The approach I develop and demonstrate in this chapter provides a new way to assess access at an individual facility-level across a wide geographic area such as a state or region. This complements more geographically-limited survey-based approaches and provides risk analysts with an approach for creating a broader understanding of post-event access recovery at the individual essential services facility level. Evaluating the overlapping recovery periods of all facilities of a certain type (e.g. supermarkets) across a geographic area (e.g. census tract) can generate resilience curves that can then be compared to resilience curves of other facility types or communities to evaluate relative resilience. A community's resilience can also be quantified based on the distance to nearest essential services facilities [28] so accounting for which facilities are open on a given day can show how access to an essential service may change over time. Finally, these access measures can be evaluated in conjunction with other resilience measures like time to return from evacuation, time to return to work, and vulnerability indices to identify how access to facilities may be correlated with these measures. Such

knowledge of the community-specific impacts of access (and in-access) to essential services facilities can empower local resilience planning and expedite recovery [65].

While I demonstrate these methods on essential services facilities, they could be extended to any other pertinent location for measuring behavioral response to service changes. For example, changes in user appearances at traffic intersections may show changes in the transportation network in response to flooding, evacuation, or other disruptions. Similarly, these methods could be applied to show behavioral response to non-hazard induced service changes, such as adoption of a new transit station, or induced demand of added highway lanes. As useful as these methods are for analyzing response to individual hazards or events, they are all the more valuable for analyzing broader research questions that may rely on comparing closure and activity behaviors across different communities, hazards, and time periods as LBS data is collected constantly and opportunistically. For example, this analysis could be used to compare behavioral responses between Hurricane Matthew, which struck Florida in 2016, and Hurricane Irma to see if a community that experienced both hazards changed their behaviors or experienced different recovery patterns between those two events. Or, this analysis could be used to compare neighboring communities experiencing the same hazard to identify differences in access and recovery. Such comparisons can reveal communities particularly vulnerable to loss of access to essential services, empowering infrastructure and policy design during non- hazard times to support those vulnerable populations and build more resilient communities. Further, this work can demonstrate how simultaneous accessibility of certain services may contribute to resilience. For example, pharmacies may be able to compensate for some need if urgent care facilities are inaccessible longer, or communities may be worse off if both supermarkets and gas stations are closed at the same time. Finally, with information around individual facility access, there are opportunities for optimizing the reopening of facilities strategically to support community-level access to essential services following a disaster [104]. These sample results demonstrate the potential for LBS data to supplement approaches such as survey-based or individual facility-based data collection to evaluate changes in access during a period of disruption.

2.7 Conclusion

This chapter presents a novel data-driven approach for identifying facility closures following a hazard event using LBS data. Transportation researchers and professionals widely trust LBS data as an accessible research resource and ongoing hazards researchers continue to demonstrate the potential of LBS data for understanding human behavior in response to hazards. LBS analysis has value in standalone analysis or to supplement survey efforts and

pertains to any type of disruptive event to enable more efficient, broad, and comprehensive assessment of community-wide access to essential services. This analysis could be applied to other episodic disruptions like earthquakes, flooding, or power failures [6] or to assess broader-scale mobility changes like mobility responses to COVID-19 pandemic policies [74] that resulted in inconsistent access to essential services [105].

I show how LBS data enables large-scale quantification of the impacts on facility of hazards on facility availability. Quantifying these impacts across disasters and communities can allow planners and decision makers to understand the criticality of various essential services facilities and the infrastructure that connects them to populations. This quantification of disaster response from LBS data can facilitate the design of infrastructure and policy interventions to support equitable access before, during, and after disruptive events, contributing to developing and sustaining equitable and resilient communities.

2.8 Acknowledgments

Thank you to Veraset LLC for providing the data for this work. This research is supported by the National Science Foundation Graduate Research Fellowship under Grant No DGE 1841052 and the Rackham Predoctoral Fellowship at the University of Michigan. The views expressed in this chapter are those of the authors and do not necessarily represent those of the sponsors or the data providers.

CHAPTER 3

Using LBS Data for Quantifying Large-scale Household-level Disaster Recovery

Disruptions such as natural disasters may result in evacuation orders, transit or road closures, power outages, school closures, and other deviations from routine that affect the ability of members of an affected household to maintain their typical behaviors. Visits to home and work, or commuter behavior, follows regular patterns on a daily or weekly basis so that disruption is detectable and has economic and social implications. However, data on how many people lost access to their home or work and how long it took people to resume their typical home and work schedules is sparse and disconnected, and, because of this, cannot show the extent of recovery across a large set of households with diverse experiences of the disruption. I show how to use location-based services data from smartphones to capture patterns in users' home and work appearances and deviations from those patterns that may indicate disruption and then recovery from a disruptive event, while maintaining anonymity. I introduce a Bayesian belief network-based anomaly detection method to identify anomalous periods indicating household-level recovery in response to Hurricane Irma in 2017. I present results showing the proportion of users experiencing an anomalous period and the average length of recovery consistent with the storm's path and validated against available survey results on evacuation. These large-scale data-driven results on household recovery can contribute to further analysis on the impacts of the hazard and social vulnerability on recovery at home and workplaces.

Keywords: location-based services data; disaster recovery; anomaly detection

Note: The research presented in this chapter was an invited submission for the journal *Computers, Environment, and Urban Systems* for the special issue *GeoAI and Location Big Data for Smart Cities*. The corresponding manuscript was submitted July 17, 2023 with co-author Seth Guikema and is currently under review.

3.1 Introduction

Social vulnerabilities exacerbate harm from disasters and inhibit the ability to recover due to lack of economic resources [2, 106, 107, 108]. Additionally, systemic inequalities and discriminatory housing policies, such as "redlining" in which race and ethnicity were explicitly used as exclusion criteria for property loans, result in vulnerable populations disproportionately relegated to neighborhoods that are particularly prone to hazards and lack access to the essential services necessary for preparation and recovery [8, 19, 28, 108]. Post-disaster infrastructure restoration has also been shown to be disproportionately lengthier for socioeconomically vulnerable communities [2, 5, 108, 109, 110]. Understanding inequities in recovery is necessary to design better preparation and response to bolster community equity and resilience, but current survey methods are cost and labor-intensive and so are not suitable for capturing the full scope of recovery experiences across a large region. There is a need for large-scale data collection to supplement surveys to capture household level recovery across a large region's full diversity of households and varying exposures to the hazard.

Location-based services (LBS) data from smartphones opportunistically captures human mobility patterns. When integrated with infrastructure networks, such data can reveal how users interact with the built environment, providing valuable insight for transportation planners and policymakers [37, 39, 41]. LBS data has already been demonstrated to show transportation behavior consistent with prior data collection methods [37, 38, 70, 72, 111]. Now, LBS data is consistently used to supplement cost and labor-intensive household travel surveys for applications beyond transportation ranging from evaluating partisan impacts on family ties and risk perception [40, 112] and modeling infectious disease [74, 75]. LBS data has been used to evaluate disaster recovery through tracking disaster evacuations [7, 63, 76, 77, 113], facility closures [76], and inequities in recovery [6]. Results from these studies show the promise of utilizing LBS data for understanding disaster response, but do not capture the full spectrum of recovery including those who do not evacuate and return to previous activities, such as work.

Commuter behavior as reflected in visits to home and work follows regular patterns on a daily or weekly basis that could be identifiable in LBS data [35, 114, 115, 116, 117, 118]. Following a disruptive event, commuting behavior is inevitably included in any affected household's change in routine. Return-to-productivity after a disruption is typically understood via surveys that capture how commuter behavior changed, such as through differences in route choice, mode choice, departure time, or working from home patterns [45]. Detecting anomalous behavior in home and work appearances reflected in LBS data can thus be used to identify return-to-productivity and assess a community's recovery time following a

disruption. This allows for much larger sample sizes than survey-based approaches, provides better coverage of the population in terms of both geography and specific sub-populations represented in the data, and reduces response time bias as behaviors are captured as they occur rather than relying on human memory days or weeks after the event.

Anomaly detection involves identification of unexpected observations or sequences of observations within time series data [82, 119]. Such deviation from expectation can be local (anomalous relative to neighboring data) or global (anomalous relative to the entire dataset) and is significant enough that the observation or series appears generated by a different mechanism [120]. Anomaly detection can identify changes in regular patterns, such as those captured by sensors monitoring systems from an individual's personal health to quality control in manufacturing to the air quality of an entire community [82, 119]. Simple linear regression models can identify anomalous behavior when receiving data outside of recognized ranges. Model complexity increases as the anomaly appears as a collection of observations rather than a single point [119, 121], as is the case for evaluating human behavior in the days surrounding a disruptive event. Unsupervised machine learning models including long short-term memory neural networks and self-organizing maps can identify and predict human behavior at fine time scales, but require vast amounts of data to define "normal" behavior [121, 122, 123]. Multivariate anomaly detection methods typically introduce even more data requirements to capture every combination of variables that may occur under "normal" circumstances [124]. Bayesian belief networks offer an unsupervised anomaly detection method suitable for coarse time scales while still including conditional dependencies between observed variables [125, 126]. Bayesian belief networks have been successfully implemented in water monitoring [127], healthcare [128, 129], and internet of things [126, 130] domains with relatively short time scales.

In this chapter, I present a novel application of Bayesian belief networks to estimate recovery and return-to-productivity times at the household level from LBS data spanning Hurricane Irma in Florida. Hurricane Irma made landfall as a category 4 hurricane in the Keys the morning of September 10, 2017 and as a category 3 hurricane in mainland Florida later that evening [86], following a trajectory shown in Figure 3.1 from [87]. All 67 Florida counties were qualified under a Major Presidential Disaster Declaration [131].

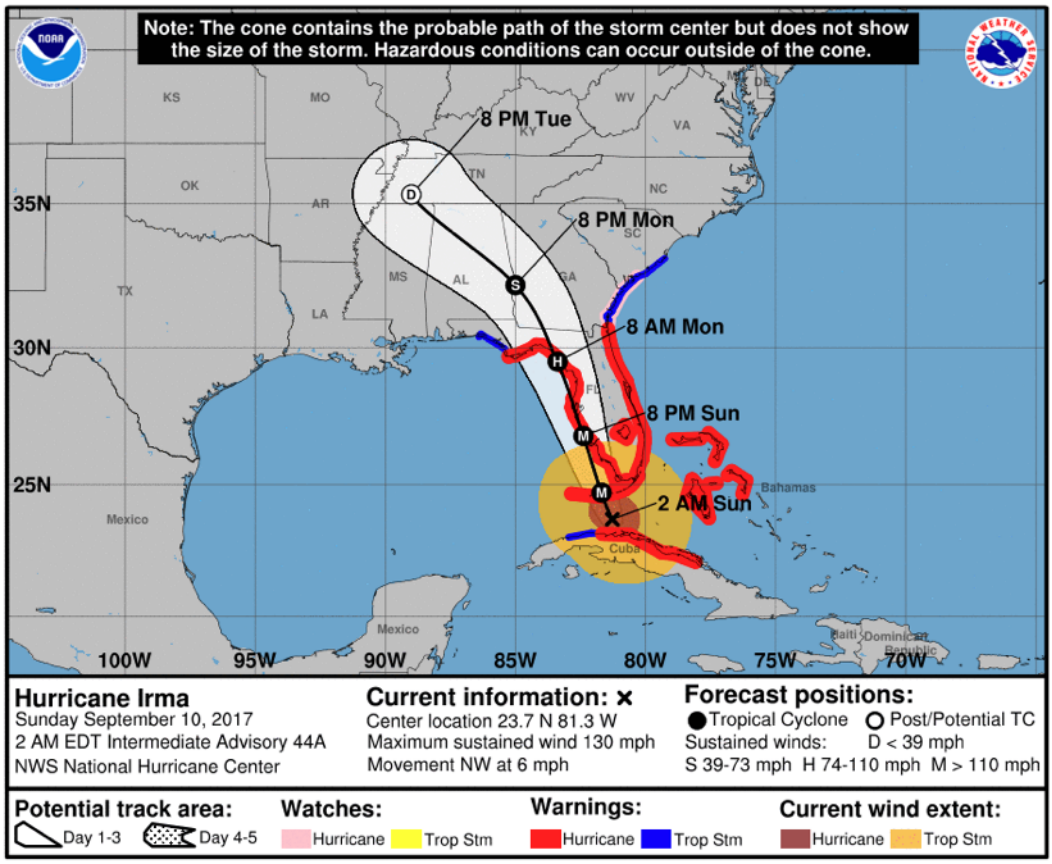


Figure 3.1: 3-day Forecast track, initial wind field and watch/warning graphic from NOAA National Hurricane Center IRMA Graphics Archive

Evacuation orders started for the Keys on September 6 and mandatory evacuations were issued in 42 Florida counties. Over 6 million Floridians are estimated to have evacuated over the next five days, one of the largest evacuations in U.S. history [52, 86]. Hurricane Irma resulted in an estimated 50 billion USD in damages, the most expensive storm in Florida's recorded history at the time [86]. Southeastern Florida sustained most of the wind damage impacts, while North Florida experienced widespread flooding of most rivers [131]. Almost every county suffered some period of power outages, with over 90% of Southeast Florida customers without power for over a week [88]. Access to the western Keys remained closed until a week after landfall still with intermittent power, sewer, and water access [132]. Public schools in Lee, Collier, and Monroe counties remaining closed for over two weeks after landfall. All of these environmental and infrastructure factors contributed to vastly variable recovery times across the state.

In the remainder of this chapter I describe a large-scale, data-driven method for estimating recovery from Hurricane Irma using LBS data and Bayesian belief networks. I describe

the dataset in detail in Section 3.2.1 and in Section 3.2.2 I explain how to identify user's home and work places from LBS data in August 2017. I then generate binary sequences based on users' daily appearances at their home and work locations and apply a novel Bayesian network-based anomaly detection method which is detailed in Section 3.2.3. From the anomaly detection results I heuristically estimate recovery duration for both home and work behaviors as a proxy for return-to-productivity, illustrated in Section 3.2.4. I present the results geographically and validate the estimates against an available survey in Section 3.3. I conclude this chapter with Section 3.4 discussing the limitations and possibilities of this novel method using LBS data to estimate hazard response and recovery.

3.2 Methods

3.2.1 Data

The LBS data was provided by Veraset LLC, a private company that collects and aggregates LBS data from a portfolio of mobile phone applications [51]. A ping, (i.e. data point) represents any exchange between the phone and a cellular network or Wi-fi, including from applications running in the background. Pings are not consistently collected at regular time intervals but vary based on the applications running on a given phone. Each ping includes an anonymized user identifier, UTC timestamp, latitude, longitude, estimate of horizontal accuracy, and device type (Android versus iOS). The horizontal accuracy ranges from 2 to 2000 meters of estimated maximum error, with 75% of points under 30 meters and 95% under 110 meters.

3.2.2 Home and Work Finding

From consistent enough LBS pings, I can identify frequently visited places, including home and work. I adapt a method for identifying home and locations from [133], who identified home and Thanksgiving visit locations from similar LBS data to identify relationships between political partisanship and family ties. First, I filter LBS data by time of day—for home I identify the most frequently visited location visited between 9pm and 6am and for work I identify frequently visited locations visited between 10am and 3pm on weekdays, defined as midday by [134]. I filter out any users who do not make appearances in those time windows at least 10 unique days within the 61 days included in the LBS dataset.

We overlay a 5e-5 degrees latitude by 5e-5 degrees longitude grid over the remaining points and generate clusters from points in neighboring grid cells, resulting in 1.5e-4 by 1.5e-4 degree overlapping clusters (the exact area varies slightly depending on the latitude of the

coordinates, ranging from approximately 54.8 ft by 44.0 ft to 54.8 ft by 49.6 ft). I ascertain the clusters visited the greatest number of unique days, and among those locations identify the cluster with the highest duration spent, where duration is calculated based on sequential appearances in the same cluster [63]. For any user where the workplace is within 0.0015 degrees (about 0.1 miles) of home, I identify the second-best location that still matches the criteria, or filter out the user if there is no distinguishable workplace that fits the previous criteria. I show this process in Figure 3.2 below (adapted from [63]).

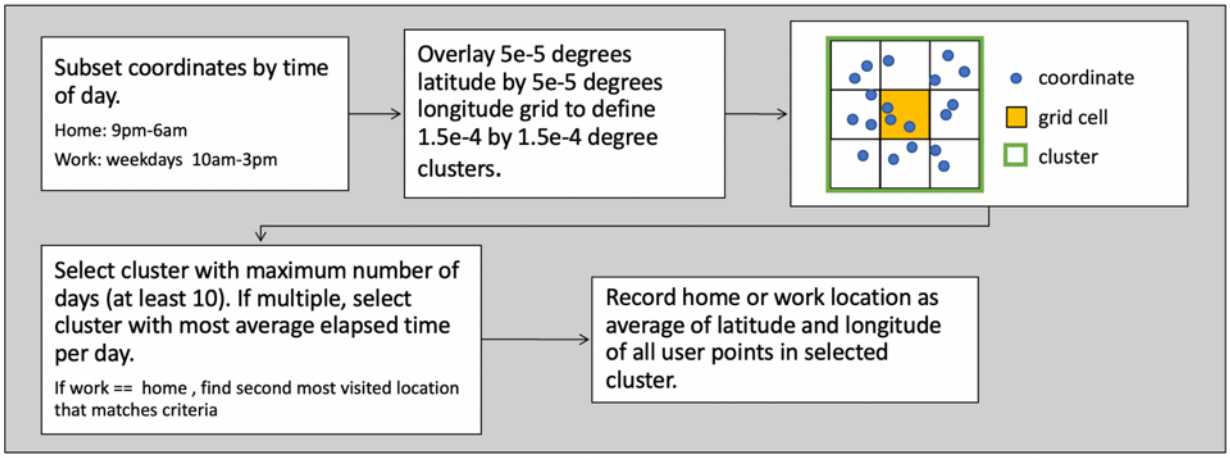


Figure 3.2: Estimating home and work coordinates

We filter out any users with a horizontal accuracy less than 1 mile (0.1% of remaining users) as well as any users who do not appear at their home a week prior to landfall (42% of remaining users). Finally, I filter out users who appear in the data set fewer than 80% of the days in August and September, 2017 (19% of remaining users), as I want to be sure I am excluding short-term visitors and that I have enough daily data to identify visit patterns. In the end, there are 123,445 unique users with an identified home and workplace. I validate the home locations based on the state of Florida 2017-2021 American Community Survey five-year population estimates. I validate workplace locations against the most recent data from the Census Tract Planning Products (CTPP) program tabulation of worker flows published in 2017. The methods and results of this validation are described further in Appendix B.

3.2.3 Bayes Network Probabilities

Of the 123,445 unique Florida users with a home and workplace, I classify all of their pings as appearing within "home" or "work" if they were within 0.0015 degrees of that location. I classified the remaining points as laying within "home county", or "other" which encapsulates

all of a user's unclassified pings within the rest of Florida. Very few users had points in their "work county" that were outside of "work" or "home county" classification and so I did not distinguish "work county" separately from the "other" category. As I am interested the number of days to recovery, any other temporal patterns within users' data add noise to this detection. For example, hourly variations in when users went to work during the day can make it more difficult for a model to identify "typical" behavior over the 61 days of data when I am only interested in if they appear at work or not. Additionally, I do not expect users to spend their free time consistently –some evenings users may spend time at the house while others they may spend doing work or leisure activities away–so I do not want to design a model with the potential high hourly variation when I am ultimately interested in daily appearances. Furthermore, the irregular inter-ping times within and across user data means that gaps in time series due to users simply not using their phones would be more prevalent. Within an hour it may not be abnormal for a user's phone to not ping if they are not actively using their phone or they set it to airplane mode during work, leisure, or sleep activities, but given the user filtering process, I would expect users to ping at least once a day unless they are outside of the geographic study area. So, I aggregate these classified points to binary sequences, with a point representing whether the user appeared in that place on that day or not. This aggregation resulted in 61 data points for each classification ("home", "work", "home county", "other Florida") for each day in August and September, 2017. I present examples of this aggregated data for a sample of anonymized individuals in Figure 3.3, where Hurricane Irma's first landfall in Florida is marked by a vertical line. For context Monroe County is made up of the Florida Keys as well as the uninhabited Everglades National Park on the mainland; Collier County is the next most southwestern county and contains Naples; Miami-Dade County is the most southeastern county as well as the most populous and contains Miami; Palm Beach County is north of Miami on the east coast of the state and is the third most populous county in Florida, containing West Palm Beach and Boca Raton.

Figure 3.3: Sample daily aggregated user data

We can see regular, but unique patterns emerge from these example users in Figure 3.3 as well as a change in patterns coinciding with Hurricane Irma's landfall. The user from the Keys appears at home every night (except August 18), at work five to six days a week, elsewhere in their home county nearly every day, and leaves their county for another location within Florida for 13 days aligning with evacuation from the islands for Hurricane Irma. They return to their home eight days after landfall, and eventually to work on Monday, September 25. The Collier County user appears at home every night in August except one, when they also appear in another county in Florida for three days, perhaps indicative of a brief work or leisure trip. This user appears at work almost every Monday through Friday and August and has more variability in their home county appearances. They seem to leave their home on Sunday, September 10, the day before landfall on the mainland and perhaps lose cell service or leave the state, though they do have an appearance in Collier County on September 13 before returning home on September 16 and the to work on September 25. I observe more variability in the Miami user's time series, where they spend several nights away from home in August, appear at work anywhere from two to five days a week, and visit another county in Florida for a weekend. Despite this variation in appearance behavior, I can identify a clear period away from their county September 7-12, perhaps attributable to the hurricane. This user returns to their home county on September 12 and resumes workplace appearances the next day, though does not appear at home again until September 16 indicating they temporarily stayed in a hotel or with family or friends which may denote a power or water outage at their home while their workplace remained accessible. The user in Palm Beach appears at home every night and did not seem to have home appearances disrupted by the hurricane, but their visits to work and other locations within their county pause in the days surrounding Irma's landfall.

While such patterns for these particular users are straightforward to see, for others with high variability in appearance behavior I may not be able to distinguish hurricane response from normal behavior. Still, for those with such discernable disruptions, I would like to be able to use this LBS data to estimate hurricane response and recovery. With 123,445 unique users, I cannot expect to parse through each user's unique data as I did in the above examples to estimate recovery periods. But with such limited data in each user's time series, I also cannot expect a deep-learning model to effectively identify spatial and temporal patterns specific to each user. I would like to develop a model that builds in the intuition about which time-based variables may impact an individual's likelihood of appearing at home, work, or elsewhere in their home county or Florida on a given day. Specifically, I want to build in the assumptions that day of week, weekend or holiday status, appearances on previous days, and appearances in other locations affect the probability of a user appearing

at home and/or work each day.

Bayesian belief networks offer a suitable method for integrating the contextual knowledge with limited user data to estimate the joint probability of an appearance conditional on these features. I do not want to assume all random variables are conditionally independent, as with Naïve Bayes classification, but I do not have enough data from each individual to calculate conditional probability for every combination of events. I therefore design a Bayesian belief network as an intermediary between a fully conditional and a fully independent probabilistic model. Figure 3.4 below demonstrates an example probabilistic graphical model where I assume a conditional probability between a home appearance and day of week, weekend, and the previous 3 days as well as independent relationship with appearances elsewhere within their home county and state.

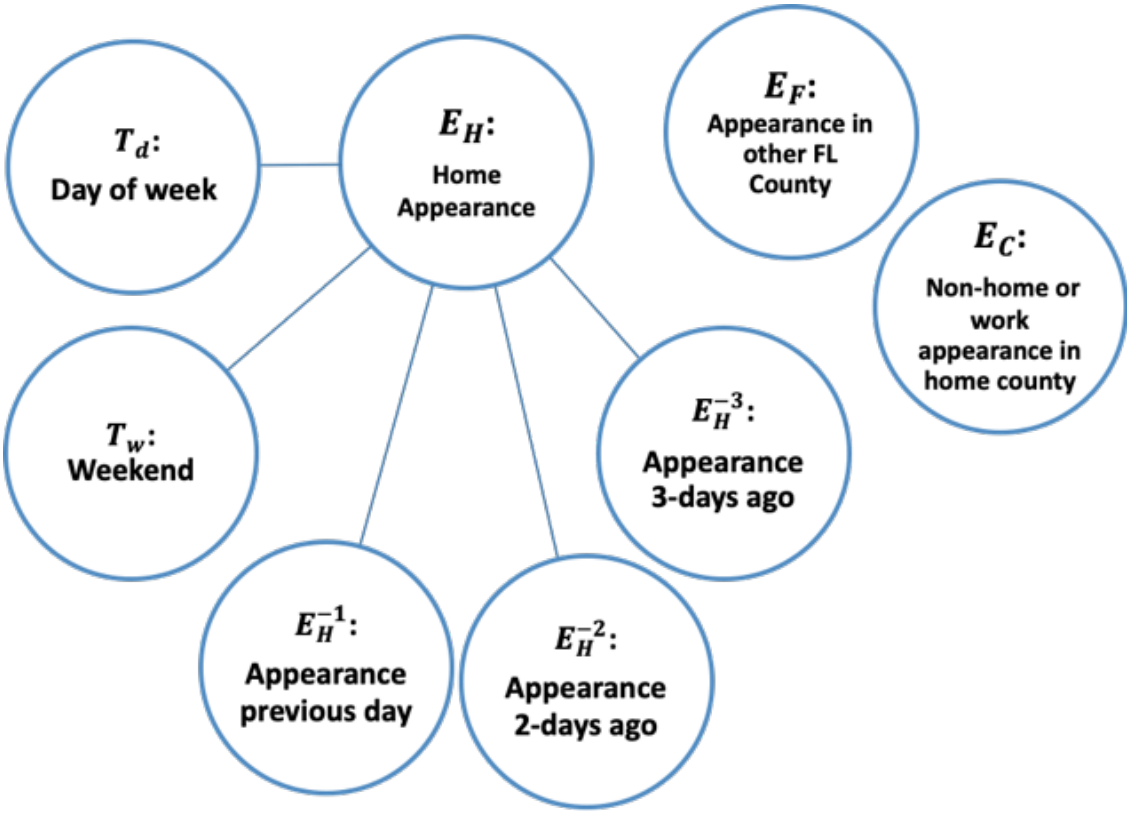


Figure 3.4: Bayesian belief network for probability of home appearance

This network can also be described through Equation 3.1 to determine the probability of each user's (u) variable values on each day as:

$$
\begin{aligned}
P_u^D &= P((E_H \& E_C \& E_F)|D) \\
&= P(E_H) * P(E_C) * P(E_F) * P(T_d|E_H) * P(T_w|E_H) * \prod_{t=1}^{L} P(E_H^{-t}|E_H)
\end{aligned}
\tag{3.1}
$$

where L is the number of previous days,]or lag days, expected to impact the likelihood of E_H and D is the set of available data $\{E_H,\ E_C,\ E_F,\ T_d,\ T_w,\ E_H^{-t}\}\ \forall\ 1 \leq t \leq L$. E_H, E_C, and E_F are binary representations of whether an appearance occurs that day or not, T_d is a categorical variable for each day of the week, and T_w is a binary variable indicating whether the day is a weekend or not. I also indicate the U.S. federal holiday Labor Day as a weekend variable, which fell on Monday, September 4, 2017 in the study period. I calculate these probability distributions separately for home and work appearances to garner separate estimates for return-home and return-to-work recovery estimates.

We determine these probabilities for all user's August data as a prior distribution to probabilistically model inherent normal-condition variability in patterns across the population. I then create an individual model for each user based on only their August data. I update each individual's probability distributions with the population-wide prior distribution weighted by some number of days W. For example, for $W = 7$ the prior would be worth a week's worth of population-wide data supplementing each user's individual August data. Then for each user, I calculate the daily joint probability of the realized variables from Equation 3.1 for each day in September. Unlikely or previously unseen combinations of values will then have relatively low probabilities, which I can identify as anomalous based on some determined threshold. I avoid zero-probability events by incorporating the prior, which ensures any combination of events observed in August across all users has a greater than zero probability.

We define a threshold for defining an anomaly a_u for each user based on their previous relative probabilities. Such a threshold must be robust enough to account for some users having more variable home and work appearances than others. For example, for users with very regular or steady behaviors, like appearing at home every day and at work the same five days every week, deviations from those norms like missing a day at home or missing three days in a row at work will result in extremely low relative probabilities that I can more easily identify as anomalous. For users with more variable behaviors, such as those who were out of town for a couple weekends in August or perhaps have a less rigid work schedule, the probability of being away from home in the days around landfall may resemble their previous August departures, so I design a threshold that will only detect behavior as anomalous if it falls far enough away from each user's previously seen data.

To select the model parameters and a_u value, I tested this Bayesian network-based anomaly detection framework with various combinations of input variables, lag-day values

in $L = \{0, 1, 3, 7\}$, and prior weights in $W = \{1, 7, 14, 28\}$. I compared different anomaly thresholds:

$$a_u = min(P_u^D \forall D \in August) \tag{3.2}$$

as well as:

$$a_u = mean(P_u^D \forall D \in August) - S * std(P_u^D \forall D \in August) \tag{3.3}$$

for $S \in \{2, 3, 4\}$. That is, I tested thresholds set to the minimum of each user's August day probability P_u^D in addition to thresholds set to two, three, or four standard deviations from the average probability seen in August. I evaluated the sensitivity of the model to each of these parameters which I share in Appendix C.

3.2.4 Defining Anomalous Periods

The results from the Bayesian network probability estimation indicates if each day is anomalous or not. I use these results to estimate the total anomalous period that could be attributed to Hurricane Irma. I set the earliest start date for a hurricane-related anomaly to Wednesday, September 6, 2017 to avoid capturing behavior related to labor day on Monday, September 4, 2017. I set the latest start date to Wednesday, September 13, 2017 to account for users who do not have observable deviations from typical behaviors until three days after landfall for those who may typically be away from home or work up to three days in a row prior to the disruption. Within this range of start dates, I identify the first anomalous day to start the anomalous period. I define the end of the anomalous period as the last anomalous day before three non-anomalous days in a row. This is to account for users who may typically take one to two days away from home or work, such as weekends, which may be recorded as non-anomalous but precede a continuing anomalous period. If there are multiple anomalous periods with a start date within the stated range, I define the anomalous period as that with the longest duration. These defined anomalous periods are then used as the measure of the recovery time for home and work.

3.3 Results

3.3.1 Sample Individuals

To showcase the estimation capabilities and limitations of this model, I present the same example individuals as in Figure 3.3 along with their anomaly detection results using the 3 standard deviation threshold for home and work in Figure 3.5.

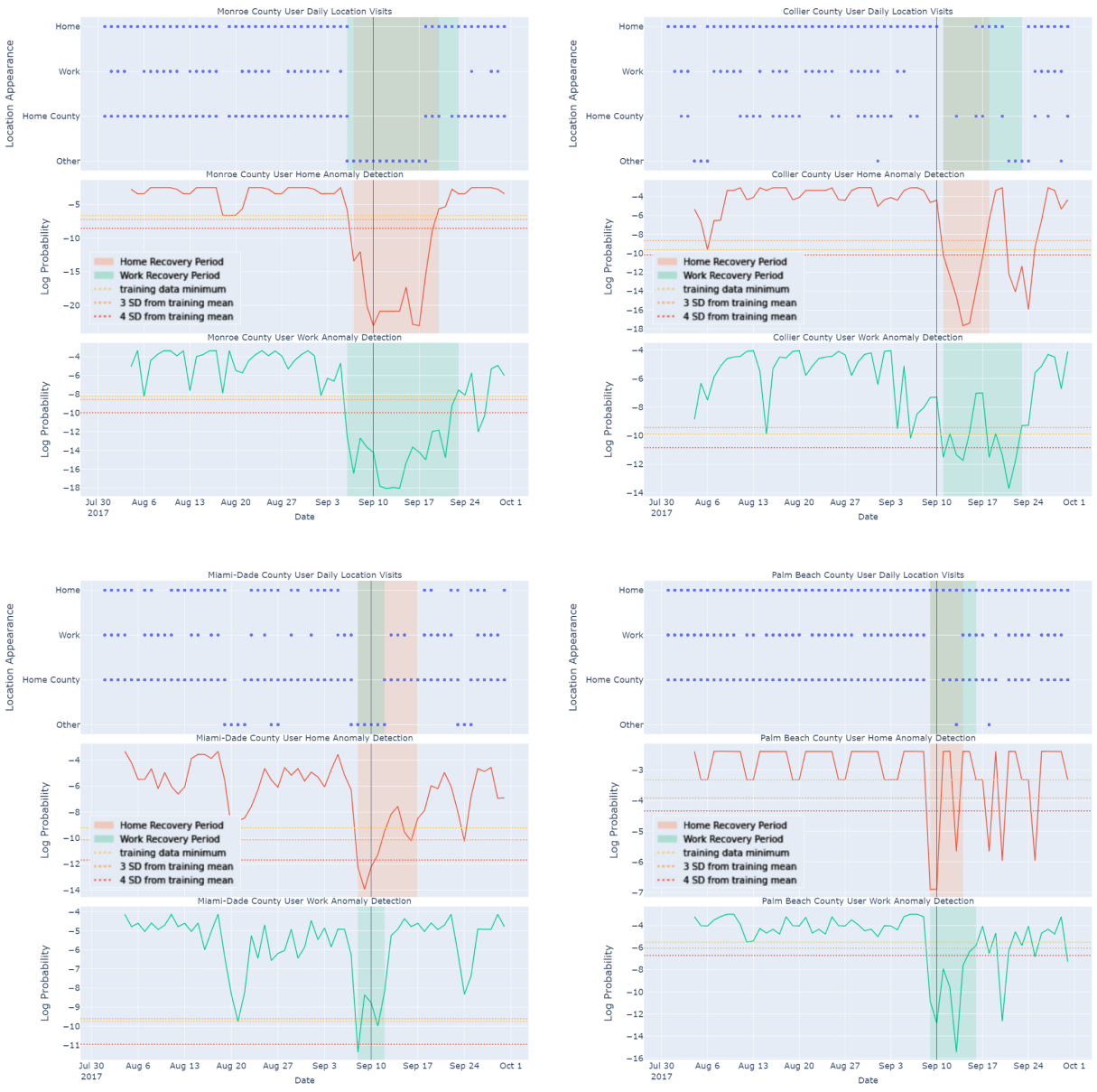


Figure 3.5: Sample daily aggregated user data with shaded home and work recovery periods

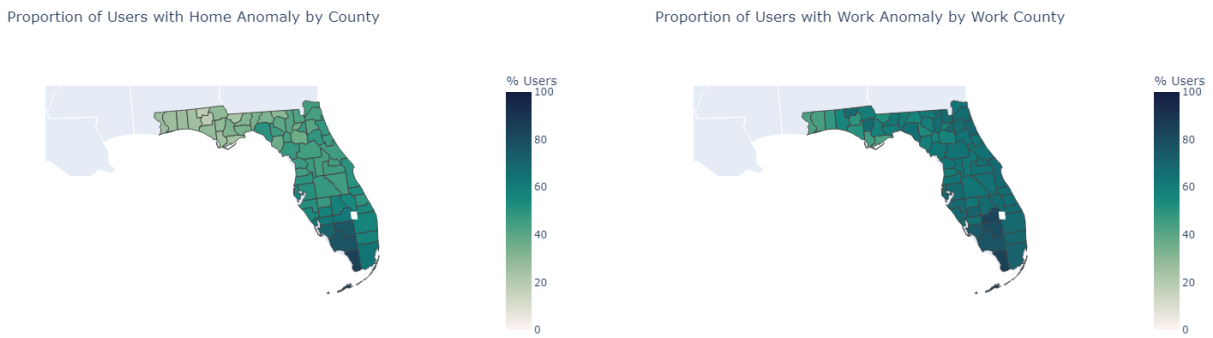
For the Monroe County user, I identify a home recovery period that lasts from their first appearance in another Florida county to two days after their return, likely do to the three-day lag variable in the network and the lack of variability in the home data prior to September 6. The work recovery period spans all missed work days, ending on a Saturday as the model is able to determine this user does not typically appear at work on Saturdays and Sundays. For the Collier County user, I see a similar effect of the three-day lag variable home recovery estimation resulting in the recovery period ending two days past when the

user first appears at home. The Collier County work recovery period starts several days after their last appearance at work, as the model determines it is typical for this user to miss several days of work. The Miami-Dade County user has more variability in their home and work appearances in the training data, and so the model is attuned to pick up only when this user does not appear in their home county as a work anomaly while their home recovery period starts two days after their previous home appearance. The Palm Beach user has a home recovery period that seems to be defined by their lack of appearances in their home county during landfall and their appearance in another county a few days later. As they rarely miss work, their work recovery period continues for two days after their first work appearance due to the influence of the three-day lag variable.

While these results do not perfectly align with first and last home and work appearances, this Bayesian belief network captures does capture the changes in user behaviors. For users appearing at home or work every day, this method is sensitive to the lag-day variable, as observable in the sensitivity analysis reviewed in Appendix C. I elected to keep the three-day lag variable to better capture users who do have regular gaps in their home and work appearances in August, as I assume many users may take some days away from the defined locations for work or leisure activities. Due to this sensitivity, I may be overestimating the duration of recovery periods for users with low variability. About 43% of users appear at home every day in August and would be most sensitive to this parameter. In Appendix C I show that this 43% is evenly distributed across counties and so I assume this source of bias is consistent across counties. With this in mind, I discuss all duration results in aggregate and relative terms, rather than using the absolute, individual values guide the findings.

3.3.2 County-level Recovery

We plot the proportion of users with any detected anomaly in Figure 3.6 based on the combination of parameters $W = 7$ and $L = 3$, and an anomaly threshold based on users' relative daily probabilities three standard deviations from the observed August mean:

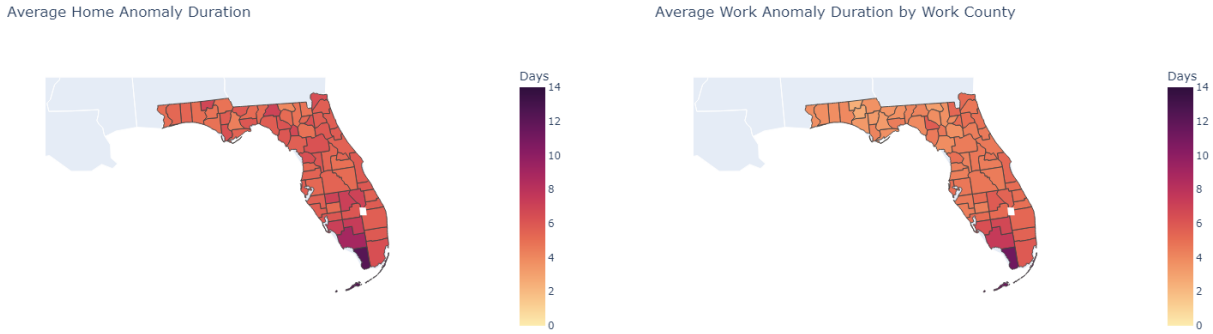


(a) Home (b) Work

Figure 3.6: Proportion of users with detected recovery period

Figure 3.6a shows the proportion of users detected as experiencing a disruption to their daily home appearance behaviors follow the storm's trajectory as shown in Figure 3.1. Compared to Figure 3.6a, Figure 3.6b shows that I identify more users experiencing an anomaly in daily work appearances.

Figure 3.7 plots the average duration of 3.7a and 3.7b work recovery periods across all Florida counties of the users who experienced a detectable disruption.



(a) Home (b) Work

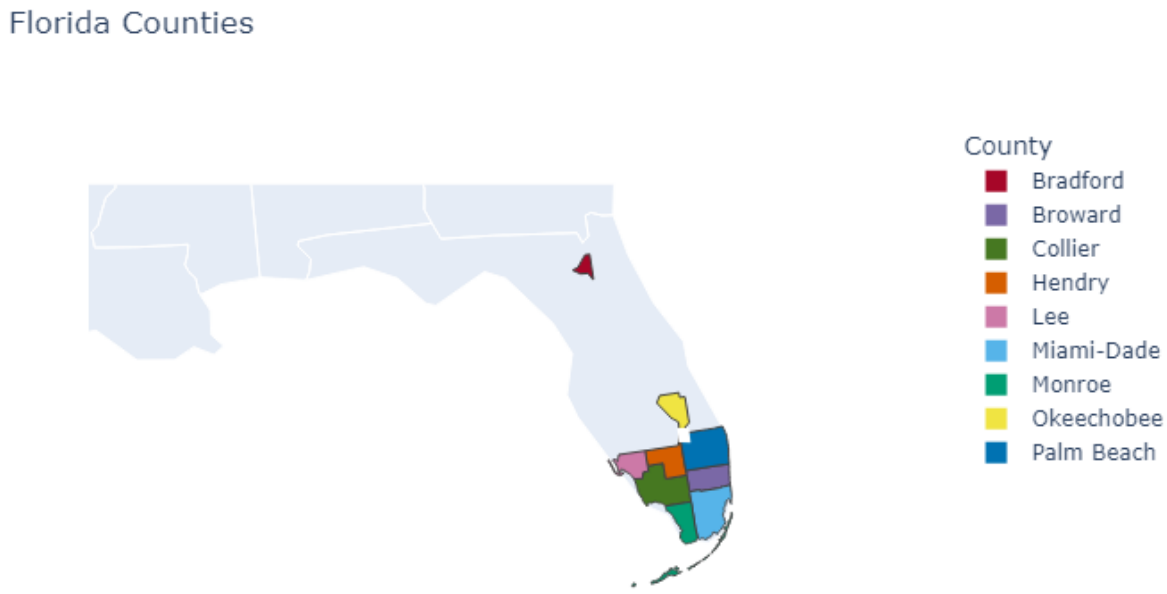
Figure 3.7: Duration of recovery period of users who experienced disruption

Figure 3.7 shows the average number of days of estimated recovery for home and work in Monroe County, which includes Key West, is almost two weeks, which mirrors the long periods of power outages particularly experienced in Monroe County. Further, the period of

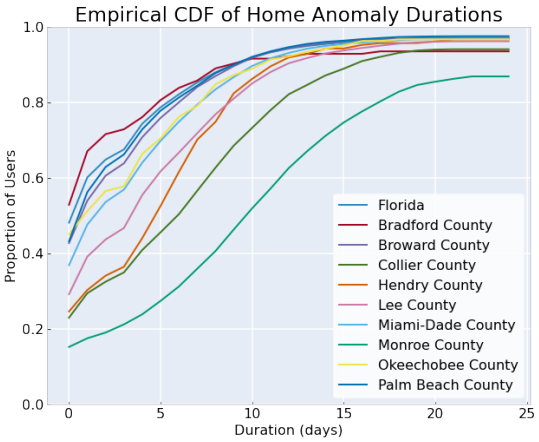
disruption for those who had a detectable home appearance recovery period is longer than the estimated recovery period for work places.

The results in Figure 3.6 indicate that more people experience disruptions in their workplace appearance patterns than at home, while Figure 3.7 indicates that the average recovery period for home disruptions was longer than recovery periods from work places. This makes sense as I can assume that many of those users with longer detected home disruptions either evacuated or faced sustained damage to their home or surrounding infrastructure networks, while those who did not face any home disruptions still may have had to miss work due to evacuation orders, power outages, or other infrastructure damage. Workplaces in Florida are also more concentrated in dense areas due to zoning policies and shared infrastructure needs, and dense areas generally see faster response times from utility companies.

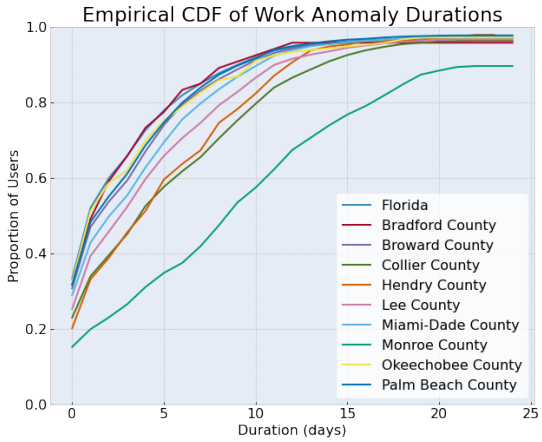
In Figure 3.8b and 3.8c I visualize home and work recovery estimated duration periods through empirical cumulative distributions of total recovery time for example counties across the state, highlighted in Figure 3.8a. I include Monroe, Collier, and Lee Counties for their proximity to the storm path; I show Miami-Dade, Broward, and Palm Beach Counties as the most populous counties in Florida and for their location on the southeast side of the state; finally, I select Hendry, Okeechobee, and Bradford Counties as a sample of inland counties that experienced significant flooding following Irma. The data represented for Florida includes all counties in the state.



(a) Select Florida counties



(b) ECDF home recovery duration



(c) ECDF work recovery duration

Figure 3.8: Empirical cumulative distributions of recovery time by county

Bearing in mind the model's sensitivity to the number of lag days parameter, I can evaluate

these county-aggregated durations relative to one another rather than as exact duration periods. Figure 3.8 shows that 50% of Florida users faced some kind of home recovery, with about 80% of users recovered within five days. I see that about 70% of Florida faced some sort of work recovery, though the recovery duration curves across all counties are steeper, consistent with the discussion from Figure 3.6 and Figure 3.7. Only about 25% of Monroe County residents and workers resumed their previous appearance behaviors in five days, with around 25% of users still unrecovered after two weeks.

These CDFs from Figure 3.8 show that even after 25 days, counties do not necessarily reach 100% recovery, particularly Bradford, Collier, and Monroe Counties for home recovery and Monroe County for work recovery. As the LBS dataset extends only through September 30, 2017, I am not able to identify users who return to their home and work patterns after September 28, 2017 as they do not exhibit at least three days of "normal" behavior to mark their return. While some of these users may return shortly after the study period, I can estimate many are users who face long-term displacement. In Figure 3.9 I show the proportion of users without a return date from the algorithm.

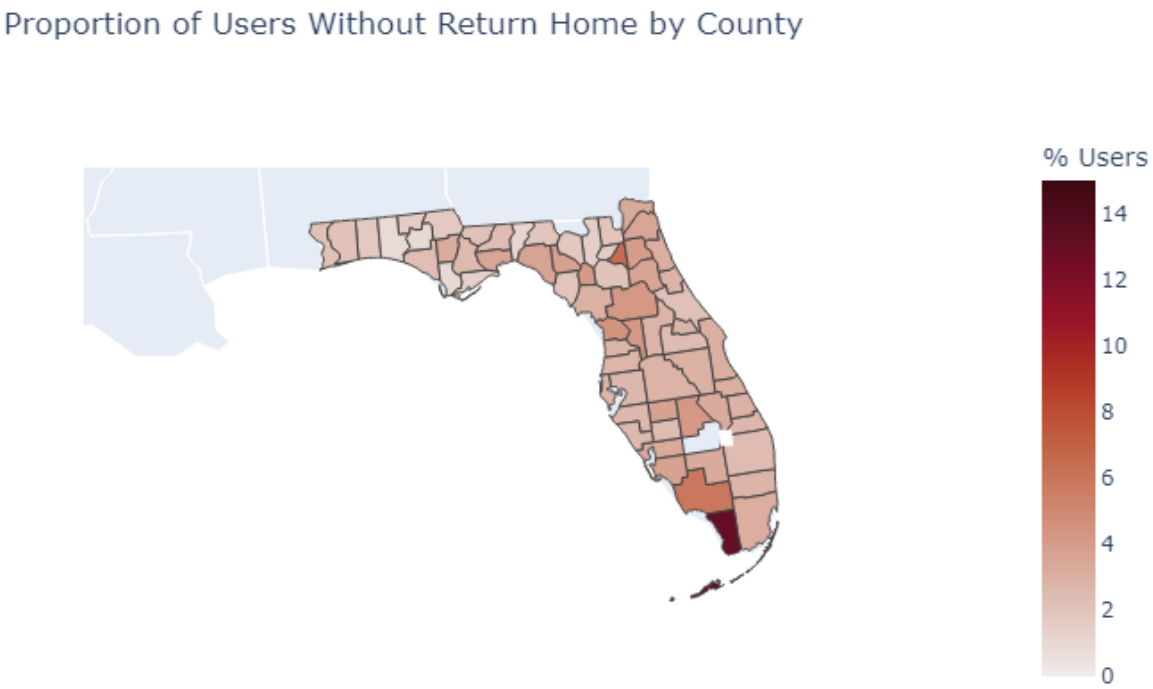


Figure 3.9: Proportion of users without detected return home

Monroe and Collier Counties again stand out by this metric, with 13.1% of Monroe County and 5.9% of Collier still unaccounted for by the end of the study period. One month after Irma's landfall, the Monroe County Mayor estimated the county would lose 15 to 20 percent of its population in the long term [135], which is supported by this 13.1% estimate. Bradford County in Northeast Florida also stands out, with around 6.5% of the 155 users not returning home by the end of the study period. Bradford County experienced the highest cumulative rainfall of the state, with 11.5 inches flooding the New River and Lake Samson, causing long term displacement in the surrounding communities [131].

3.3.3 Recovery Over Time

In addition to looking at total duration, these recovery results can be plotted by date to portray resilience curves, as shown in Figure 3.10.

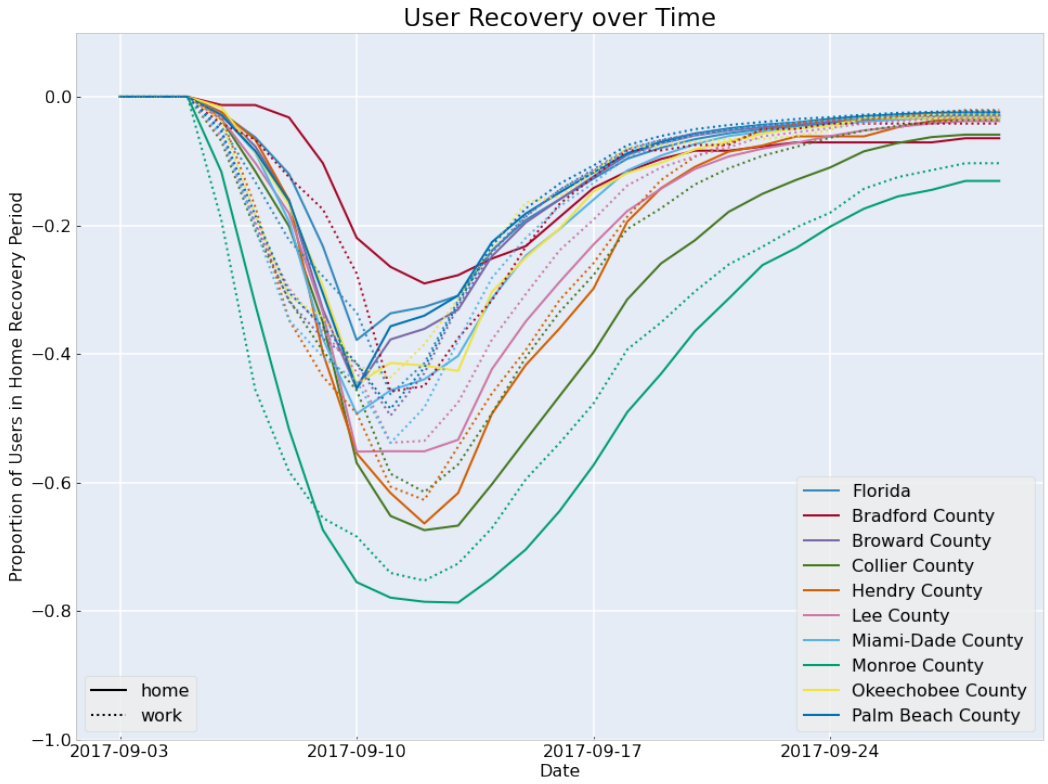


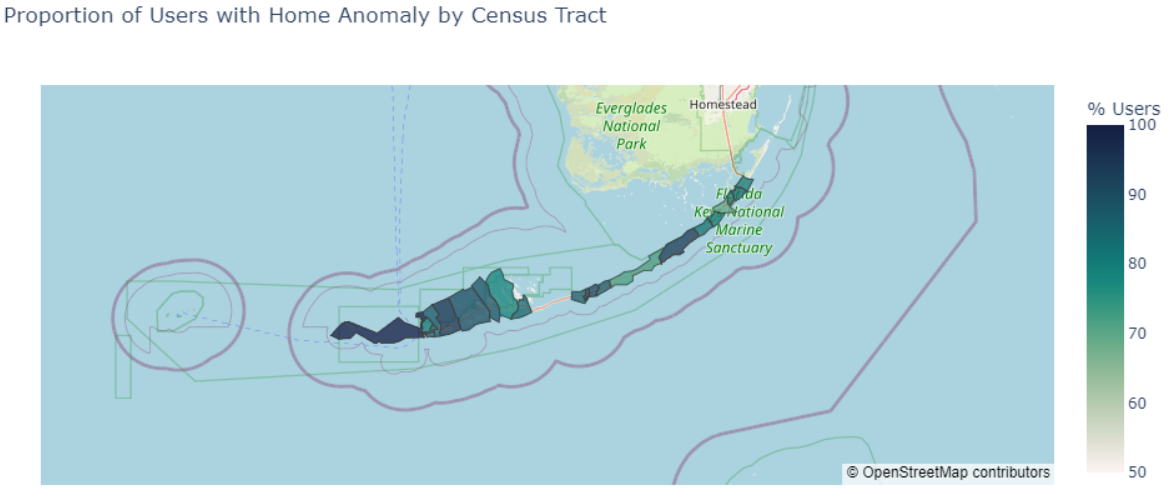
Figure 3.10: User recovery over time by county

Again, given the model's sensitivity to the lag days variable, these aggregate results should

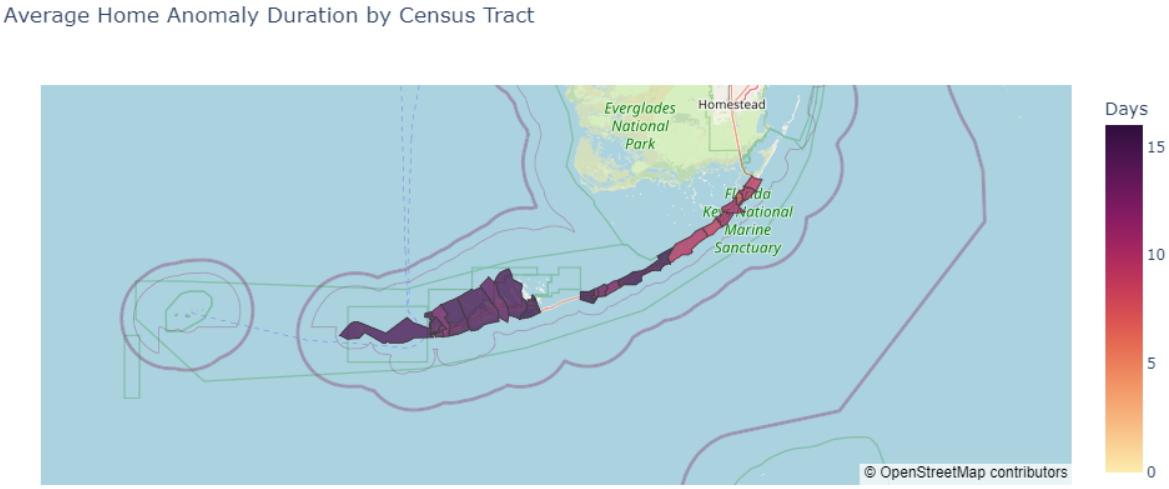
be evaluated relative to one another and are necessarily representative of exact recovery dates. Figure 3.10 shows more about the unique user experiences of recovery based on home and work counties over time. Monroe County recovery periods start earlier, consistent with evacuation orders starting September 6, while mainland counties started evacuating days later [136, 137]. Work recovery generally starts and ends earlier for all counties, though has greater maximum magnitudes of recovering users for all counties except the hardest hit Southwestern counties Monroe, Collier, and Lee. A smaller proportion of users with homes in the inland counties Bradford, Hendry, and Okeechobee face recovery, but those who do see it later and have longer durations than those in the more coastal and densely populated Miami-Dade, Broward, and Palm Beach Counties. This suggests these users faced longer impacts of flooding and power outages, perhaps because repairs were prioritized for more populated and dense communities. Further, large-scale utilities and denser communities are more capable of investing in resiliency and storm-preparation measures that would facilitate faster recovery [5]. Decision makers can utilize such recovery curves at different geographic scales to compare communities and possible identify discrepancies in recovery for sub-populations that may be worthy of additional investigation. For example in Fig. 3.10, Bradford county appears to have fewer residents facing household recovery than other counties, but towards the end of the month up to 5% of those residents remain unrecovered. Further investigation shows that Bradford county faced inland flooding and long-term damage of a significant number of homes, even though the county is quite inland and did not face the eye of the storm, and so may otherwise go unnoticed if surveys are focused only on the hardest hit locations. Further, this plot shows differences in home versus workplace recovery– those counties for which home recovers faster than work may require different interventions than those with workplaces recovering faster than homes. This sort of data is not typically available in surveys, particularly over a large population and geography, and so is another unique benefit of using LBS data for hazards recovery analysis.

3.3.4 Census Tract-level Recovery

Using LBS data facilitates finer-scale analysis to gain insight into community-level response. In Figure 3.11, I show the proportion of recovering users and average detected home recovery duration for census tracts in Monroe County (The Keys) with at least 10 users.



(a) Proportion of users with detected recovery period for home



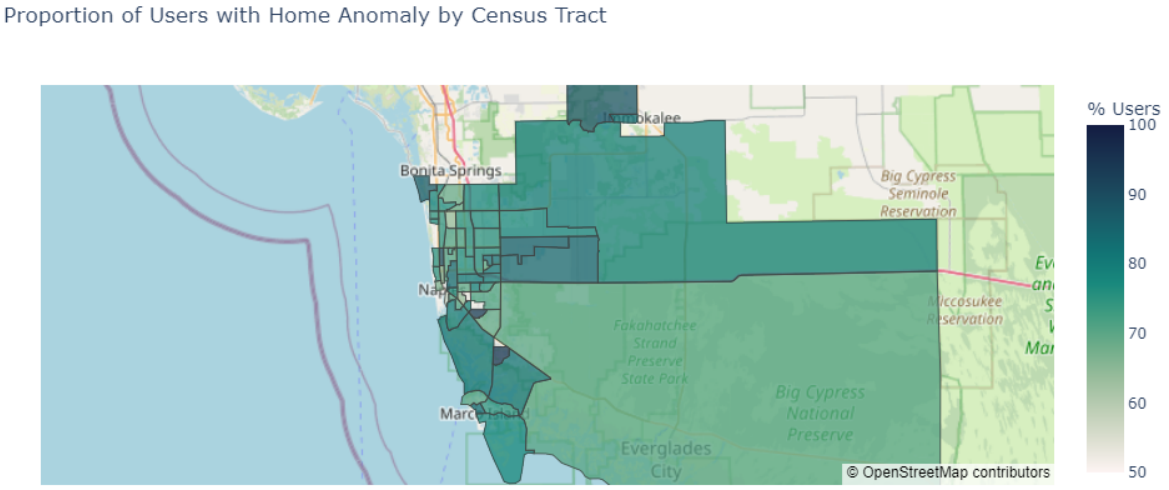
(b) Duration of home recovery period of users who experienced disruption

Figure 3.11: Florida Keys recovery by census tract

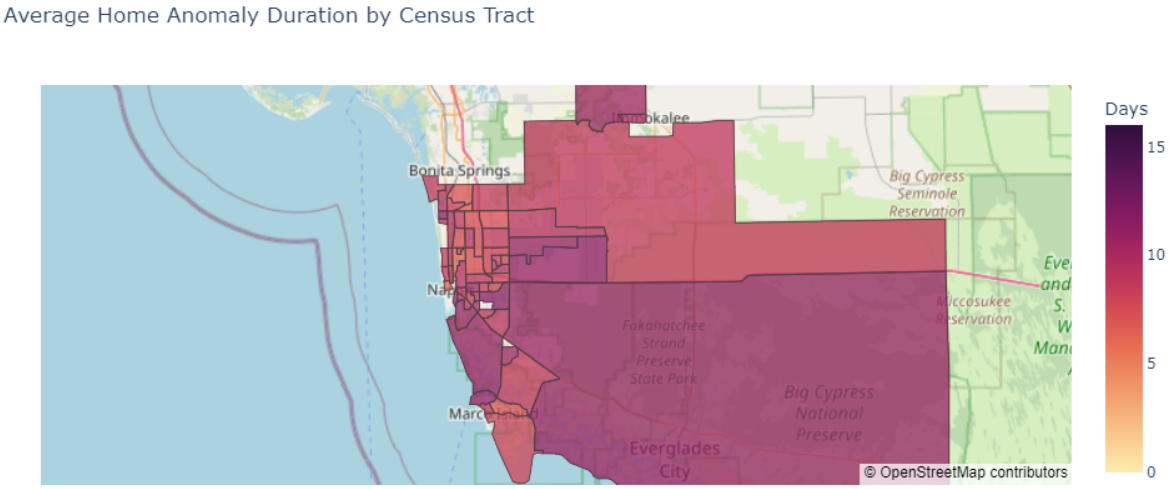
In Figure 3.11a the proportion of users experiencing detected recovery varies somewhat across the county, with the 11 census tracts with over 90% recovering user population distributed across the county. In Figure 3.11b, though, there is a noticeable pattern of the tracts east of Long Key taking an average of eight-ten days to recover, while those western tracts

average 12-16 days of recovery time. This reflects the reopening timeline of the Keys where access to the upper Keys (Key Largo to Islamorada) was reopened on Tuesday, September 12, 2017 [138], while access the lower Keys was not permitted until a week after landfall on Sunday, September 17 [132].

We show the same metrics in Figure 3.12 for Collier County, the southwestern most county in mainland Florida that includes the Naples and Marco Island communities that faced storm surge up to 10 feet.



(a) Proportion of users with detected recovery period for home



(b) Duration of home recovery period of users who experienced disruption

Figure 3.12: Collier County recovery by census tract

No obvious spatial patterns stand out in Figure 3.12a or 3.12b, but there is perceptible difference in how these smaller neighborhoods and communities experienced recovery. This suggests the dynamics driving disruption and recovery that cannot be reduced to proximity to the coast or urban versus rural areas, but other possible drivers such as socioeconomic

vulnerabilities, power or water outages, and access to essential services like supermarkets and home improvement stores [28].

3.3.5 Survey Validation

As stated previously, the ability to assess large-scale localized disaster recovery is out of scope for household survey methods, and so data available for validation is scant. For Hurricane Irma in Florida, I am able to compare these results to regional evacuation data from survey analysis conducted by [52]. In Table 3.1 I compare the survey's report of percent evacuated from 645 responses from the survey defined regions shown in Figure 3.13, to the corresponding proportion of LBS users with a detectable home recovery period when I set parameters to $W = 7$ and $L = 3$, and an anomaly threshold of three standard deviations from users' observed August means ($a_u = mean(P_u^D \forall D \in August) - 3 * std(P_u^D \forall D \in August)$).

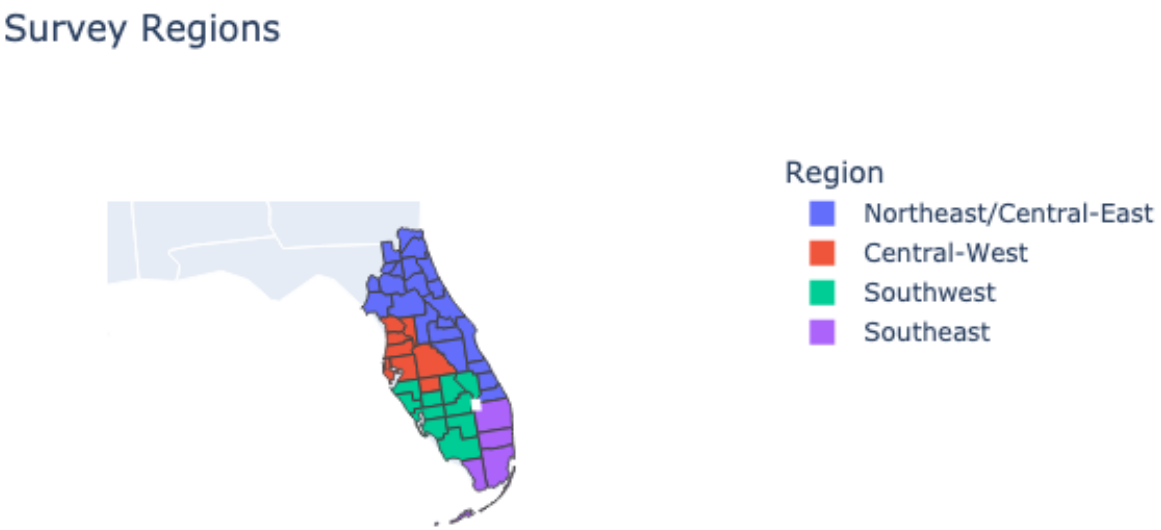


Figure 3.13: Map of regions survey defined regions

Region	Survey		LBS		%
	% Sample	% Evacuated	% Sample	% Evacuated	Difference
Northeast/ Central-East	54.7%	46.2%	36.2%	46.3%	0.3%
Central-West	2.9%*	73.7%	20.1%	51.4%	30%
Southwest	32.6%	72.4%	12.8%	64.0%	11%
Southeast	9.8%	61.9%	30.9%	60.1%	3%
Total	100%	57.1%	100%	53.9%	6%

*This sample size is deemed by the survey team as too small for providing descriptive statistical conclusions

Table 3.1: Survey evacuation versus LBS home recovery

Other than the Central-West region, which does not have a large enough sample size from the survey, the LBS estimates are close to the survey evacuation values, particularly the Northeast/Central-East region where the survey had the most responses. I compare these survey results to additional model results with other anomaly thresholds in Appendix D, but based on these results identify that $a_u = min(P_u^D \forall D \in August)$ and $a_u = mean(P_u^D \forall D \in August) - 3 * std(P_u^D \forall D \in August)$ most closely align with this survey's results.

While I cannot validate these results at a more granular level than these large survey-defined regions, this method is effective at providing estimates for household and workplace recovery that align with the few aggregated survey values reporting evacuation. To truly evaluate the parameter choices and results of the model, household and workplace recovery periods should be compared to survey results with the same definition of "recovery period" for a representative sample across the state of Florida. In lieu of such data, the results provide a baseline estimate for evaluating the relative recovery periods across communities at a finer spatial scale than available survey results.

3.4 Discussion and Conclusions

This work demonstrates the ability to estimate household level and return-to-work recovery periods for large regions, such as the state of Florida. I show how to estimate behavioral probabilities using a Bayesian network with only two months of LBS data and how to subsequently identify anomalous behavior that may correspond to a hazardous event. These results show how this method can be used to identify spatial and temporal trends from local census tract to state-wide scales, beyond the scope of typical household post-event surveys. Thus, this methodology can supplement surveys to evaluate recovery periods along

with storm parameters and social-vulnerability indicators to assess the localized factors contributing to a household or workplace's ability to recover from a disruptive event.

This methodology is limited by its overarching definition of "recovery period" that does not indicate the cause of the anomaly. In the case of hurricane response, a user may present an anomalous period if they evacuate, experience power or telecommunications outages, or conveniently left for vacation during the same time as landfall. For Hurricane Irma in particular, the time-based heuristic for identifying an anomalous period is susceptible to the concurrence of Labor Day the week before landfall. The selected parameters defining a recovery period may undercount those who perhaps evacuated their homes prior to Labor Day and did not return after the storm. I may be overcounting recovering households by including those who were not appearing at home or at work because they had planned vacation time even before news of Irma's path. From the sensitivity analysis shared in Appendix C, I also know this model is sensitive to the number of previous days' data included in calculating the conditional probabilities. With more days of data capturing users' normal behaviors, this model could be more attuned to those typical behaviors and thus better at identifying anomalies, but I show how to estimate recovery with only one month of pre-event data.

Further, the results of this model are sensitive to the definitions of home and work that may not be inclusive of all daily behaviors that are disrupted by a hazard. These definitions of home and work do not include populations without a smartphone, houseless populations, incarcerated populations, and/or those employed at a job not regularly attended during the prescribed time period on weekdays. These populations may exhibit different responses than those captured by the algorithms, but are still critical for evaluating community recovery. For all of these reasons, I present this method as a large-scale data-driven approach to supplement surveys that can include more diverse populations and capture more detailed metrics related to household and workplace recovery.

This method of identifying household and work place recovery following disruptions can be applied for evaluating any hazard that may include evacuations, power outages, or transportation network closures that require large portions of the population to alter their home and work routines. Applying anomaly detection to LBS data expands the literature on anomaly detection for stochastically collected spatiotemporal data as well as extending Bayesian network and anomaly detection methodologies to an urban informatics setting. Quantifying recovery through LBS data offers new opportunities for assessing risk and response at the household level. With this kind of quantification, I can analyze recovery periods by geography, socioeconomic status, and access to essential services to evaluate possible inequities in recovery to better understand the characteristics that lead to a resilient community.

3.5 Acknowledgments

Thank you to Veraset LLC for providing the data for this work. This research is supported by the National Science Foundation Graduate Research Fellowship under Grant No DGE 1841052 and the Rackham Predoctoral Fellowship at the University of Michigan. The views expressed in this chapter are those of the authors and do not necessarily represent those of the sponsors or the data providers.

CHAPTER 4

Access and Recovery

Access to essential services, including water, power, healthcare, education, food, and cultural amenities are necessary for individuals to recover and adapt to disruptive events. Identifying the dimensions of access across these services is complex in and of itself, in addition to the task of collecting relevant data across a swath of service providers. Thus, the relationship between access and resilience is not easily quantified. I develop random forest models estimating recovery based on household measures of access to essential services prior to and in the days following landfall of Hurricane Irma in southwest Florida demonstrate the impact of access or in-access on communities' resilience, or capacity to recover. First, I estimate recovery periods of individual households and essential services facilities from cell phone data. I then use open source routing software to collect travel times by car, transit, and walking from households to identify proximity to the nearest open essential services facility in the days following the hurricane's landfall. I incorporate affordability by including social vulnerability index (SoVI) variables representing household income and housing prices for an individual's home census tract as well as car ownership and population insured. Finally, I incorporate storm impact parameters including wind speed, storm surge, and cell network and power outages. The outputs of my random forest models reveal interpretable and quantified measures of importance between observed access to essential services facilities, storm damage, SoVI metrics, and estimated recovery time. Power, cell service, and school outages all rank highly in importance, followed by measures of access to essential service facilities following Irma's landfall. These results reveal the importance of including access metrics, even over currently used SoVI measures, in estimating community recovery and resilience. This work also contributes to literature on evaluating access beyond proximity, but by incorporating multiple dimensions including travel time across modes, amenity availability, and economic access to essential services to capture the more complex but more realistic household recovery experiences across these intersections.

Keywords: location-based services data; disaster recovery; machine learning; access

4.1 Introduction

4.1.1 Disaster Recovery and Resilience

According to the United Nations Office for Disaster Risk Reduction's 2022 Global Assessment Report, "risk creation is outstripping risk reduction" [1]. The frequency and impact of natural and man-made disasters are increasing annually along with climate change and growing economic inequality [2]. The concept of resilience involves the capacity of a system to absorb, recover from, and transform after a disruptive event [17, 20, 23, 24, 25, 26]. Frameworks for evaluating resilience rely on on quantifying the disruption and recovery of a system, historically to its prior state [47, 139, 140, 141]. More recently, understanding resilience also includes capturing mitigative and adaptive transformations following a disruption to avoid "resilience traps" of exacerbating inequality and vulnerability that were inherent to the pre-disaster state [20, 108, 140, 142]. Furthermore, conversations about resilience typically fall into two distinct approaches: evaluating community capacity versus infrastructure functionality. Understanding resilience based on community capacity relies on indicators representing qualities that enable a community to mitigate, prepare for, respond to, recover from, and adapt following a disruption [27, 46]. Such qualities include generalizing the population, economic assets, housing infrastructure, governance, and environmental characteristics of a community, as reported through the Social Vulnerability Index (SoVI) [106, 143, 144, 145]. Though SoVI is highly relied upon for emergency management and disaster response, it has been criticized for internal and theoretical inconsistencies and falls short in actually helping predict or plan for disaster outcomes [146, 147, 148]. Furthermore, the direct relationship between such indicators and disaster recovery are difficult to quantify, particularly in response to discrete disruptive events rather than over long time horizons. Infrastructure resilience is often modeled in terms of the built environment, specifying individual buildings or systems such as power or transportation networks that may lose functionality during a disruption [139, 141, 149, 150]. While these models have a physical basis for evaluating resilience, they do not capture the social consequences that impact a community when function is lost. There is need for multi-disciplinary approach to quantify resilience that captures the inter-dependencies across infrastructure, social, and economic system recovery and transformation.

4.1.2 Access

Access to essential services is proposed as a solution to integrate community vulnerability and engineering infrastructure-based methods for evaluating community resilience [28]. Access

is defined as the degree of fit between a system and the users of that system and extends beyond spatial proximity, but is also a function of attributes of service provision such as availability, affordability, accommodation, acceptability, and awareness [49, 50]. Access to essential services, including water, power, healthcare, education, food, and cultural amenities are necessary for individuals to recover and adapt to disruptive events [27, 28, 46, 47, 69]. These essential services are not available at the same standard across all members of a community even outside of disastrous periods [151, 152, 153, 154, 155, 156]. High mobility and access are linked to socioeconomic development including higher income and education rates as well as lower deprivation and unemployment rates [48, 152]. When buildings, infrastructure, and transportation networks are damaged by a disruptive event, access declines and decision makers do not necessarily prioritize equity in their restoration efforts [5]. Hence, access is inherently connected to community capacity and vulnerability as well as the physical infrastructure that enables access.

Identifying the dimensions of access across all these services is complex in and of itself, in addition to the task of collecting relevant data across a swath of service providers before, during, and after a disruptive event. Thus, the direct relationship between resilience and access is not easily quantified, particularly beyond aggregated populations even though access is experienced at the individual level [33, 65]. There is a need to understand what services should be considered essential at what stages of recovery, relative to the timeline of the disaster and other recovery efforts.

4.1.3 LBS Data

Location based services (LBS) data from cell phones offers opportunity to capture the relationship between behavioral responses to disruptions and access to essential services facilities before, during, and after a disruptive event. Historically, understanding behavioral response to disasters has relied on surveys [36, 37]. But capturing diverse experiences of a disaster over a time horizon that captures all preceding and following dynamics requires an extensive sample size or focus on a very specific community as well as either the foresight to initiate a survey months in advance or reliance on retroactive reporting, which is subject to recall bias. LBS data comes from opportunistically collected user locations from cell phone interactions with a network or WiFi, thus providing a mobility profile of all included users with a sample size several orders of magnitude larger than can be feasibly obtained by surveys. Furthermore, LBS data has been demonstrated to be sufficiently representative of large populations and so can capture diverse users across a large geography, socioeconomic distribution, and path of a disruptive event despite not including any user demographic data

[7, 38, 40, 63, 76, 77, 112]. Such a vast number of observations is conducive to statistical modeling even with a large number of features, enabling new possibilities of inferential and predictive modeling to understand the underlying behaviors captured in the data.

4.1.4 Contribution

In this chapter I utilize LBS data from Veraset, LLC spanning August 1- October 3, 2017 in the Southwestern Florida counties: Collier, Hendry, Lee and Monroe. On September 10, 2017, Hurricane Irma made landfall in Southwest Florida as a Category 4 Hurricane, devastating the Keys in particular but causing widespread power outages and facility closures throughout the state [53]. Power restoration took up to three weeks in some communities and schools in Monroe County closed for three weeks, from September 6 until September 27 [92]. I identify essential services facilities from OpenStreetMaps that include supermarkets, libraries, convenience stores, gas stations, pharmacies, clinics, hospitals, and home improvement stores [94]. I apply an LBS-based functional closures algorithm to identify when the facilities were closed or otherwise inaccessible during the period surrounding Hurricane Irma as described in Chapter 2 and [157]. I also apply an LBS-based algorithm to identify user home and work locations as presented in Section 3.2.2 with validation described in Appendix B, and a Bayesian-network based anomaly detection algorithm to estimate when users experienced recovery from their home and work visit patterns as described in Section 3.2.3 and Section 3.2.4. I then calculate the travel times between each home and work location and the previously identified essential services facilities using OpenTripPlanner [158]. From these results, I determine each user's travel time from their home and workplace to the nearest open essential services facilities each day in the month of September 2017 to capture changing access to these essential services facilities. Finally, I build out a database to include these daily access metrics and recovery status as well as socioeconomic and SoVI data from user's home and work census tracts, storm variables capturing wind and storm surge, and utility data including power and cellular network outages. From this database I develop a statistical model to estimate the impact of all of these variables on recovery status to demonstrate the role of access to essential facilities on recovery over time.

4.2 Methods

4.2.1 Facility Closures

I first identify facilities that may have closed or otherwise became inaccessible as when access was restored. Lack of access may result from the facility declaring itself to be closed (e.g., due

to direct damage, lack of employees, or loss of infrastructure services such as water or power), from disruption to the connecting transportation system, or from so few people accessing the facility because of nearby evacuation or stay-at-home orders. Using OpenStreetMap [94] and the OSMnx Python package [159], I select all facilities in the state of Florida tagged with the following categories: supermarket, school, pharmacy/chemist, hospital, doityourself (e.g. home improvement and hardware stores), and library. To include as many locations as possible, I downloaded all facilities represented by points, polygons, and multi-polygons in the year 2023, which include 34% more locations within those categories as in 2017, the year of Irma's landfall. To generate a footprint for the facilities represented as points, I calculated the average footprint area of the polygons and multi-polygons that were available. I calculated a radius for each category to apply a buffer around the associated points that would create a footprint equal to that category's average area. Note that if there is a bias in the OSM data, like if smaller facilities are more likely to be recorded as points, then this may introduce a systematic error like overestimating the footprints of smaller facilities. Next, I applied the LBS-base functional closure detecting algorithm (as described in Chapter 2 and [157]) to these footprints to generate a time-series of unique users appearing within the footprints on each day in August 1- October 3, 2017. I filtered out any facilities that did not have 10 or more users on at least 19 days in the study period (e.g. had regular users at least three days per week prior to landfall). This filter removed 37% of convenience stores, 28% of coded libraries and clinics, and 25% of gas stations. As library locations are centrally published on county web pages, I wanted to ensure their inclusion and so manually added locations for seven libraries not included in the OSM data. Through this filtering process, any facilities that were constructed after 2017 would not have enough detectable visit patterns, and so would be discarded through the algorithm. Any facilities that were repurposed after 2017 may be misrepresented in the locations data and so may be a source of error in subsequent analyses.

To identify closure periods, I applied a pruned exact linear time (PELT) anomaly detection method [98] available through the 'changepoint' R package [99] to each facility time series, excluding schools (given county-level data on closure dates was available for public schools, I encoded school closure periods manually). PELT is a dynamic program that identifies changepoints within a time series based on a statistically significant change in the linear parameters of the series. For a Poisson distribution, like these time series counting user appearances, this means a statistically significant difference in both the mean and variance of neighboring segments. I applied the PELT algorithm and flagged all segments that contained September 10, 11, or 12. I determined a closure period as a flagged segment that had a mean less than both the prior and following segments and that started between September 3, 2017

and September 12, 2017. Only 24 of the 626 non-school facilities had no detected closure, zero of which were in Monroe County and only six of which were in Collier County. Of those six, four were hospitals that each showed an increase in visits in the immediate days surrounding landfall, indicating continued access to these facilities.

Below in Figure 4.1 I show a histogram of the detected percent change on either side of the identified closure segment for all facilities. The percent change is consistently quite large (greater than 40%), which instills confidence that this algorithm is successfully identifying periods of remarkably lower access to these facilities.

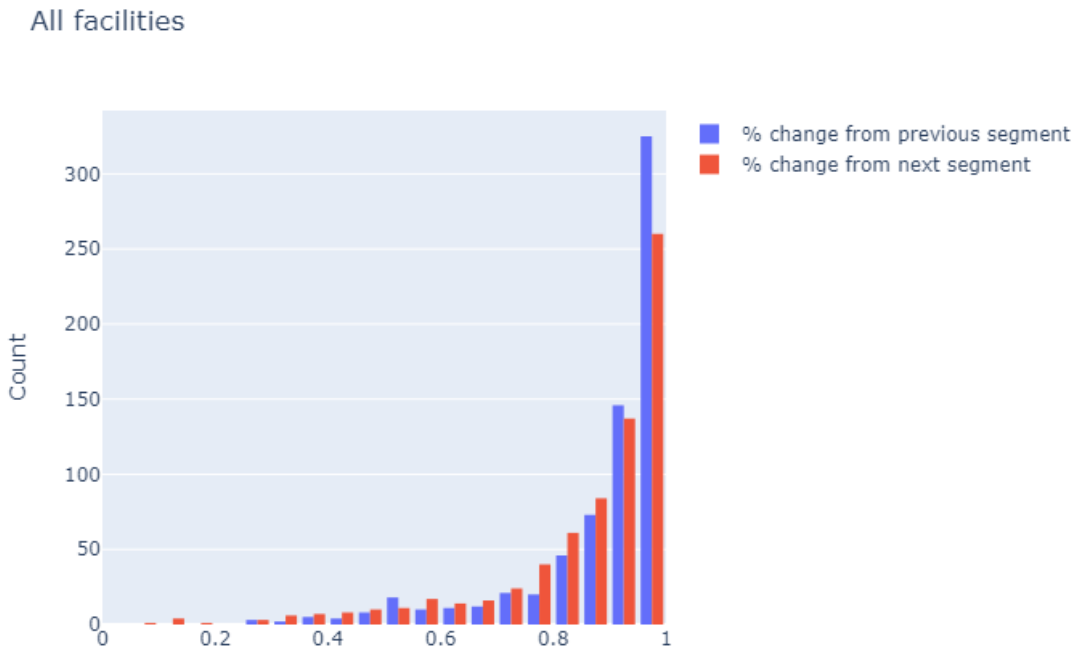


Figure 4.1: Percent change from previous and following changepoint segments for essential service facilities time series

I show in Figure 4.2 a histogram of the duration of the closure segments by county. Monroe County appears to consistently have longer closure periods.

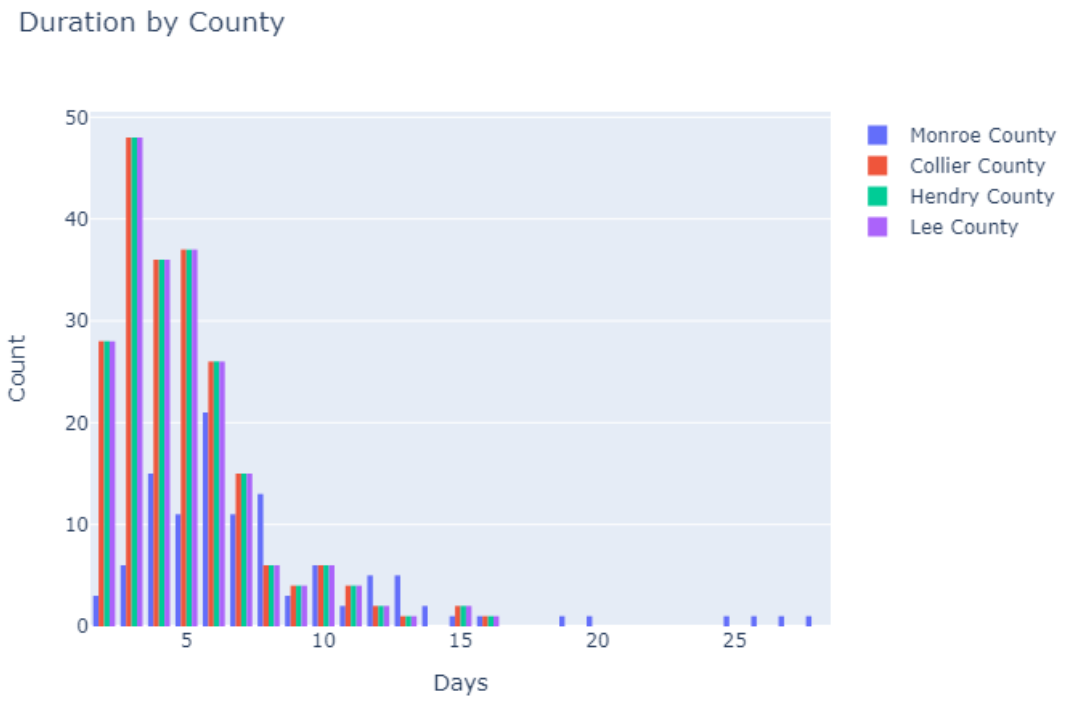


Figure 4.2: Duration of facility closure by county

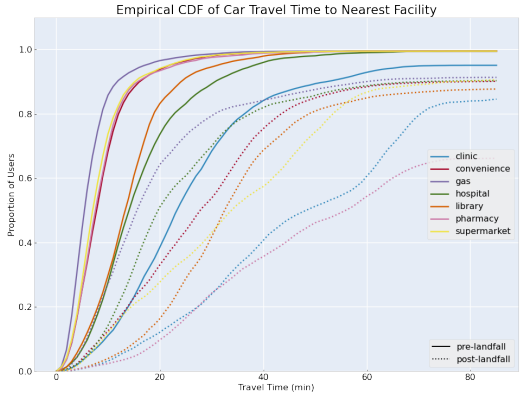
4.2.2 Household Recovery

I utilized the household and workplace recovery periods as determined in Chapter 3. As described in Chapter 3, I determined the locations of homes and workplaces from LBS data using an algorithm developed by [63], using most frequently visited places between 9pm and 6am for determining homes and between 10am and 3pm for determining workplaces. For each user, I generated time series based on whether they appeared at home, work, somewhere else in their home county, or"other" for somewhere else in Florida. I developed a Bayesian belief network incorporating day of week, whether the day was a weekend, and appearances on previous days and at other locations to estimate the probability of an appearance at home (or work) on a given day. Using each user's August data to train this probability distribution, I then applied the model to each user's September data. Any extremely low probability event (three standard deviations from users' observed August mean) was identified as anomalous and I classified series of anomalous points as anomalous segments. I then designated recovery periods as the longest of those segments with a start date between September 6, 2017 (the first day of evacuation notices) and September 13, 2017.

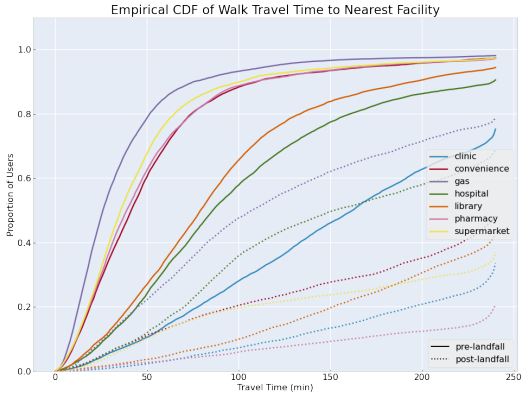
4.2.3 Travel Time

From every household and workplace, I found the travel time by car, walking, and transit to every facility within 30 miles using the open source routing software OpenTripPlanner (OTP) [158]. I obtained the Florida road network from OSM [94]. I sourced transit route and schedule data from General Transit Feed Specification (GTFS) files from Collier Area Transit [160], LeeTran [161], Miami-Dade Transit [160], and City of Key West [162]. Note, Hendry County did not have an operating fixed-route system in 2017. For each household (or workplace) and facility pair, I determined the travel time at 12:00pm on a Wednesday. For transit, I also determined travel times for 12:15pm, 12:30pm, and 12:45pm to account for fixed schedules. I allowed up to a mile of walking for transit riders and required a minimum of two minutes to allow for any transfers. I also calculated the travel time between each user's home and workplace. From this same data, I calculate the minimum travel time by each mode to each service type and the number of each type of essential services facility that is available within 10, 20, 30, 40, 50, and 60 minutes from each home and workplace.

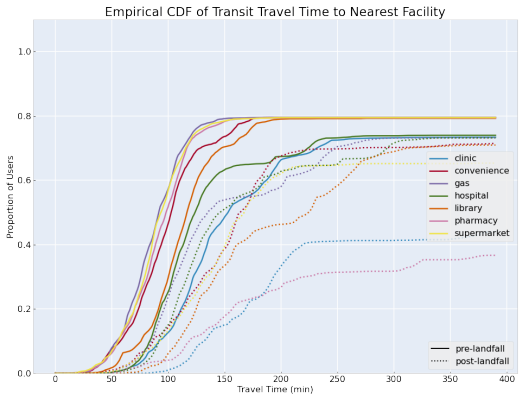
I overlay the daily functional closures results to find the nearest (by travel time) open essential services facilities over time. In Figure 4.3 I show empirical cumulative distribution plots (ECDF) of how travel time to each type of essential facility changes from prior to landfall versus September 11, 2017 by each mode.



(a) Car travel time (min)



(b) Walking travel time (min)



(c) Transit travel time (min)

Figure 4.3: Empirical cumulative distributions of travel time by mode

From Figure 4.3, it is clear that access in terms of travel time and availability drastically decreases the day following Irma's landfall. Even before Irma, transit travel times for most individuals were very high, with only about 25% of home and workplaces within 60 minutes of a supermarket, gas station, convenience store, or pharmacy and fewer than 10% of home and workplaces within 60 minutes of a library. Notably, the increase in transit travel time following Irma's landfall does not account for transit services being closed. Walking to essential services is also notably inaccessible even before landfall, as walking more than 60 minutes is not an accessible for most individuals for most trips. About 50% of users have a gas station, supermarket, pharmacy, or convenience store within 30 minutes, but libraries, hospitals, and clinics are much further.

I show the change over time of the mean travel time to each facility type by each mode

type in Figure 4.4.

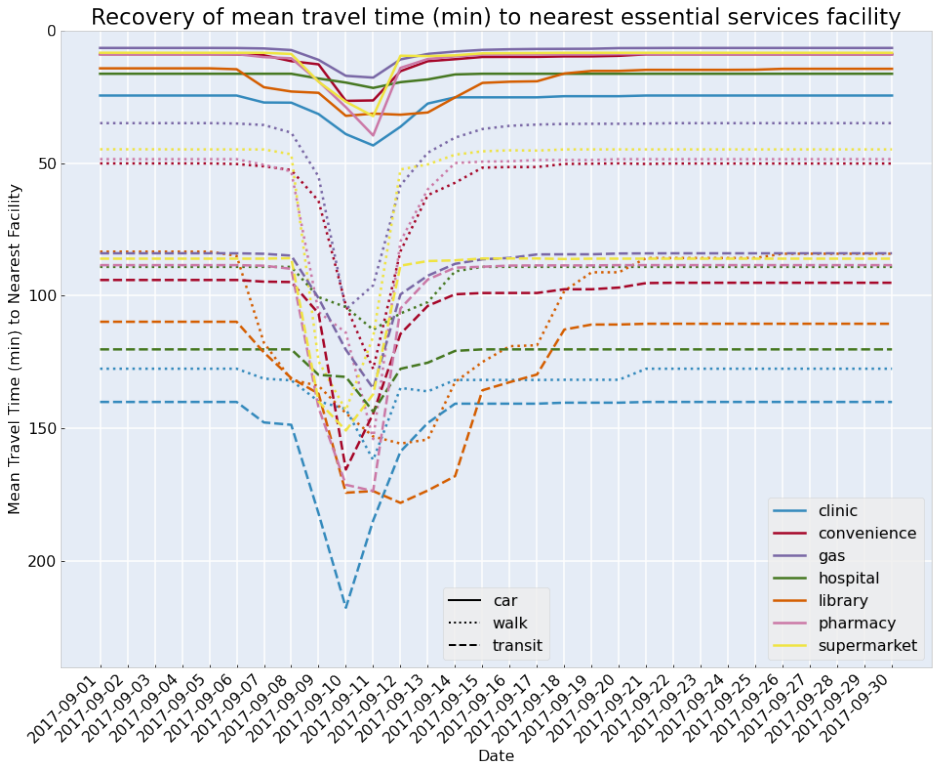


Figure 4.4: Travel time recovery curves by facility type and travel mode

Figure 4.4 shows how access to each facility type recovers in the days following Hurricane Irma's landfall, with gas stations showing the least decrease in access and quickest recovery, while libraries had relatively long recovery periods. Walking access had the greatest change between pre-and post- landfall travel times, tripling for most facility types. Recovery curves broken dine by facility type and travel mode can show decision makers differences in how access is being restored, which may be useful for determining priorities for recovering individual facilities and connections in the transportation network.

4.2.4 Statistical modeling

In order to evaluate the importance of access to essential services and recovery, I built out a dataset where each observation represented one day for each user in the month of September to estimate whether that user was in recovery that day, or not. Features included the travel time to the nearest facility by each mode and the total number of each facility type by each mode within 10, 20, 30, 40, 50, and 60 minutes. I also generated a variable to represent

the pre-landfall travel time and for the difference between the travel time to a facility on a given day and the pre-landfall travel time. Other variables include the travel time between the user's home and workplace, days since landfall (positive for days before September 10, negative for following days), whether public schools in that county were closed, the day of week, whether the location was home or work, the density of the census tract, whether the location was defined as urban or not based on the Florida DOT classification [163], the evacuation status of the user based on Florida evacuation zones and orders [164], and the county (Monroe, Collier, Lee, and Hendry). I also included variables for the previous day's state for that user and the state of the user's other location (home or work). I obtained power outage data from [165] to determine variables representing each county's service level at noon, as well as variables for time until power was restored for 25%, 50%, 75%, 90%, 95%, and 99% of the population in units of days and hours. I also included daily county-level cellular outage data (in %) based on the FCC's Hurricane Irma communications status reports [166]. I obtained variables for each location's wind exposure in terms of maximum gust wind speed, maximum sustained wind speed, gust duration, sustained duration, and maximum 3-second gust wind speeds from [167]. I also obtained storm surge estimates for Collier and Monroe Counties in the form of variables representing maximum flow speed, maximum significant wave height, maximum force per length, maximum unit discharge, and maximum depth from [168]. Finally, I incorporated variables from the CDC's 2018 SoVI [106, 145] including raw values as well as percentage estimates, percentile estimates (relative to all census tracts in the United States),"flag" variables for denoting the variable is the 90th percentile, and aggregate "rank" variables that aggregate percentile scores across the themes: socioeconomic, household composition/disability, minority status/language, and housing type/transportation. A table of all of the included variables along with their variable name, units, and source is included in Appendix F along with the associated correlation matrix.

To model the relationship between the aforementioned variables and the state of recovery, I opted to develop a random forest model. Random forest classification is an ensemble decision-tree based modeling method that involve estimating the outcome from "majority vote" determined by n individual trees. Each individual tree is developed from a random bootstrap sample of observations from the dataset, then for each branch the a random subset of features is selected and the optimal split feature is selected from that subset. Random forests are known to be highly accurate and highly interpretable. While they do not output a coefficient estimate for each input variable, they can produce the importance of individual variables relative to the other features in addition to partial dependence plots, which show the overarching relationship of each feature with the predicted outcome. Random forests are

well suited for high dimensionality data as they account for high correlation and are robust to overfitting due to their inherent random selection involved in selecting observations and features for each tree. As this dataset includes 373 variables and many of the access variables are highly correlated with each other, a random forest is advantageous for estimating the importance of the included access measures for predicting recovery.

To develop a random forest model, I used a random holdout method for cross-validation. For each model scenario, I trained the model on 90% of the data (training set), then evaluated its performance on the other 10% of the data (testing set) using outputs for accuracy, precision, recall, AUC-ROC, F1, and brier-score as well as a confusion matrix and variable importance ranking. I repeated this 10 times, where the assignment of the data to the training or testing set was random each time. I compared model scenarios based on the average across the 10 iterations.

As I only have storm surge estimates for Collier and Monroe Counties, some tested scenarios include only those counties. Other scenarios include Lee and Hendry Counties, as well, but exclude the storm surge variables. Because the observed recovery states data are zero-inflated (about five times as many 0s as 1s), I also tested various class balancing methods including manually weighting the 1s observations by five or by 25, and a "balanced subsampling" method, which balances the class observation each tree's random selection of observations. Finally, I evaluated the inclusion of different variable combinations. In some scenarios I removed all variables representing the number of facilities within 10, 20, 30, 40, 50, and 60 minutes as they are highly correlated with each other and with the travel time measures. I also experimented with removing the power variables for restoration times up to some % of the population time, as they were highly correlated with the variable representing % outage at noon. In some scenarios, I removed all transit variables, as transit access was fairly limited across most users and did not include transit service closures. Finally, in some scenarios I removed the aggregated SoVI values, including percentiles and flags as well as any raw values that were redundant of percentage estimates. These scenarios are cataloged in Appendix G along with the output evaluation measures.

4.3 Results and Discussion

Cross-validation results were consistently highly accurate across all tested scenarios, with average test set accuracy ranging from 94.3% to 94.5%. The confusion matrices revealed the main discrepancy between values was their accuracy with predicting 1 values, as all models predicted "recovered" (0) correctly about 96% of the time, while predictions for "recovering" (1) were correct between 86% and 89% of the time. This slight discrepancy

makes sense, given the higher volume and variability of the "recovered" observations (which include pre-disruption observations) but is still a promising result for predicting "recovering" observations. The models that included only Monroe and Collier Counties performed slightly better in terms of ROC-AUC than those including all four counties, but again the range was small from 0.979 to 0.984. Models that used balanced subsample weighting tended to perform slightly better than the other weighting methods. For both the two-county and four-county results, removing all of the designated groups of variables yielded better accuracy for "recovering" predictions as measured by the F1 scores. With these results in mind, I trained a model using the full dataset for just Monroe and Collier Counties with balanced subsampling and without the within-x minutes, transit access, time to % power, and aggregate and percentile SoVI features. The importance ranking of this model is plotted in Figure 4.5.

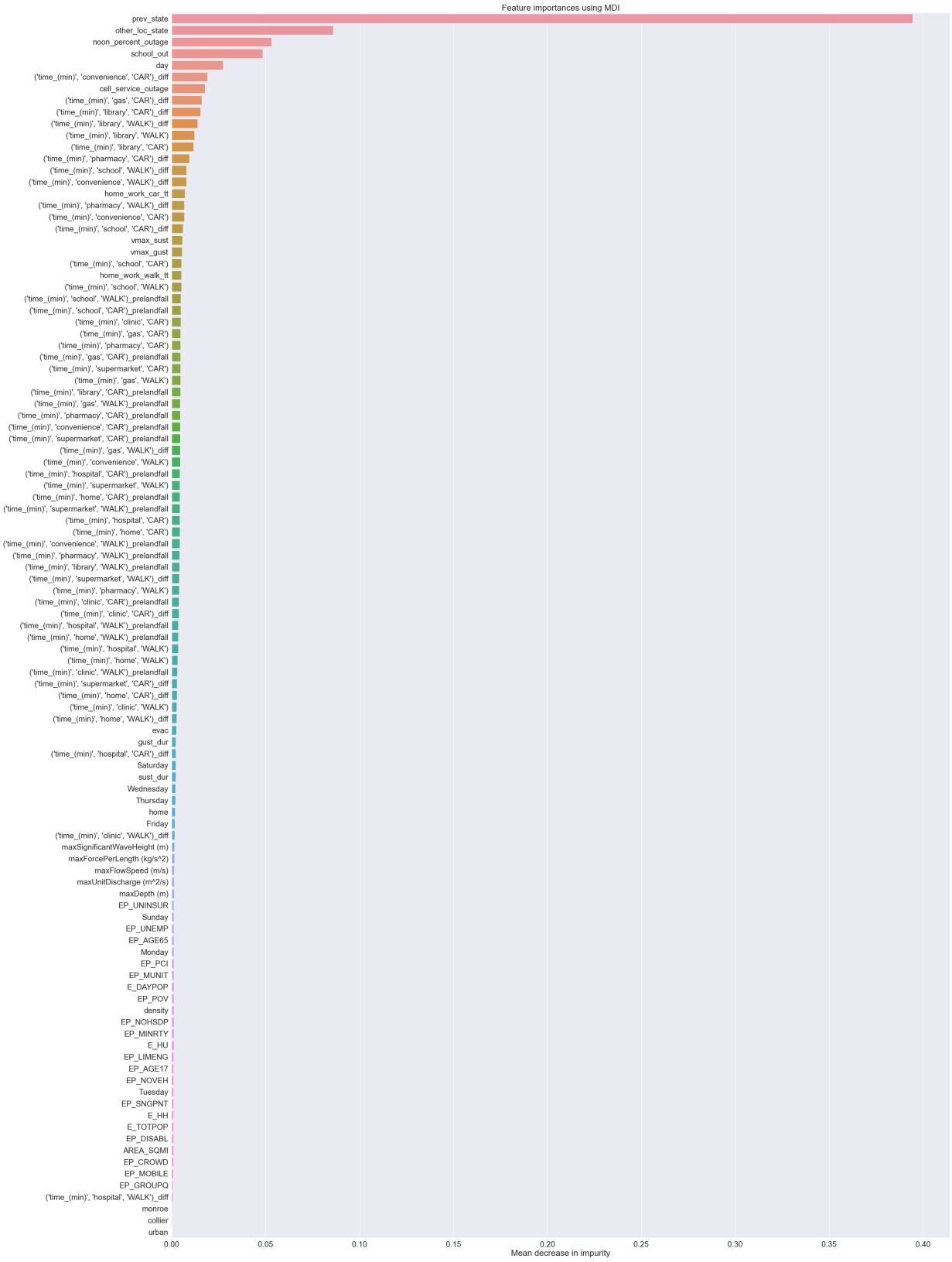


Figure 4.5: Importance for random forest model including only Monroe and Collier Counties

The importances of previous state and the user's other location state stand out, followed by power outages, cell network outages, school closures, and days since landfall. These variables are then followed by several access measures denoting the change in access from pre-event travel times. Wind variables are ranked 20, 21, and 22 followed by many more access variables until evacuation status is 57th in importance, then down further are variables representing day of week and whether the location was home or work. The storm surge variables all had very low importance, followed by the SoVI values.

As the storm surge variables are ranked relatively low in importance (perhaps due to correlation with the wind and power variables, see Appendix F), I expanded the model to the additional two counties and removed the storm surge features to see how the importance measures might change with the additional observations. The importance ranking for this model is shown in Figure 4.6.

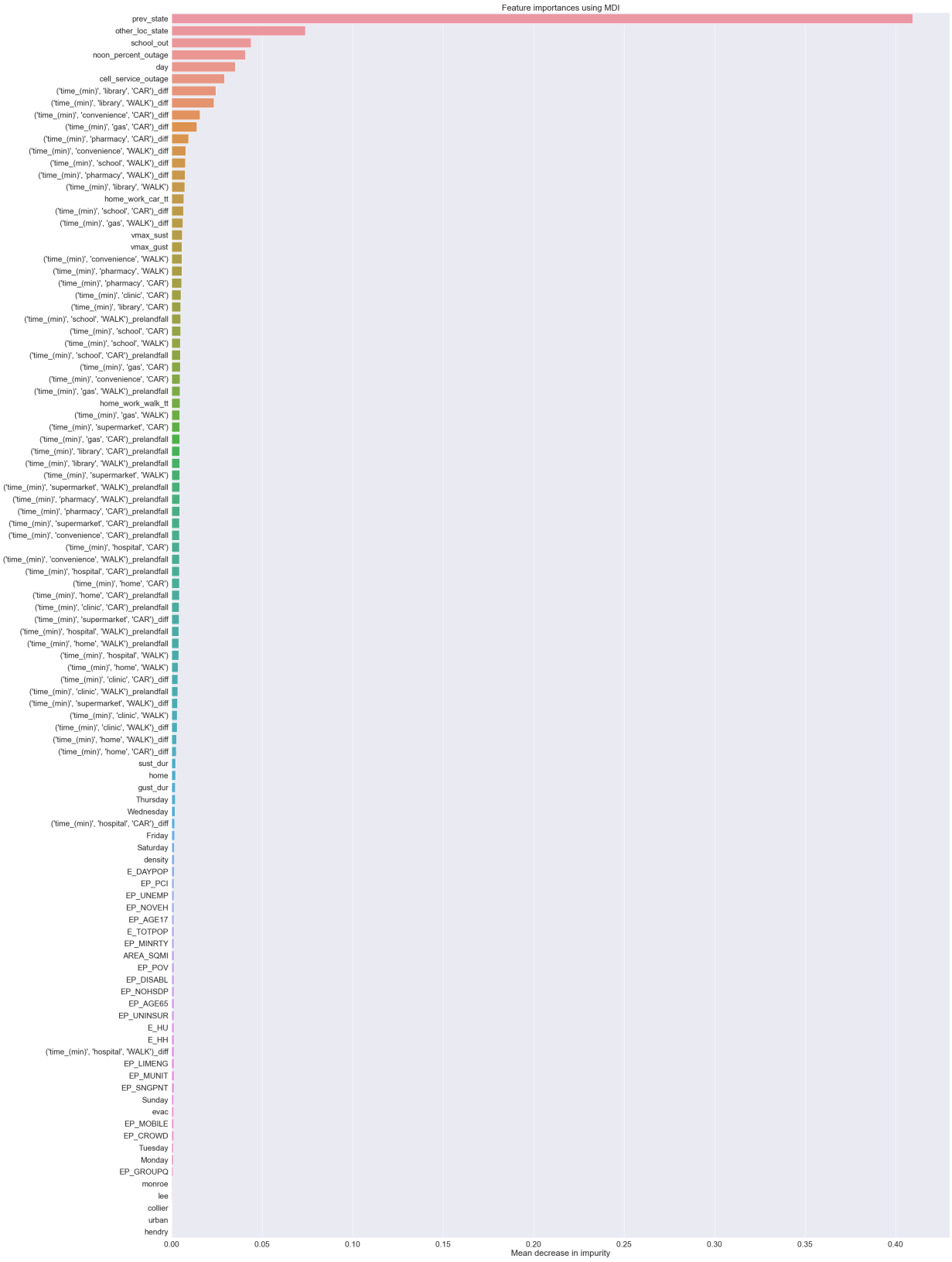


Figure 4.6: Importance for random forest model including Monroe, Collier, Lee and Hendry Counties

When expanding the dataset to include Lee and Hendry Counties, school outages become more important, moving to the third highest importance ranking over power outages and cell phone outages. The access variables shift around slightly, but for the most part have a similar importance value whether estimating all counties or only Monroe and Collier. The partial dependence plots for this model are shared in Appendix H to demonstrate the impact of each individual variable on the model's estimation.

As the importance rankings are dominated by the previous state and the user's other location state variables, I test the sensitivity to excluding each of these variables from the previously described model with all four counties. I include the cross validation results from these sensitivity testing scenarios in Appendix G. The importance values generated from excluding previous state features are shown below in Figure 4.7 and the importance values generated from excluding other location state feature in addition to previous state feature are shown in Figure 4.8.

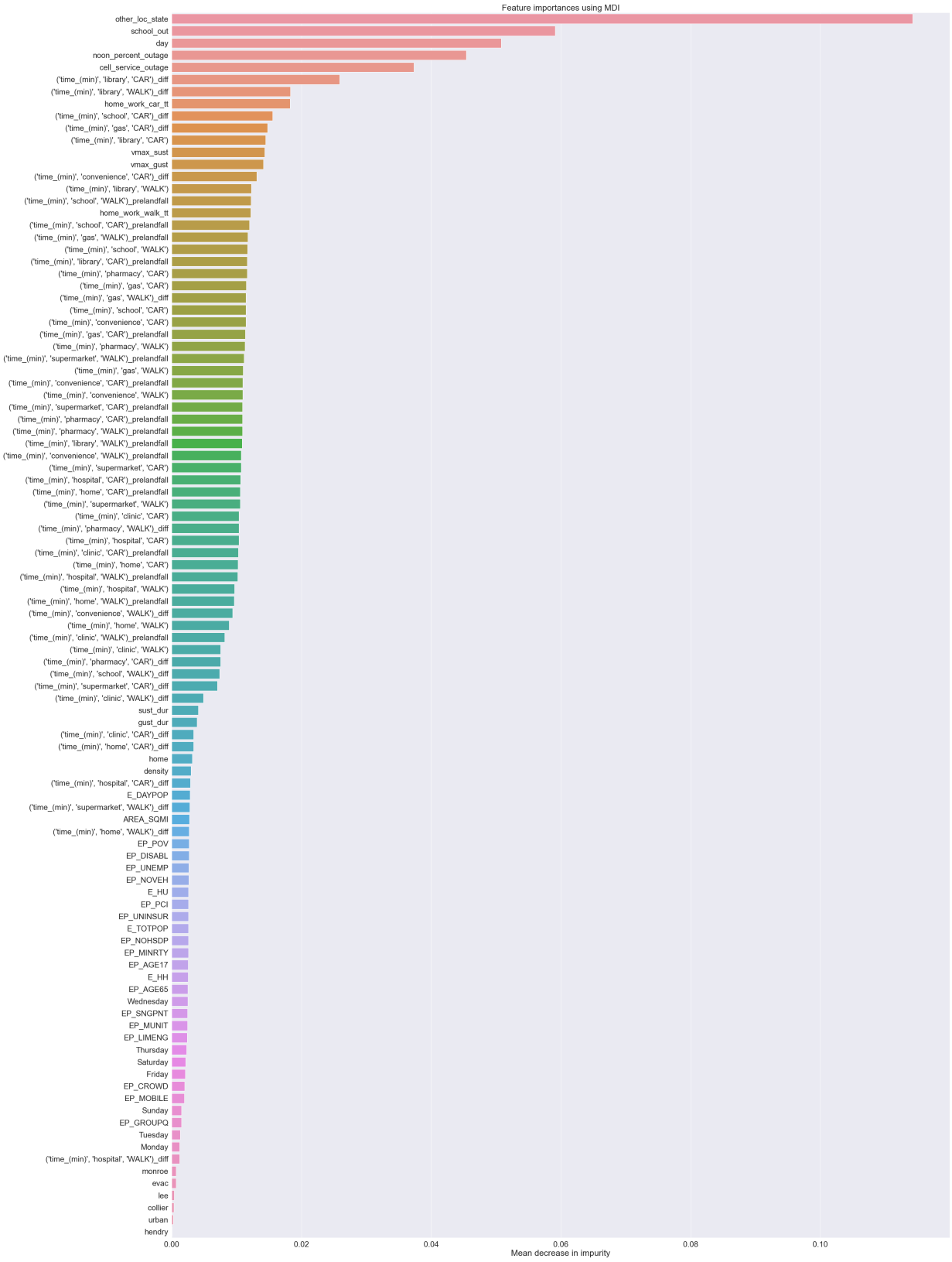


Figure 4.7: Importance for random forest model including Monroe, Collier, Lee and Hendry Counties and excluding previous state feature

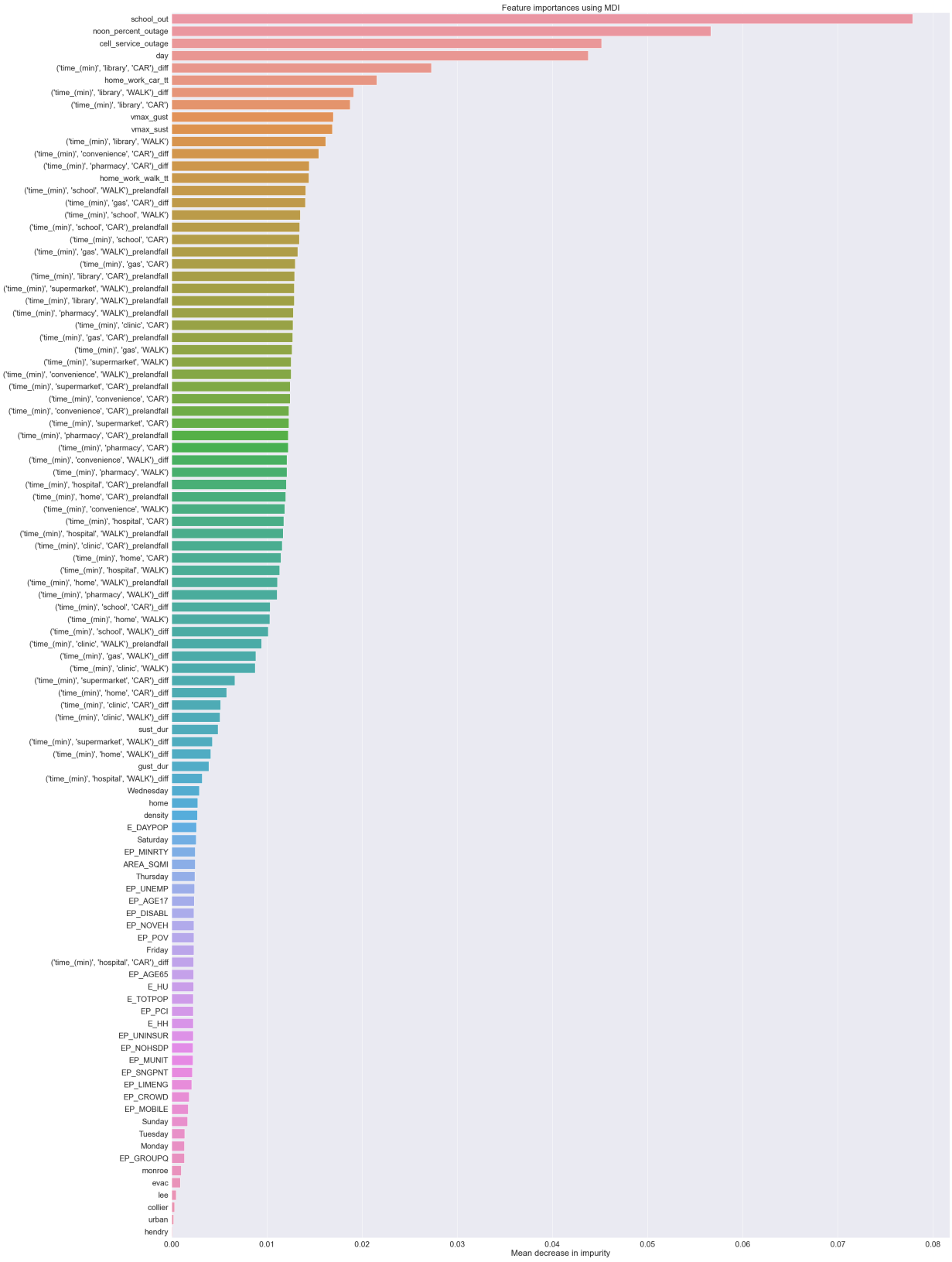


Figure 4.8: Importance for random forest model including Monroe, Collier, Lee and Hendry Counties and excluding previous state and other location state features

When excluding the previous state variable, the accuracy goes down to around 91% and the ROC-AUC value decreases by 0.019. Still, the ranking order shifts relatively little other than wind gaining slightly more importance and access measures are still positioned much higher than the SoVI measures. When the other state variable is excluded, the ROC-AUC value decreases by 0.037, but again the ranking metrics stay relatively similar other than the wind variables shifting up in importance ranking. These results instill confidence in the predictive contributions of all other included features.

Finally, I tested sensitivity to the inclusion of aggregated and percentile SoVI features instead of raw percentage estimates. The cross validation results for these sensitivity tests are shared in Appendix G. The importance rankings for this scenario are in Figure 4.9

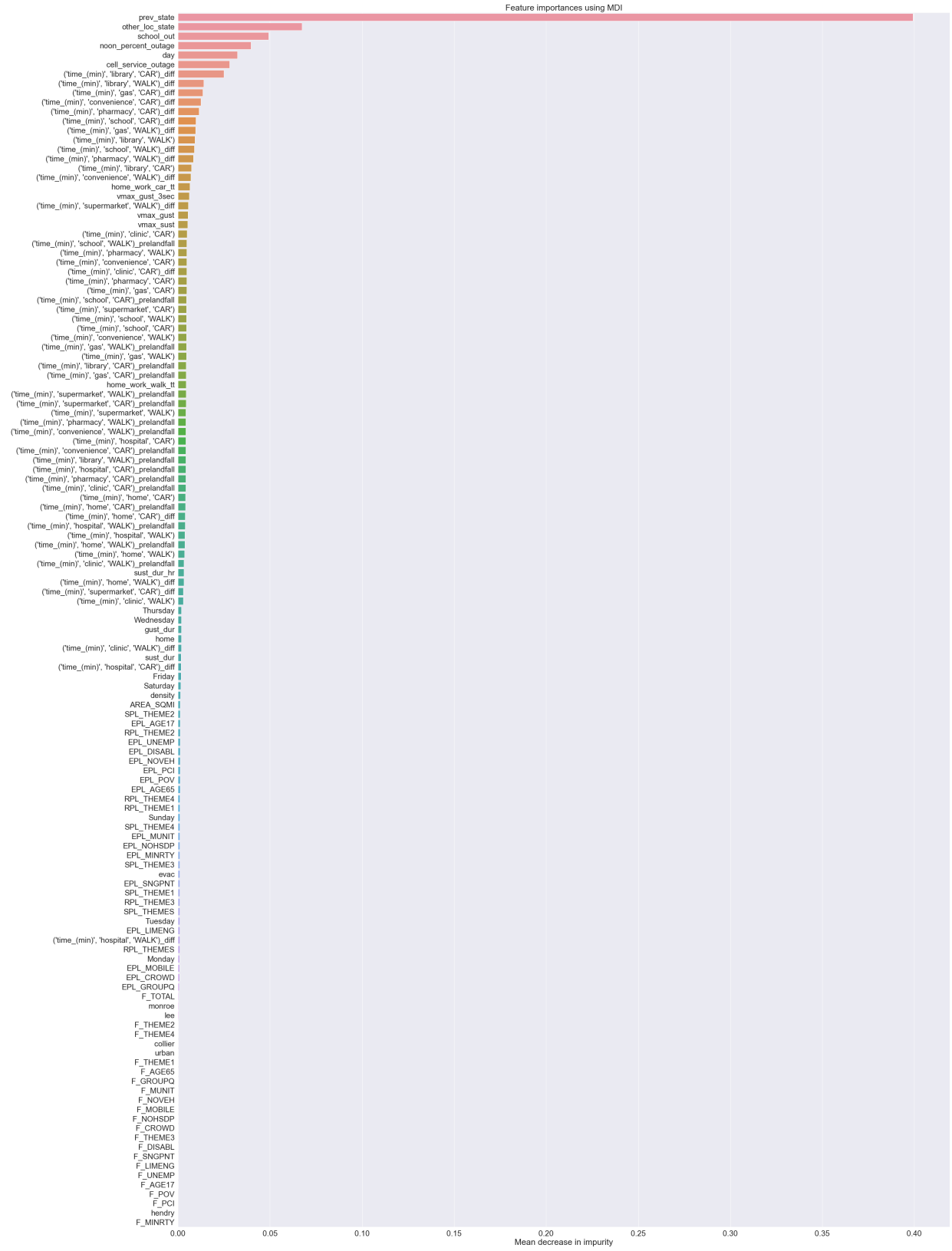


Figure 4.9: Importance for random forest model including Monroe, Collier, Lee and Hendry Counties with aggregated SoVI measures

The difference across all evaluative metrics was less than 0.01 and the importance rankings did not change, so including the aggregated measures of social vulnerability does not change their role relative to the included access measures.

Finally, I conducted statistical comparison of the full model versus a version without SoVI variables and another without access variables. A two-sample t-test comparing the ROC-AUC values for 10 cross validation sets for each model revealed a statistically significant, though slight, increase in performance when SoVI variables were excluded (p-value = 0.021). When access variables were excluded, there was a much larger and statistically significant decrease in performance (p-value $< 0.\ 001$). These results are shared in Fig. 4.10 below.

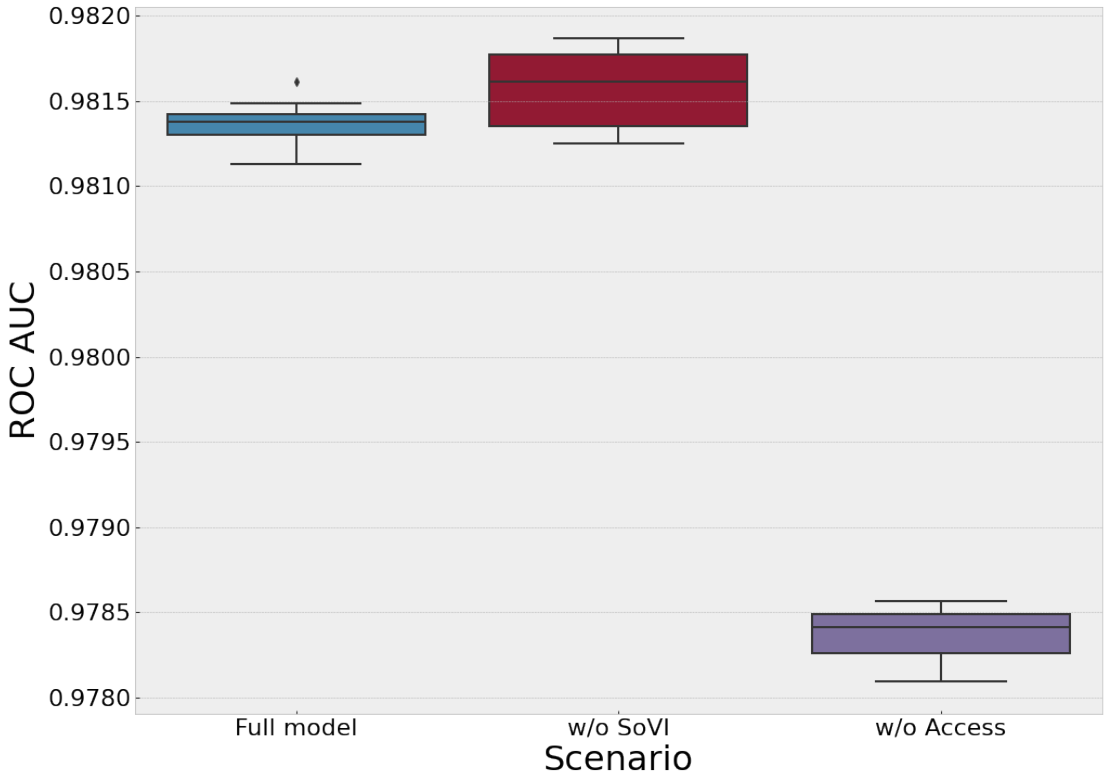


Figure 4.10: Comparison of ROC AUC performance for models with and without SoVI and access variables.

Fig. 4.11 shows the comparison of F1 scores from the 10 cross validation sets for each model. Again, the model including access variables but excluding SoVI variables performed statistically significantly better (p-value $< 0.\ 001$). The model excluding access variables

but including SoVI variables did not have statistically different F1 scores from the full model (p-value = 0.10).

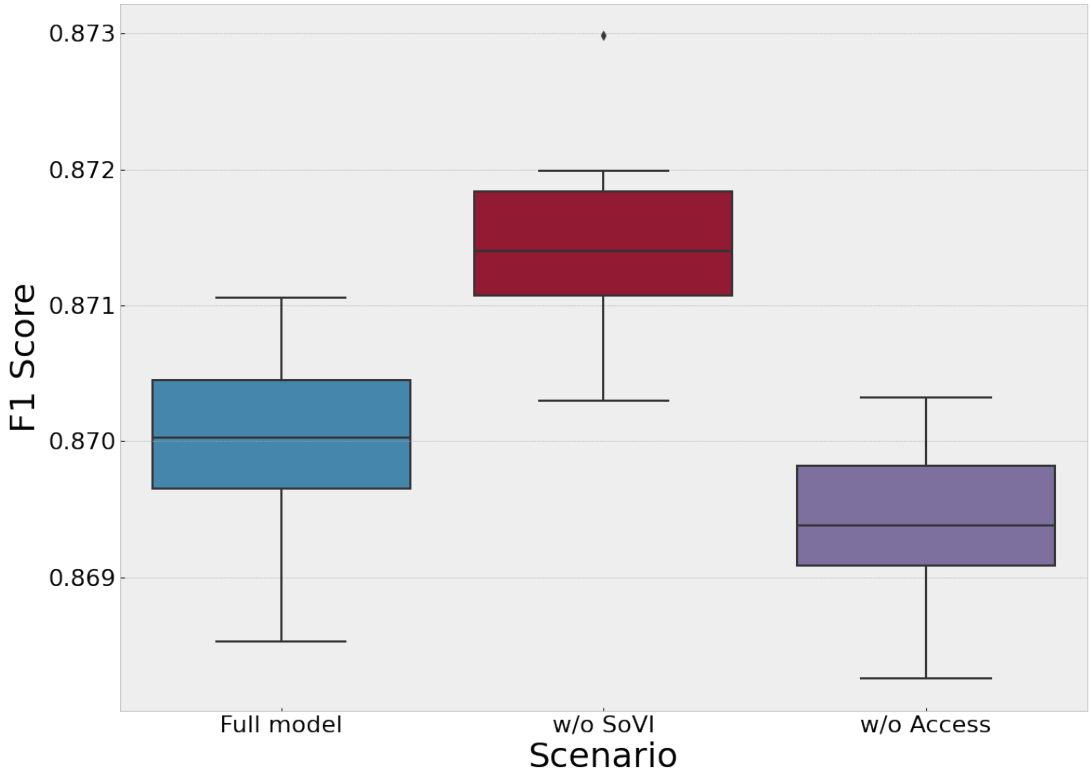


Figure 4.11: Comparison of F1 performance for models with and without SoVI and access variables.

These results indicate the value of including access variables over more traditional social vulnerability measures. These results do not discount the impact of social vulnerability on hazard experience. But, as static values that are not location nor hazard specific, perhaps SoVI values do not provide helpful information for predicting recovery. The relative contributions of any social vulnerability characteristic to resilience and recovery may vary drastically by location and by hazard [148]. The household and vulnerability in one community may not even be consistent with other communities in the same region [148, 169], and so these metrics are not useful for quantitatively evaluating resilience and recovery. Meanwhile, access variables capture the systemic inequities that manifest as difference in access that may be specific to a given community. For example, racial disparities in food access are prevalent in many communities, but race not consistently a good predictor of hindered food access

everywhere [68, 170]. Access measures captured the realized impacts of social vulnerability specific to a location and provide actionable steps for decision makers during all phases of a hazard. During the preparation phase, knowing the communities with already low access or with high likelihood of diminished access can help with prioritizing the distribution of supplies prior to an event. In the recovery phase, access metrics can guide decision making around the restoration of infrastructure and essential services facilities, as well as guide where to create temporary pop-ups for providing supplies for immediate needs. Over the long term, hazard mitigation and adaptation may involve long-term planning around equitable access to essential services through prioritizing facility openings and transportation options based on how they contribute to improved community resilience.

4.3.1 Limitations

This method is limited by the available data and tools. These results rely on open source data and software, including OpenStreetMaps and OpenTripPlanner, which may be incomplete and suboptimal. The functional closures algorithm only serves as a proxy for what facilities were inaccessible and when, the true closure periods of those facilities and of transportation networks are not known. Furthermore, this model does not currently incorporate access to shelters or pop-up suppliers that are known to provide emergency assistance in the days following a disrupted event. Restoration timelines of infrastructure networks varied highly within communities [5] but I only include power and cell network outages estimated at the county level and so cannot speak to the more granular impacts of infrastructure damage and outages on recovery. Finally, the exclusion of transportation and transit network outages is a critical missing component in this study, particularly for estimating travel times as the shortest route to an open facility as estimated by OpenTripPlanner may not be available. These limitations encourage support for enhancing open data and software platforms to achieve greater precision and accuracy, as well as for agencies and utilities to disseminate local outage and closure data throughout the duration of their restoration efforts.

4.3.2 Future Opportunities

This research inspires opportunities for developing these ideas and models further. Additional time and effort could be spent on tuning the hyper-parameters of the random forest models, such as the number of trees and the number of features considered at each node split. These models could be expanded to include interactions between variables, or could be segmented further by estimating recovery separately for home versus workplaces, urban versus suburban versus rural households, majority minority census tracts, and low car own-

ership census tracts. Workplaces could also be classified further based on the land-use of the location, especially in states or counties where land-use is highly regulated. Other access measures could also be incorporated, like the availability and accessibility of supermarkets accepting SNAP and WIC as an additional measure of affordability. Finally, this work could be extended to capture the relationship between access to essential services and recovery for other geographic regions and hazardous events. Of special interest are communities with higher transit dependence and transit service levels, which would be expected to include other behavioral responses to disruption, as was the case with transit use and recovery following Hurricane Sandy [45].

4.4 Conclusion

To help risk analysts, planners, and emergency management offices better define and act on resilience, I propose access to essential services as a critical contributor to recovery. These results presented in this chapter showcase the high importance of access measures over SoVI data in estimating recovery from Hurricane Irma in Southwest Florida. While these presented results do not differentiate much between the importance of different facility types or between walking and driving, they show that change in access may play a critical role in recovery for some communities. These results support incorporating access measures into disaster planning efforts, considering both access to essential services before an event as well as subsequent loss and recovery of access over time.

Part of the promise of reframing resilience around access to essential services is the potential for action. Mapping access necessarily incorporates infrastructure networks and the layout of a community, providing spatially explicit support for decision making [28]. With the development of digital-twins for cities, hazards simulation models, and improved forecasting models for predicting damages to the built environment comes opportunities to simulate and forecast changes in access. Such modeling can help identify communities that are particularly vulnerable to changes in access or isolation and empower planners to take action to reduce that vulnerability before hazards strike. Following a hazard, measuring and tracking access to essential services in a community can help with the prioritization of reopening [104]. Properly attributing the crucial role of access in enabling recovery is imperative for distributing funding and resources to bolster resilience and transform communities both before and after a disruptive event takes place.

4.5 Acknowledgments

Thank you to Veraset LLC for providing the data for this work. Map data copyrighted by OpenStreetMap contributors and available from https://www.openstreetmap.org. Thank you to Jiani Shen for her efforts in collecting power outage and cell network outage data. Thank you also to Zaira Pagan-Cajigas for her contribution of the wind features data and to Benjamin Nelson-Mercer for his contribution of the storm surge features data. This research is supported by the National Science Foundation Graduate Research Fellowship under Grant No DGE 1841052 and the Rackham Predoctoral Fellowship at the University of Michigan. The views expressed in this chapter are those of the author and do not necessarily represent those of the sponsors or the data providers.

CHAPTER 5

Conclusion

5.1 Summary of Contributions

Overall, this dissertation contributes to the greater machine learning literature for modeling human behavior and developing data-driven solutions for urban planning. Advances in complex systems planning and machine learning methods enable intelligent and efficient use of data such as LBS, but require a transdisciplinary approach to operationalize for effective application. My work expands literature on machine learning methods and applications for stochastically collected spatiotemporal data while drawing on urban planning and risk analysis to characterize the highly localized, dynamic, and reciprocal relationship between social vulnerability and the built environment.

For Chapter 2, I developed a novel data-driven approach for identifying facility closures following a hazard event using LBS data. I explained the process of identifying facility locations, time periods, and users of interest to generate time series from LBS data. I demonstrated this approach on a sample of supermarket, school, home improvement, and urgent care facilities in the months surrounding thee landfall of Hurricane Irma in Southwest Florida in 2017. I applied the PELT anomaly detection method to automatically identify extremely low visit patterns in those time series that corresponded with Irma's landfall. I termed these anomalous periods as functional closures to indicate in-access to the facility by LBS users, whether due to the facility itself being closed or other barriers in access, such as disruptions to the transportation network, low inventory, or perceived lack of safety. All of these barriers contribute to residents being unable to meet their resource needs, and so are important to evaluate to restore access to the communities who need it most when they need it most. I also applied this method to all clinics, convenience stores, gas stations, home improvement facilities, hospitals, libraries, pharmacies, and supermarkets available through OpenStreetMap for the state off Florida. I mapped the proportion of facilities experiencing closures and the duration of closures by county to show spatial and temporal patterns of

in-access across the state. This large-scale data-driven approach can complement smaller-scale surveys to provide broader coverage of analysis of facility availability. Knowing when facilities are available and when is a key step in evaluating resilience in terms of access to essential services facilities [28]. Such knowledge of the community-specific impacts of the availability to essential services facilities can empower local and region-wide decision making for strategic restoration of services [65, 104].

In Chapter 3 I described a method for identifying household and workplace recovery periods from LBS data using a novel Bayesian network approach. I constructed a Bayesian probabilistic network based on geographic and time-based variables to identify extremely low probability user appearance behaviors that may indicate a period of recovery from Hurricane Irma in 2017. I scaled this algorithm to include user households and workplaces across the state of Florida to reveal county-level patterns of recovery prevalence and duration as well as census tract-level analysis of Collier and Monroe counties, beyond the scope of current household post-event survey methods. This method could be applied to identify recovery periods for any sort of large-scale disruptive event that would impact infrastructure networks, public health and safety, or other factors that would lead members of a population to alter their routine home and work appearances. Quantifying household recovery through LBS data offers new opportunities for assessing risk and response at the household level based on geography, socioeconomic status, and access to essential services to evaluate possible inequities in recovery and to better understand the characteristics that lead to a resilient community.

In Chapter 4 I utilize the methods described in Chapters 2 and 3 to build out a dataset including travel time to available essential services facilities and household and workplace recovery periods. I also include data for power and cell network outages, modeled storm surge and wind speed variables, and census-tract level Social Vulnerability Indices (SoVI) metrics [46]. With this dataset and using cross validation and variable selection processes, I developed highly accuratee random forest models to predict recovery status for every household and workplace each day during the month of September, 2017. The results from these models, in the form of variable importance rankings, indicate that travel time to the nearest available essential services facilities (including clinics, convenience stores, gas stations, home improvement facilities, hospitals, libraries, pharmacies, and supermarkets) were more important predictors of recovery than evacuation status, wind duration, storm surge, density, and every single SoVI variable. I conducted statistical comparisons of the full random forest model against a model without SoVI variables and a model without access variables, and found no statistically significant difference in performance when SoVI variables were excluded, but a statistically significant reduction in performance when access variables were

excluded. These results do not indicate that social vulnerability should be disregarded in understanding recovery, but that in their current form, these index values are not helpful for estimating the drivers of recovery. These indices are static in nature and are not hazard specific, and the relative contributions of any one of these variables to resilience and recovery may vary drastically by location and by hazard impacts [148]. Vulnerability is contextual to location, time, and event and for regional-level single-event analysis, these measures do not necessarily capture the differences in hazard experiences across vulnerable groups, as what constitutes vulnerability in one community may not even be consistent with other communities in the region [148, 169]. Access, however, provides a metric that captures the vulnerability of households in terms of their ability to get to resources they need when they need them. Access varies over geography and time and is operational, so decision makers can actually change the access measures of a community through mitigation and adaptation efforts before a hazard, and restoration plans during the recovery phase. The work in this chapter motivates including access measures in hazards and resilience planning.

The interpretations from each of these facility closure detection, return-to-work detection, and inferential modeling methods can shed new light on how various dimensions of access to different essential services influence recovery. This work contributes to literature on evaluating access beyond proximity by incorporating multiple dimensions including travel time across modes, amenity availability, and economic access to essential services to capture the more complex but more realistic individual experiences across these intersections. Quantifiably demonstrating the relationship between access and community resilience can motivate improvement in access for the communities who need it most before, during, and after a disaster. These methods and results from using LBS data present opportunity to translate theory on access to essential services into quantitative modeling for localized, evidence-based, and data-driven planning around community resilience.

5.2 Future Research

Future work could involve extending these methods to sensor and internet-of-things technology data to integrate data feeds and capture more dimensions of human behavior to identify more behavioral responses to disruptive events. These methods can also extend beyond disaster response for analysis of disruptions caused by local transportation network changes, migration patterns, or even pandemics. My method for detecting functional closures from LBS data from Chapter 2 could be applied to internet of things and other sensor data. Parking lots, for example, make use of sensors and cameras that may provide data that can be converted to similar time series data to complement the data generated through LBS

to show patterns of in-access across a community. This method could also be applied to identify behavioral responses beyond facility closures, such as changes in user appearances at traffic intersections may show disruptions to a transportation network. Anomaly detection methods like PELT could be applied to LBS-based time series show behavioral response and increased or decreased traffic in response to non-hazard service changes, such as transit service levels or induced demand from added road network capacity. Future analysis could also show how this method might contribute to facility reopening policies [104].

The household and workplace recovery detection described in Chapter 3 could be further assessed over longer time horizons and with additional locations of interest for a more complete picture of LBS users' routine activities. Future work with longer-term data can also tune this method for more precise treatment of lag days, which my iteration of the model is currently sensitive to and which may contribute to over- or under-estimated results. Further, future work involving this method should investigate more inclusive definitions of workplaces, as the definition described in the workplace finding algorithm excludes populations without a smartphone, houseless populations, incarcerated populations, and/or those employed at a job not regularly attended during the prescribed time period of 10am to 3pm on weekdays. These community members may exhibit different responses than those captured by the algorithms and are critical for a comprehensive understanding of recovery.

Both methods in Chapters 2 and 3 should be further validated and tailored to complement specific surveys for evaluated future hazard response and recovery. The greatest value from each of these methods may be the capacity for comparative analysis and broader research questions that involve investigating community resilience across different locations, hazards, and time periods. Future research should apply these algorithms for the purpose of directly comparing communities' response to a hazard, or directly comparing the same community's response to multiple hazards. For example, this analysis could be used to compare behavioral responses between Hurricane Matthew, which struck Florida in 2016, and Hurricane Irma to see if users or communities that went through both events exhibited similar or different behaviors and outcomes. These methods could be used to compare neighboring communities experiencing the same hazard, controlling for storm parameters, to identify differences in access and recovery that may reflect differences in local policies, interventions, or strategies that influence recovery. Such comparisons can reveal local and hazard-specific vulnerabilities in terms of essential service facility, household, and workplace recovery patterns.

Much more work is needed for quantifying the relationship between access to essential services and community recovery and resilience. The results of Chapter 4 provide a promising start, but questions remain about who needs access to what services and when? Further research, including evaluating additional communities and hazards with other statistical

models, may illuminate more descriptive location- and hazard-specific relationships between access and recovery. Additional dimensions of access beyond proximity and availability should also be considered. For example, affordability may be incorporated by identifying the food providers that accept SNAP and WIC benefits. Investigating other data-driven means for capturing affordability as well as accommodation, acceptability, and awareness should be the subject of future research. Future analysis should also consider communities with higher transit dependence and transit service levels, which may influence different responses and vulnerabilities to disruption, as exhibited by the relationship between transit service and recovery in New York City following Hurricane Sandy in 2012 [45].

Once the relationship between access to essential services and community resilience is better understood, such measures could be developed into tools for decision makers. The advent of digital twin technologies for cities offers opportunities to simulate hazards as well as policies and interventions to evaluate changes in access due to infrastructure, transportation, and facility outages. Repeated simulations could reveal locations that are particularly vulnerable to changes in access that may not stand out otherwise. Such simulations could also reveal critical facilities or transportation pathways that should be further bolstered to mitigate harm from future disruptions. Access and digital twin research could complement research around x-minute cities [171] to identify which services should be available and at what proximity to relate this planning concept of access to hazards scenarios.

5.3 Implications for Practice

Current resilience indices, including the Social Vulnerability Index as well as Baseline Resilience Indicators for Communites (BRIC), are widely used for evaluating the capacity and resilience of communities, including for the U.S. Federal Emergency Management Agency's National Risk Index [46, 143, 144, 148]. However, these measures are not event or hazard specific and lack temporal components. Further, they assume that each of the included metrics equally contribute to the vulnerability of every community. Community vulnerability is place- and hazard-specific, but these index values do not capture important contextual differences [146, 148, 169, 172, 173, 174]. Furthermore, these index values are not particularly actionable, as they reflect stationary demographic, infrastructure, and environmental characteristics of a community [64, 148, 174]. Measuring resilience in terms of access offers new opportunities for integrating infrastructure and social capacities within a community system. FEMA already has GIS-based tools through HAZUS for predicting hazard risk and impacts. The work presented in this dissertation motivates building out similar tools to include access measures for a more comprehensive and place-specific portrait of hazard risk. With such

tools in hand, planning and emergency response decision makers can shift the paradigm of recovery from evaluating property loss to evaluating access to essential resources.

The work described in this dissertation can also motivate local decision makers to evaluate access in their own communities with known context of the specific social vulnerabilities that may exacerbate risk. Resilience as a function of access presents opportunities for operationalization by local decision makers at every stage of a hazard, but requires transdisciplinary buy-in and expertise for effective implementation. For the hazard mitigation stage, access to essential services, equity of access, and redundancies and robustness of access are all quantifiable and actionable metrics that can be evaluated to determine where additional essential services facilities or transportation connections may improve resilience. For preparation, knowing specific locations with already low access to certain resources or with high likelihood of decreased access can aid the prioritization the distribution of provisions. During the response and recovery phases, anomaly detection methods like those presented in Chapters 2 and 3 may show which facilities and households are recovering and in need, and can provide a real-time picture of changes in access to essential services over time. This work also motivates the need for additional real time data streams on access, as LBS data is helpful for retrospective analysis but may not be available in real time to practitioners. With this information on hand, emergency responders can know where to place temporary supply stations, while decision makers can prioritize infrastructure and facility restoration to improve access where it is most needed. Transformative sustainability planning can involve incorporating equitable access to essential services in long-range planning to evaluate facility openings and transportation options based on how they contribute to community resilience. With such knowledge at each hazard stage, decision makers can introduce people-centered, place-based interventions and policies to diminish the scale of disruption in the first place while promoting more equitable and resilient communities. To get to that point, more work is needed to determine exactly which services are essential for who and at what time. Successful fulfillment of this work will require cooperation across social scientists, planners, policy makers, engineers, and data scientists to ensure quality and inclusive data and methodologies that account for the breadth and depth of human experiences to develop effective tools and policies that contribute to a more resilient future.

APPENDIX A

Functional Closures of Sample Essential Services Facilities

For each facility category (supermarkets, elementary schools, secondary schools, urgent care facilities, home improvement facilities) I map the selected facility locations in Figures A.1, A.3, A.4, A.6, and A.8. I tabulate the latitude and longitude of these locations along with the estimated closure and recovery dates in Tables A.1, A.2, A.3, A.4, and A.5. I then show the plotted results from PELT changepoint detection for all facilities with enough unique users (excluding elementary schools) in Figures A.2, A.5, A.7, and A.9.

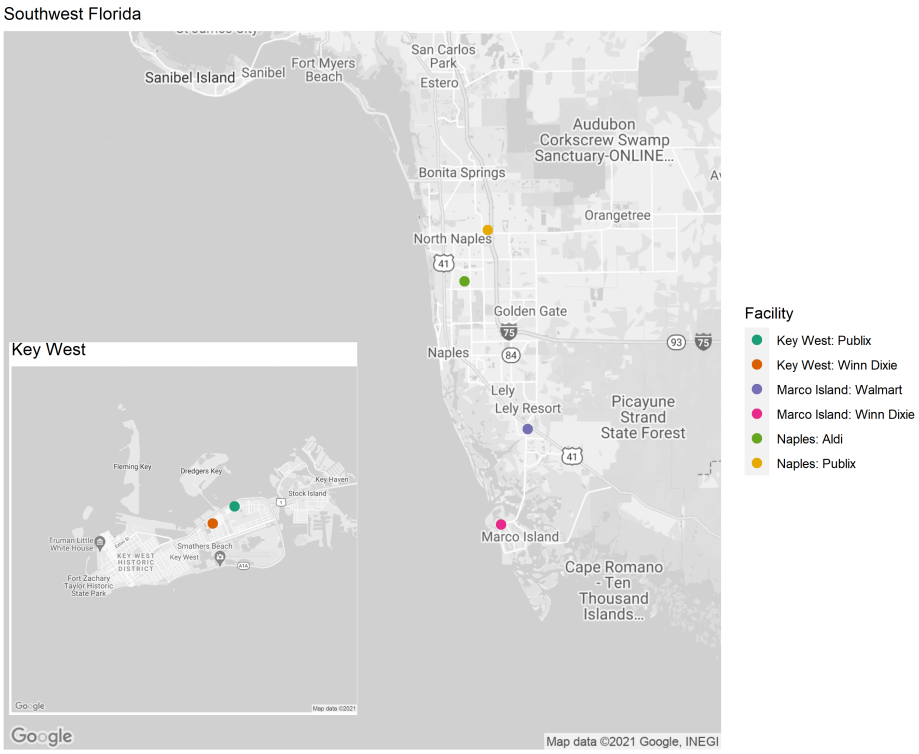


Figure A.1: Selected supermarket facilities locations

Location	**Facility**	**Latitude**	**Longitude**	**Closure Date**	**Recovery Date**	**Duration (Days)**
Key West	Publix	24.56940869	-81.76391392	9/8/17	9/15/17	7
Key West	Winn Dixie	24.56445487	-81.77081773	9/8/17	9/15/17	7
Marco Island	Walmart	26.0567864	-81.69696217	9/9/17	9/15/17	6
Marco Island	Winn Dixie	25.95182147	-81.72933746	9/9/17	9/12/17	3
Naples	Aldi	26.21890343	-81.77430144	9/8/17	9/14/17	6
Naples	Publix	26.27532576	-81.745717	9/9/17	9/13/17	4

Table A.1: Selected supermarket facilities recovery periods

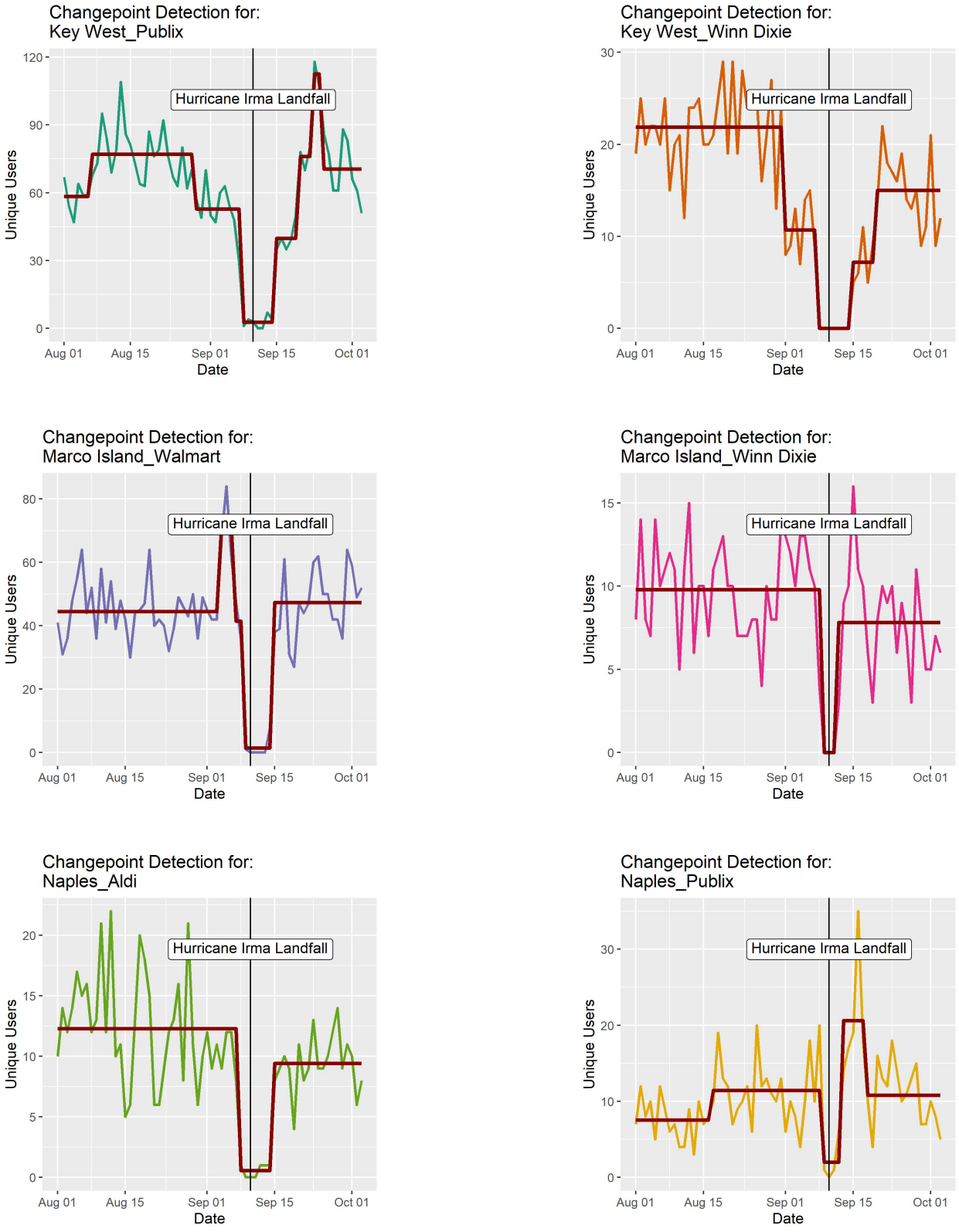


Figure A.2: PELT changepoint detection for selected supermarket facilities

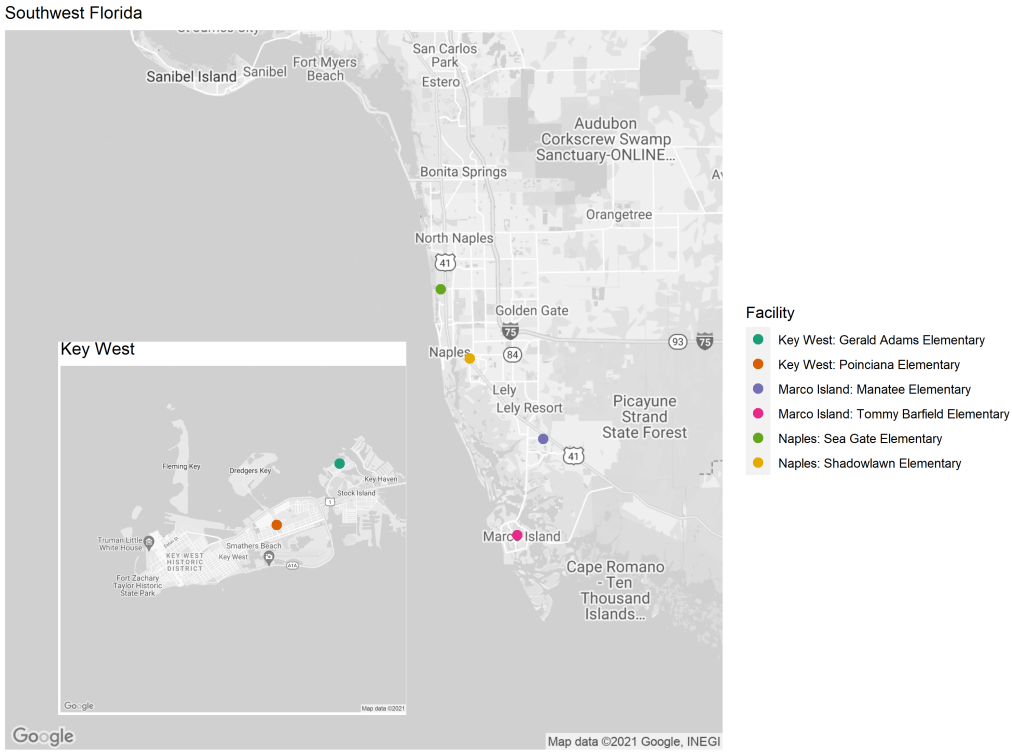


Figure A.3: Selected primary school (grades K-5) facilities locations

Location	**Facility**	**Latitude**	**Longitude**	**Closure Date**	**Recovery Date**	**Duration (Days)**
Key West	Poinciana Elementary	24.56393853	-81.76607656			
Key West	Gerald Adams Elementary	24.58165525	-81.74607789			
Marco Island	Tommy Barfield Elementary	25.93974535	-81.7111098	Not enough unique users		
Marco Island	Manatee Elementary	26.04532811	-81.67917478			
Naples	Shadowlawn Elementary	26.13361211	-81.7693005			
Naples	Sea Gate Elementary	26.20976737	-81.80497323			

Table A.2: Selected primary school (grades K-5) facilities recovery periods

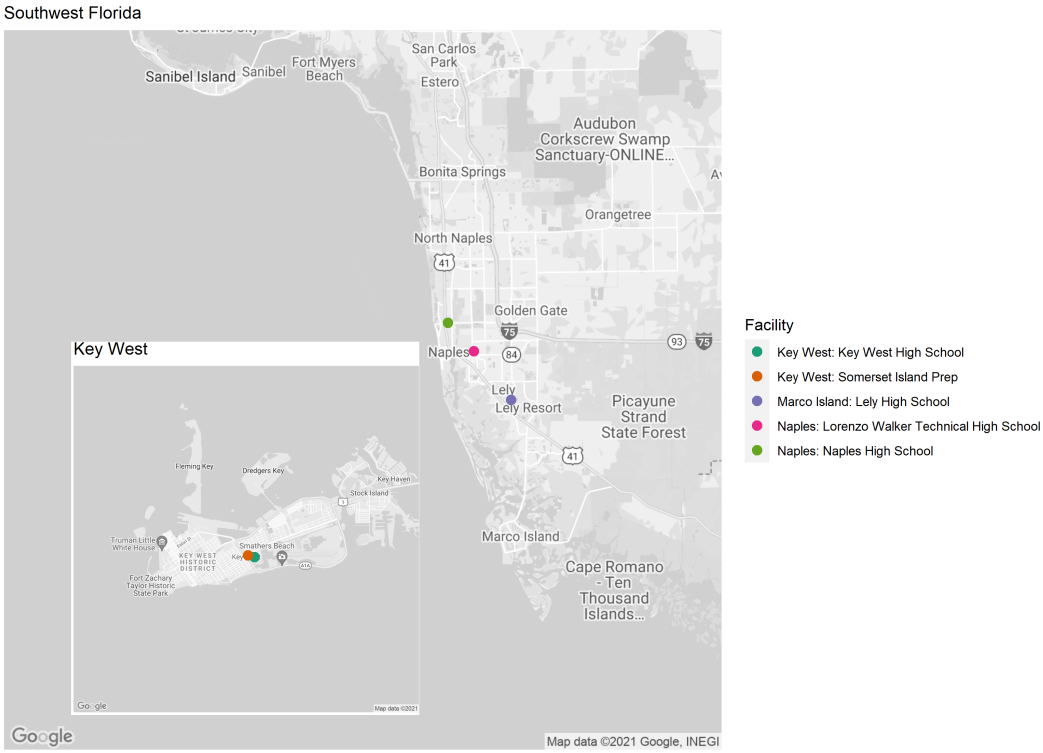


Figure A.4: Selected high school (grades 9-12) facilities locations

Location	Facility	Latitude	Longitude	Closure Date	Recovery Date	Duration (Days)
Key West	Key West High School	24.55469806	-81.77721592	9/6/17	9/27/27	21
Key West	Somerset Island Prep	24.55514959	-81.77932354	9/6/17	9/20/17	14
Marco Island	Lely High School	26.08832206	-81.71741543	9/7/17	9/25/17	18
Naples	Naples High School	26.17287689	-81.79477825	9/7/17	9/25/17	18
Naples	Lorenzo Walker Technical High School	26.14151171	-81.76286184	9/7/17	9/25/17	18

Table A.3: Selected high school (grades 9-12) facilities recovery periods

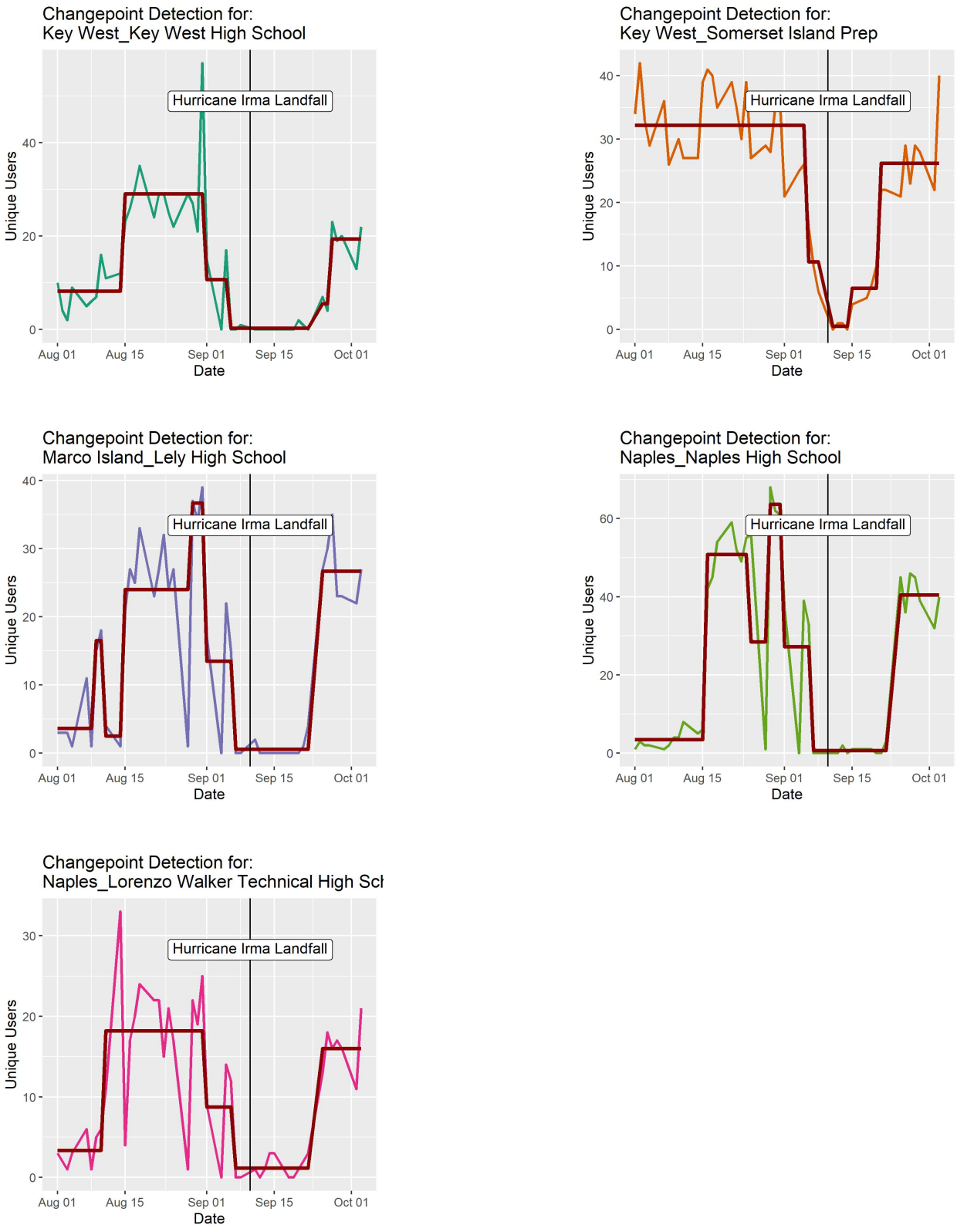


Figure A.5: PELT changepoint detection for selected high school (grades 9-12) facilities

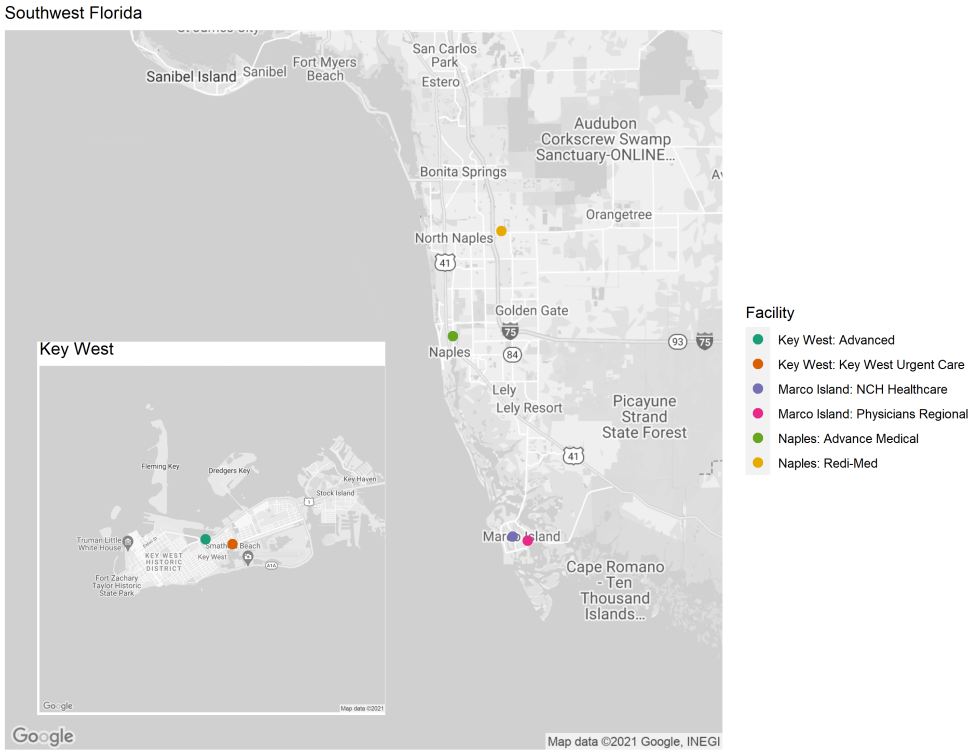


Figure A.6: Selected urgent care facilities locations

Location	**Facility**	**Latitude**	**Longitude**	**Closure Date**	**Recovery Date**	**Duration (Days)**
Key West	Advanced	24.55983526	-81.78201053	9/7/17	9/15/17	8
Key West	Key West Urgent Care Inc.	24.55842623	-81.77343188	9/7/17	9/15/17	8
Marco Island	Physicians Regional	25.93388095	-81.69799277	Not enough unique users		
Marco Island	NCH Healthcare	25.9380368	-81.71674392			
Naples	Redi-Med	26.27351711	-81.73055093	9/9/17	9/14/17	5
Naples	Advance Medical	26.15833019	-81.78983111	9/10/17	9/16/17	6

Table A.4: Selected urgent care facilities recovery periods

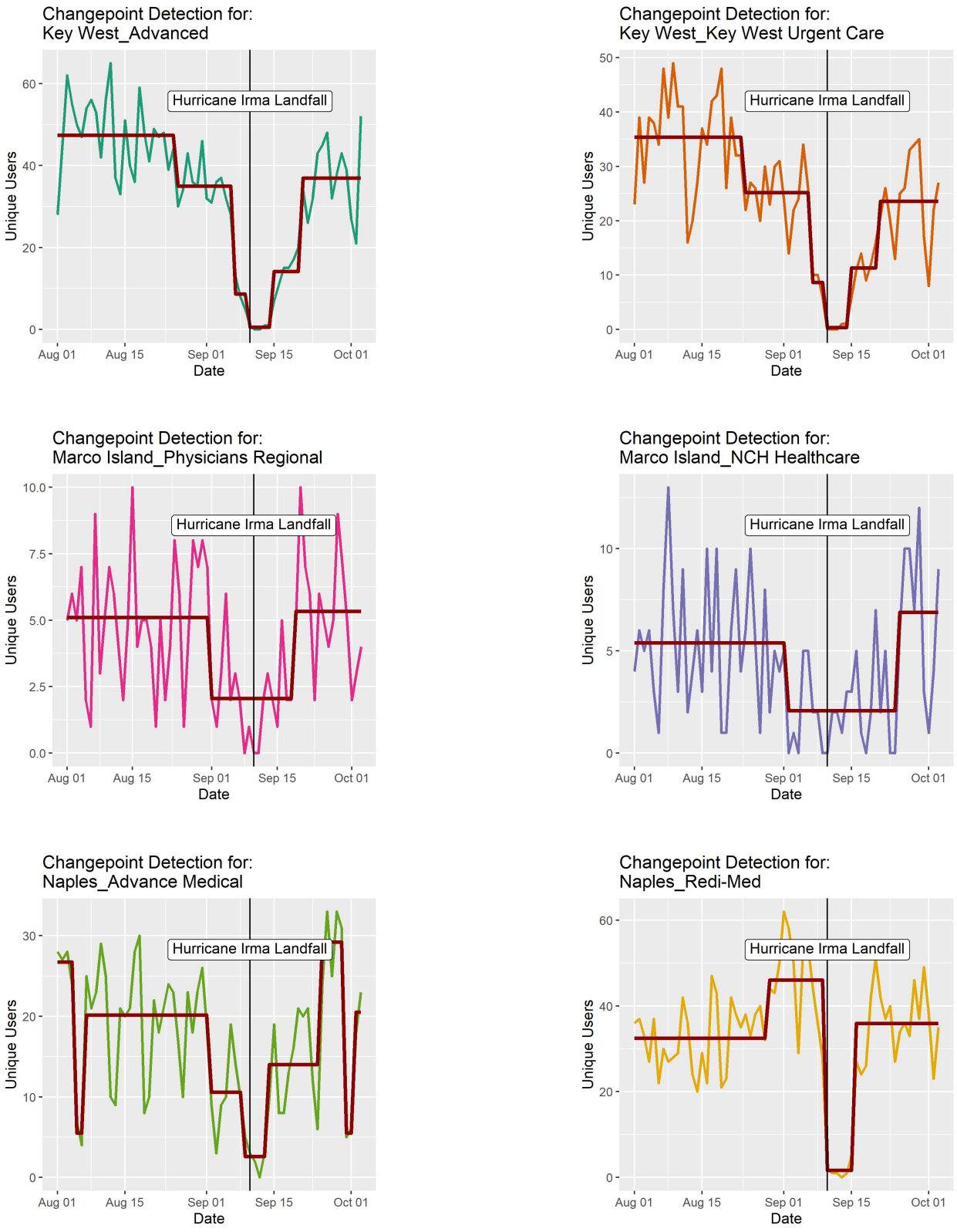


Figure A.7: PELT changepoint detection for selected urgent care facilities

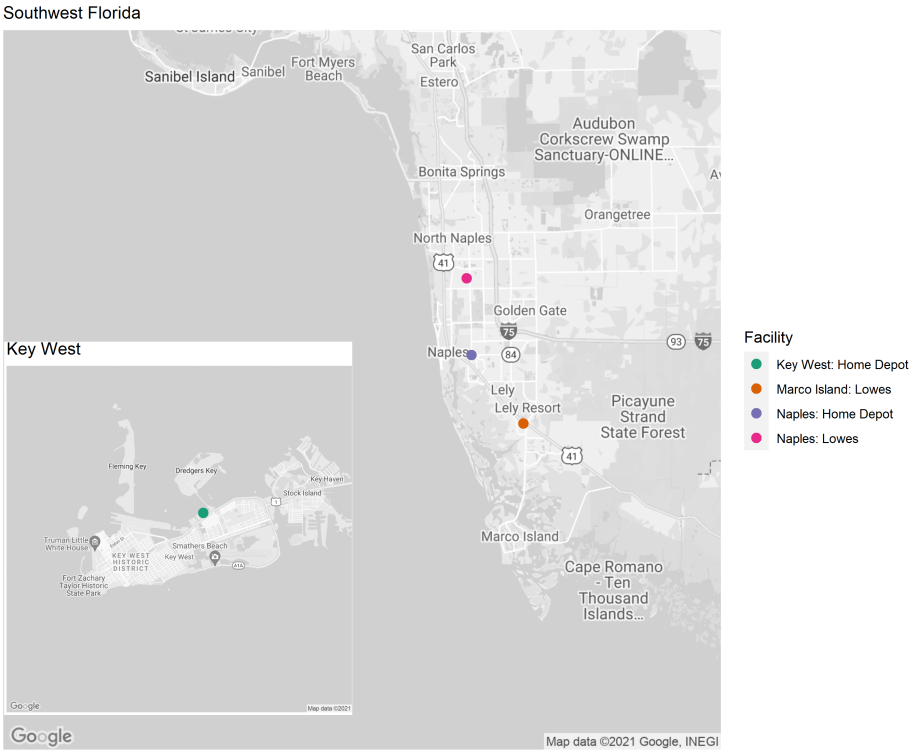


Figure A.8: Selected home improvement facilities locations

Location	**Facility**	**Latitude**	**Longitude**	**Closure Date**	**Recovery Date**	**Duration (Days)**
Key West	Home Depot	24.5674623	-81.77222628	9/7/17	9/16/17	9
Marco Island	Lowes	26.06247058	-81.70142496	9/9/17	9/11/17	2
Naples	Home Depot	26.13752046	-81.76513321	9/9/17	9/12/17	3
Naples	Lowes	26.22148867	-81.77097095	9/9/17	9/13/17	4

Table A.5: Selected home improvement facilities recovery periods

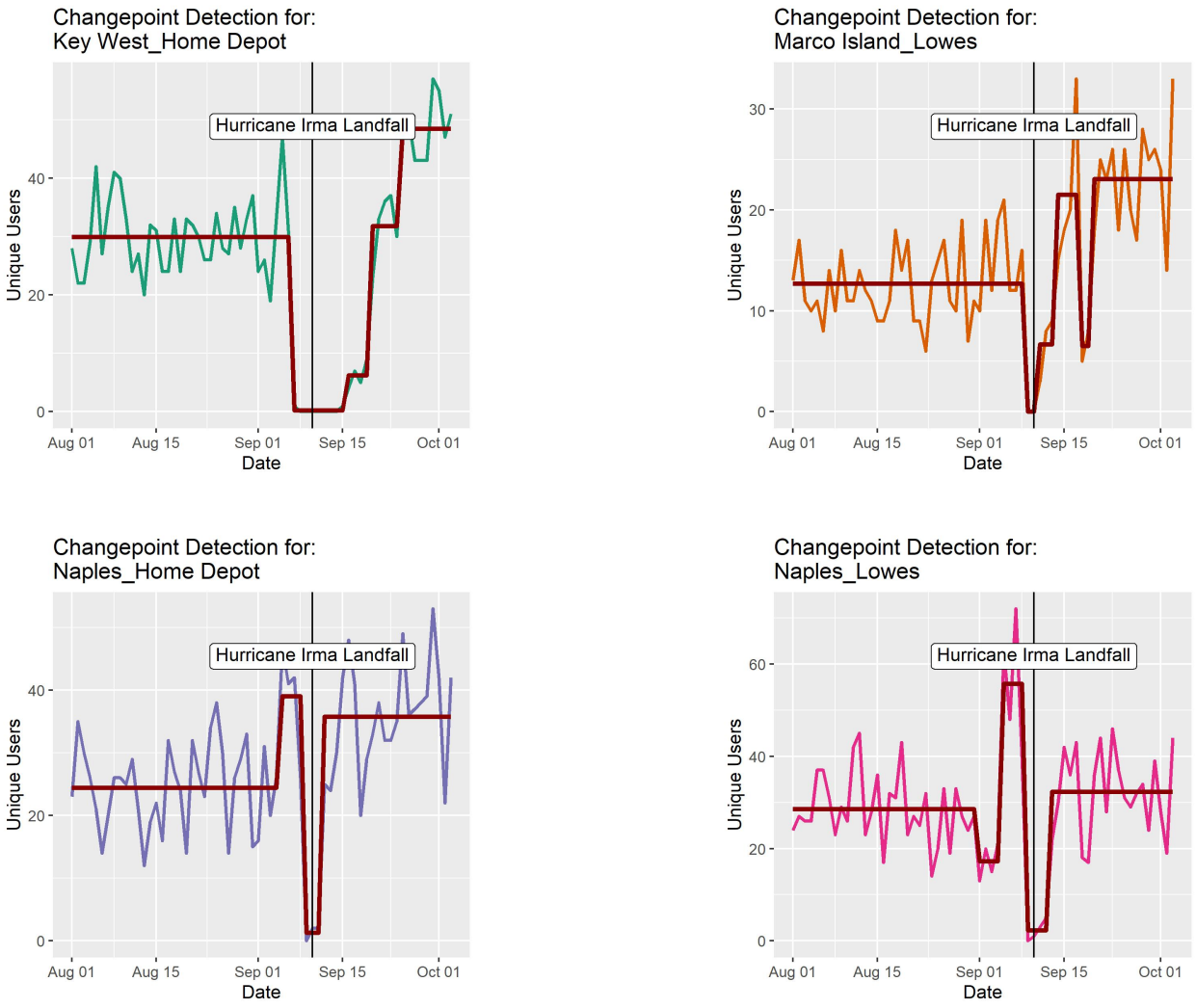


Figure A.9: PELT changepoint detection for selected home improvement facilities

APPENDIX B

Household and Workplace Finding Validation

The method I used to identify home locations is consistent with the work of (Washington, 2023) and (M. K. Chen et al., 2018) who used LBS data from the same provider and have conducted their own validations with Census and land use data to affirm this method is representative of home locations in their respective study areas. I build on this validation by aggregating the identified Florida home locations to census tracts and comparing them to the 2013-2017 American Community Survey (ACS) 5-year population estimates. I converted the home estimates and the tract population data to density in order to compare the variables on a continuous scale consistent with other academic comparison of cell phone data stay locations to ground truth (Calabrese et al., 2011). The difference between the ACS census-tract level population density versus the factored cell phone user home location density resulted in a linear trend shown transformed to a log scale in Figure B.1. The R^2 value of 0.8798 indicates 87.98% of the variability in the LBS user homes density can be explained by the actual population density of Florida census tracts according to the ACS.

Figure B.1: Validation of home finding algorithm

I compared identified work locations to the latest available Census Transportation Planning Products (CTPP) Program tabulation of worker flows based on 2012 – 2016 5-year ACS. The census tract density comparison transformed to a log scale produced a linear trend with a R^2 value of 0.9007, as shown in Figure B.2.

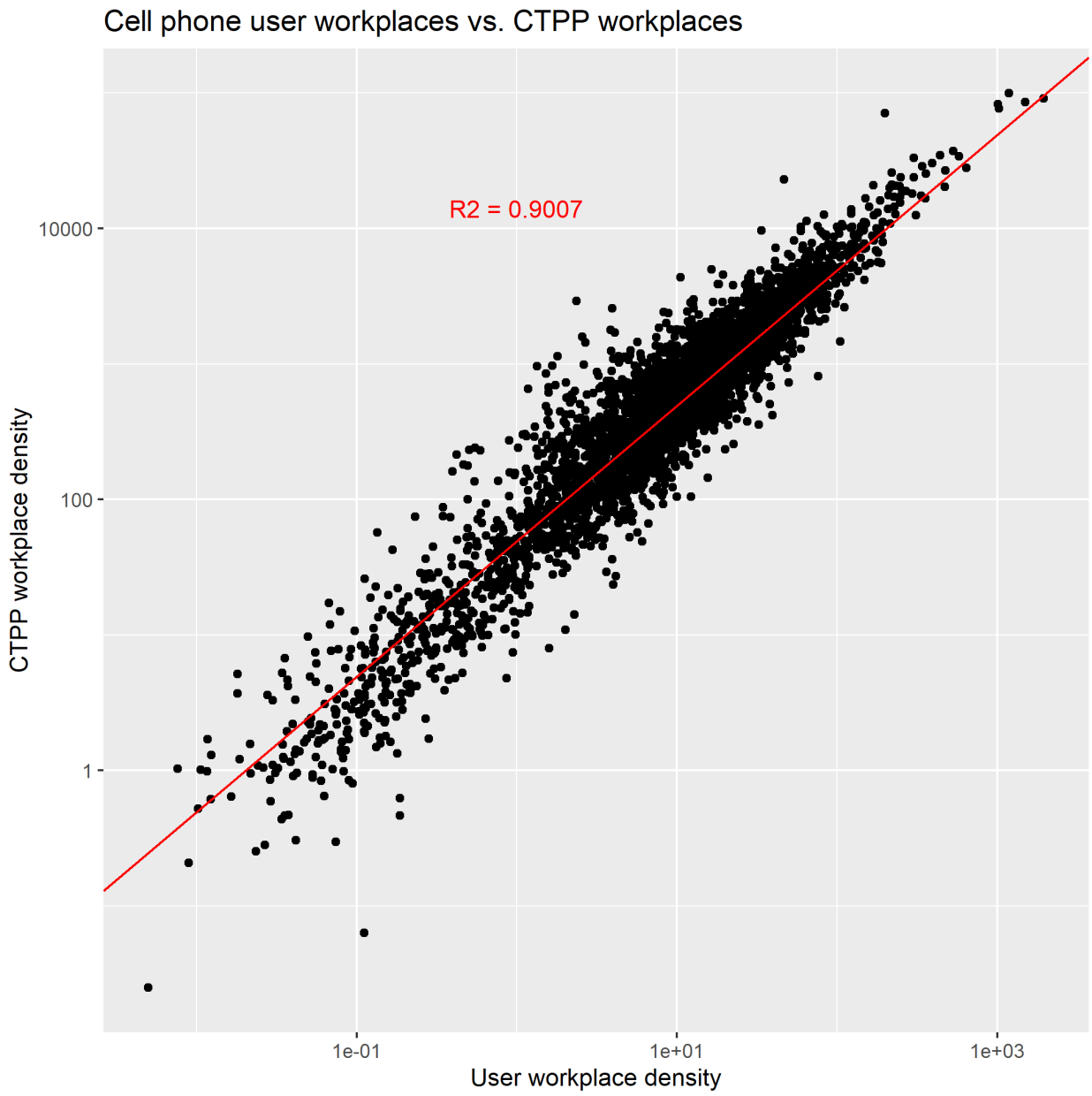


Figure B.2: Validation of work finding algorithm

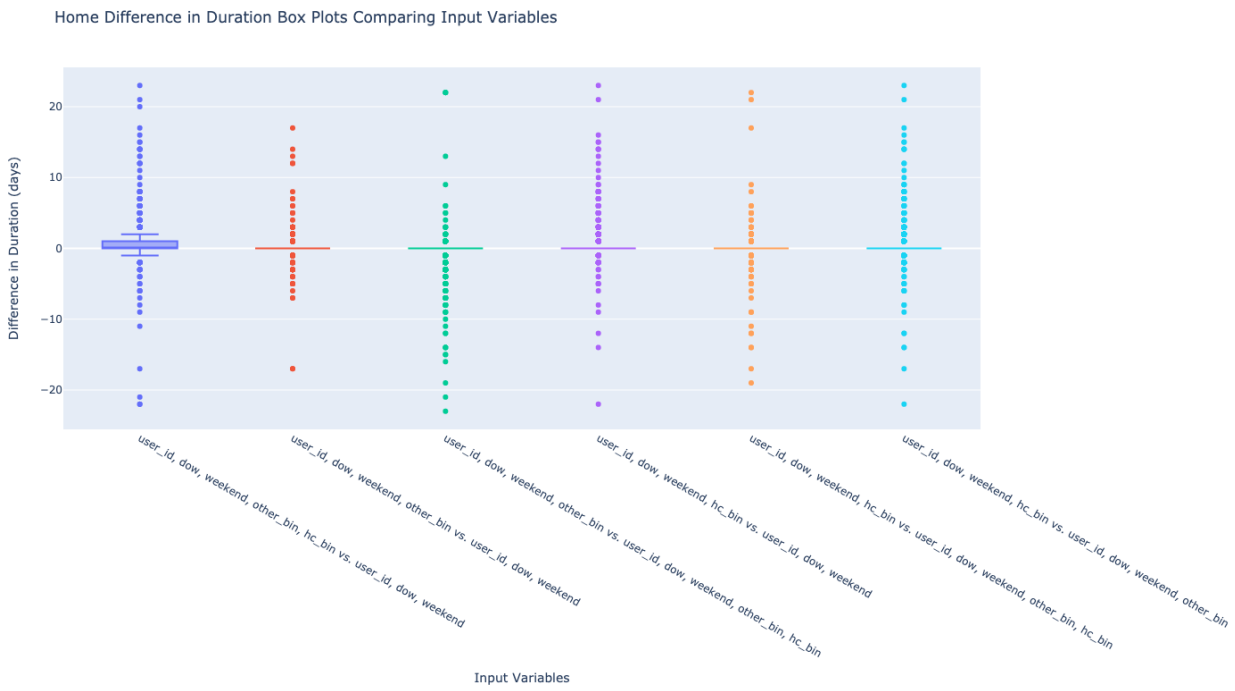
These results in addition to the those of (M. K. Chen et al., 2018; Washington, 2023) affirm that the LBS is adequately of the study area home and workplace locations to provide confidence for subsequent recovery modeling and analysis.

APPENDIX C

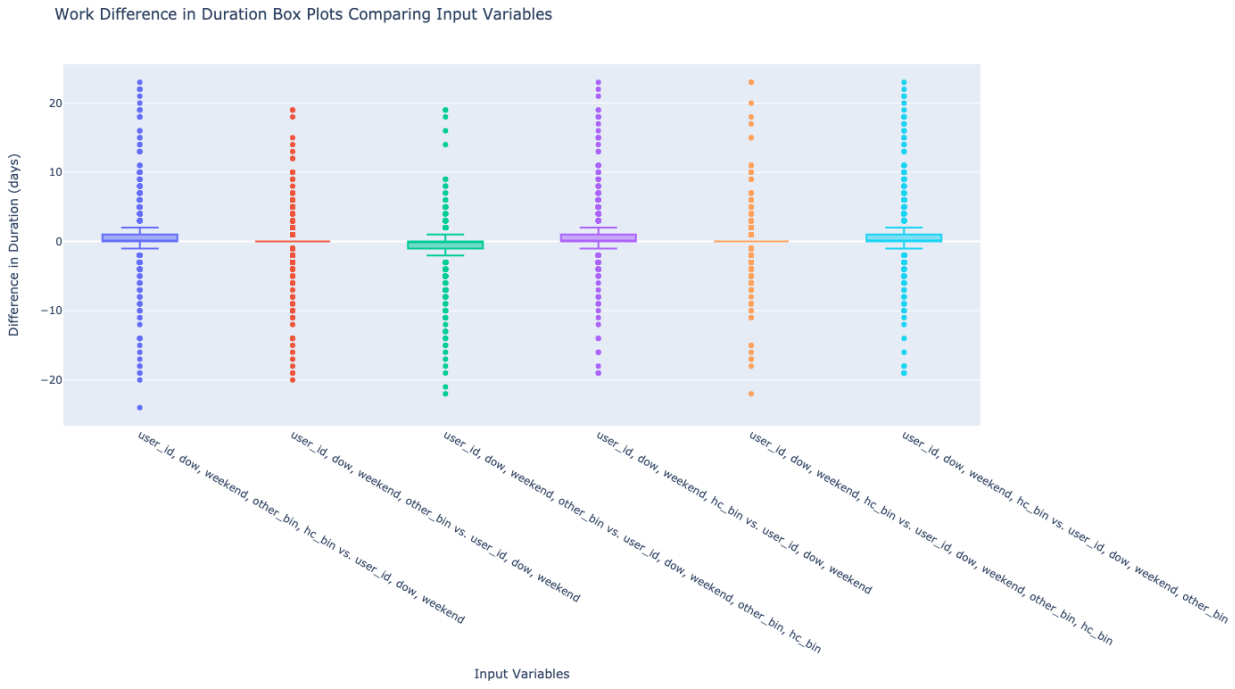
Recovery Period Anomaly Detection Sensitivity Analysis

I introduce several parameters in this Bayesian belief network including: the number of previous days to include (i.e. the number lag days) L and how much to weigh the prior data W. Additionally, I introduce variables representing days of the week, weekends and holidays, and user appearances outside of home and work but within their home county and other counties in Florida. To evaluate these parameter and variable choices, I conducted sensitivity analysis and compared the model using across different combinations of variables and parameter values. As running each combination on the full user dataset was computationally expensive, I compare the results from models run for a semi-random sample of 1000 users. I oversampled from counties in closer proximity to Hurricane Irma's path as I assumed more of these users would experience a recovery period. I weighed the random user selection process to sample at least 100 users each from Monroe and Collier county and at least 50 users each from Lee, Henry, Miami-Dade, Broward, and Palm Beach Counties.

Below in Figure C.1 I show box plots of the difference in calculated durations when different input variables are used in the model setting $L = 3$, $W = 7$, and a_u is the minimum of the user's observed August data. In the plots "dow" stands in for "day of week" while "hc_bin" and "other_bin" denote locations outside of home and work in the user's home county and in other Florida counties, respectively.



(a) Home recovery sensitivity to input variables

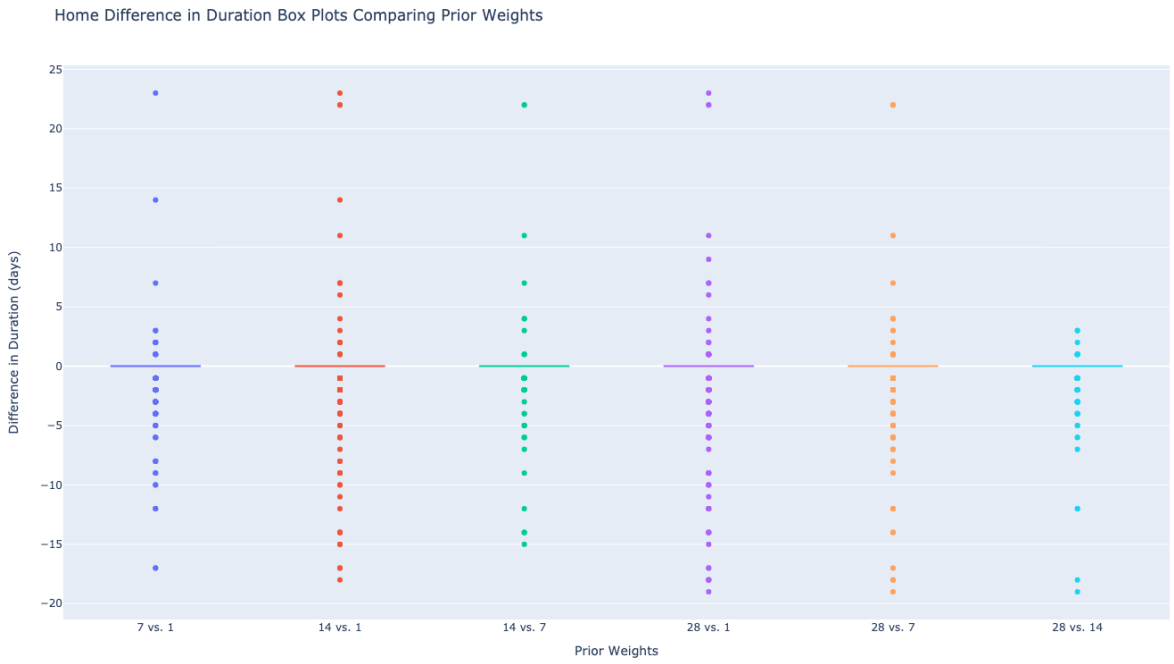


(b) Work recovery sensitivity to input variables

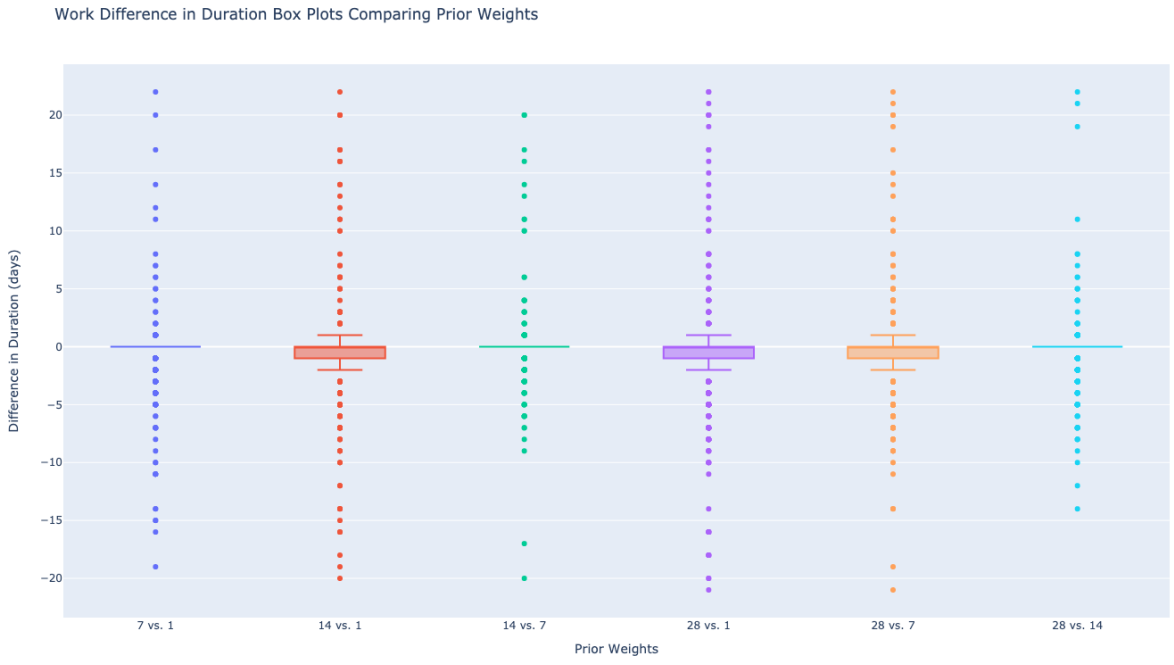
Figure C.1: Sensitivity test comparing difference in duration when modeling with different input variables

These results show generally normally distributed differences with outliers distributed evenly on either side of 0. I expect many of these differences would be reduced with the inclusion of all users' data in developing the prior rather than just the sample of 1,000 users. Still, note that the inclusion of "hc_bin" results in a wider distribution skewed towards 1 extra day of duration when estimating work recovery periods. This may indicate that changes in appearances in home county are related to changes in appearances at work. This makes sense as pings could be collected as users travel or make additional stops between home and work, and that the same barriers in access to work, like road closures or power outages, would impact other destinations like supermarkets or child care providers. I opt to include both variables "hc_bin" and "other_bin" in subsequent modeling to include as much data as possible, but note that the model is sensitive to including user appearances outside of home in work in their own home county. Therefore, the definition of home recovery period should reflect users who change behavior not just by appearing more or less frequently at home and work, but those who appear more or less frequently in their own home county.

Next, I evaluate model sensitivity to the prior weight, W holding $L = 3$, a_u as minimum of the user's observed August data, and including all variables including home county and other Florida county appearances. I evaluate weight values of 1, 7, 14, and 28. This is equivalent to weighing the prior data for 1 day, 1 week, 2 weeks, or 4 weeks in addition to each user's August data. I present the results from this sensitivity test in Figure C.2.



(a) Home recovery sensitivity to prior weight

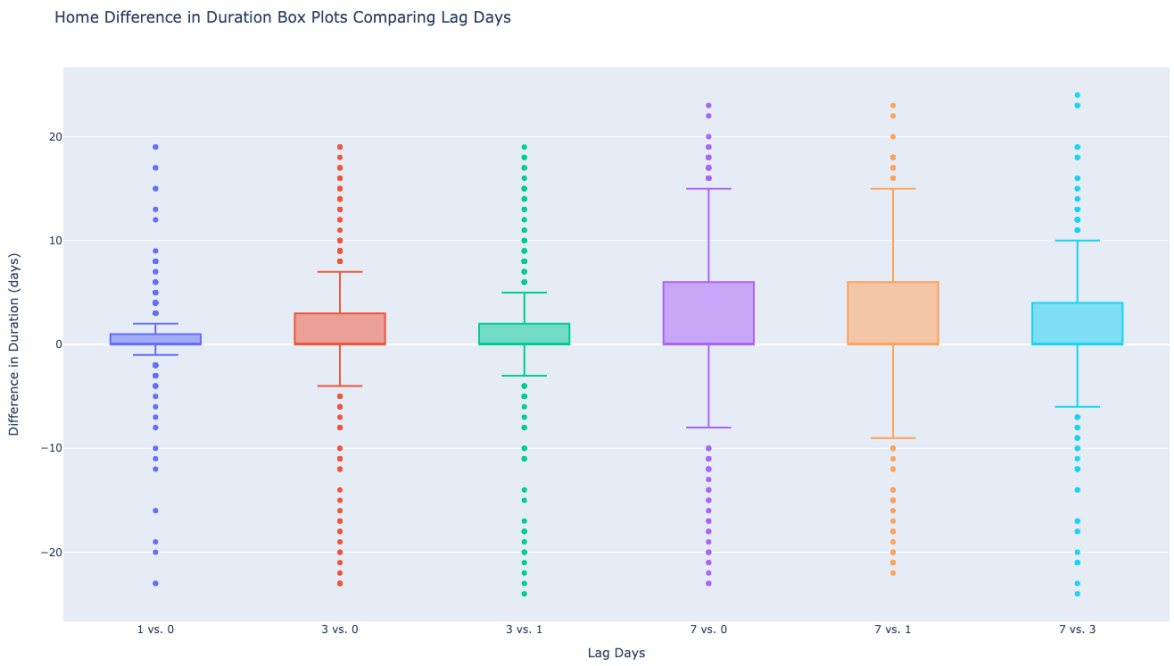


(b) Work recovery sensitivity to prior weight

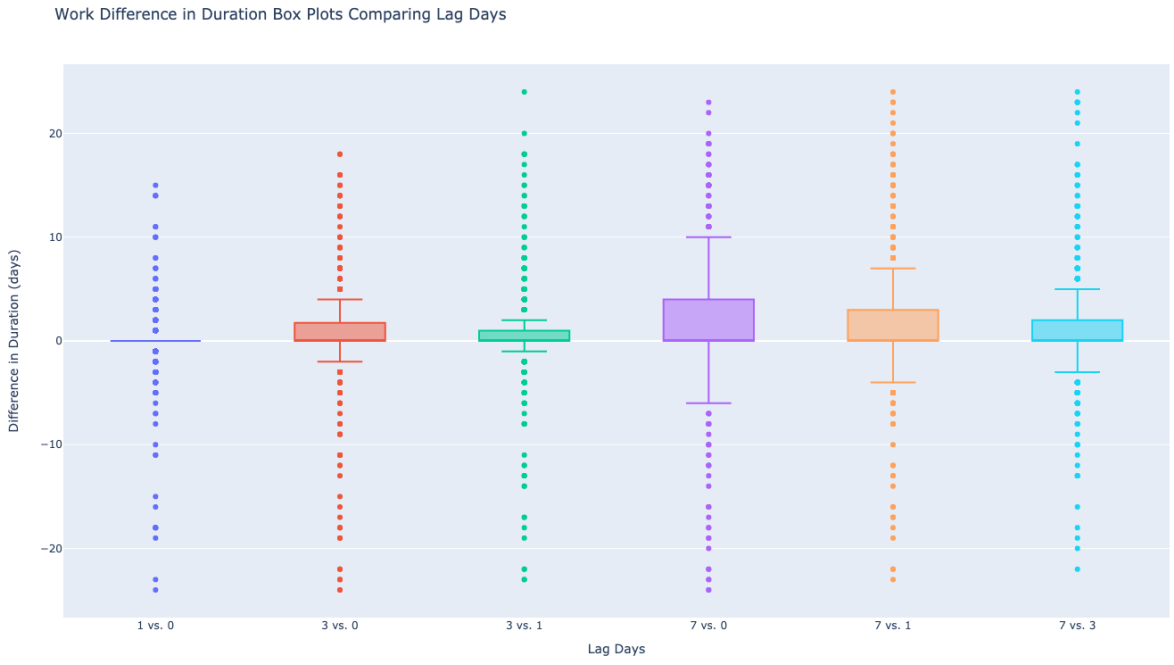
Figure C.2: Sensitivity test comparing difference in duration when modeling with different prior weight values

For home recovery, the weight value does not make a difference in detected duration for over 97% of users. For work recovery, lower weight values result in wider distributions and slightly longer estimates for duration. I elect to proceed with $W = 7$ to ensure that individual users' results are primarily influenced by their own data while maintaining the meaning of the day of week variables in the prior distribution. In other words, by setting $W = 7$ I set the prior conditional distribution for each day of the week to be weighed exactly one day relative to the user's individual daily data.

Finally, I evaluate the number of lag days to include in this Bayesian belief network. The point of including this variable is to account for expected variation in user home in work appearances. While some users may appear at home every day during the month of August, others may take frequent work or vacation trips. If a user regularly leaves their home county for a day or two, I do not want to mistakenly attribute this regular behavior to a recovery period. For example, if I include two lag days in the model, then a user who always appears at home in August will have extremely low values for the probability of appearing at home given they did not appear at home yesterday or the day before. For users who regularly leave town for exactly one night, they will have an extremely low value for the probability of appearing at home given they did not appear two days ago. For users who regularly leave town for exactly two nights, these probability values will not be as extreme, so will not reduce the joint probability as much and will be less likely to result in classification as an anomaly. For these users, identifying an anomalous period will be more heavily based on probability values from the other day of week and appearance variables. Importantly, generating a lag day variable requires at least that many days of previous data, so having 10 lag days for example means I cannot include days 1-10 for each user as I would have empty values. I show the sensitivity results for the number of lag days 0, 1, 3, and 7 included in Figure C.3.



(a) Home recovery sensitivity to lag days



(b) Work recovery sensitivity to lag days

Figure C.3: Sensitivity test comparing difference in duration when modeling with different numbers of previous days

These results for both home and work indicate high sensitivity of the model to including lag days, with home having slightly wider distributions than the corresponding results for work. The inclusion of more lag days results in higher estimations for duration of recovery period, particularly for home where users are more likely to appear every day versus at work where they are intuitively more likely to spend more days away on a regular basis. While ideally one would be able to include all seven days of lag-day variables to capture every combination of absentee days within a week, the limited time scale of this data means there are only 24 days of training data. I use $L = 3$ in the full-data modeling to maintain an equal number of each day of the week in the training data (exactly four full weeks) while still including as much past behavior as possible for each user. Given the sensitivity of the model to this parameter, future steps in the development of this method should evaluate the sensitivity to lag days using all available users and for a much longer study period to reduce the impact of losing data.

Sensitivity to this parameter will mostly affect users with low variability in their home and work appearances. I found that 52,677 users appear at home every day, roughly 43%. This value was consistent across counties, shown in Figure C.4.

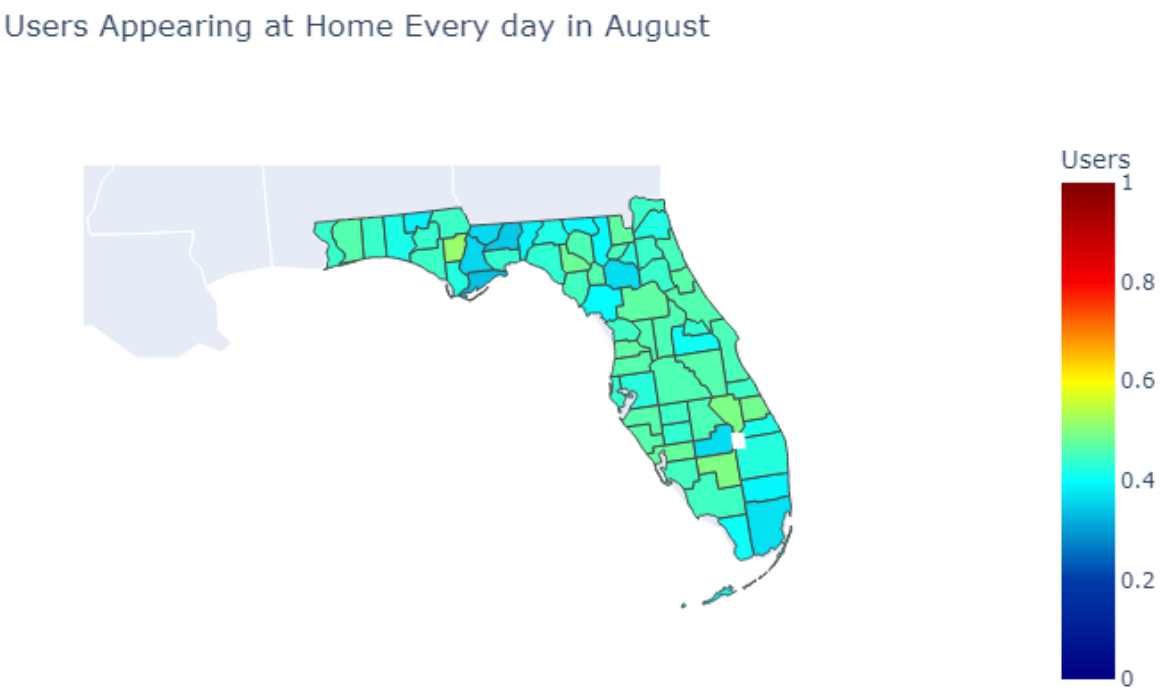


Figure C.4: Proportion of users appearing at home every day in August 2017

Most of these users still have some variability in their home county or other Florida county appearances. Only 4.5% of users also have no variability (that is, they show up every day or not at all) in their home county or other Florida county appearances during the month of August. These users would be the most influenced by this lag variable. Again, I found these users were distributed evenly across Florida, shown in Figure C.5.

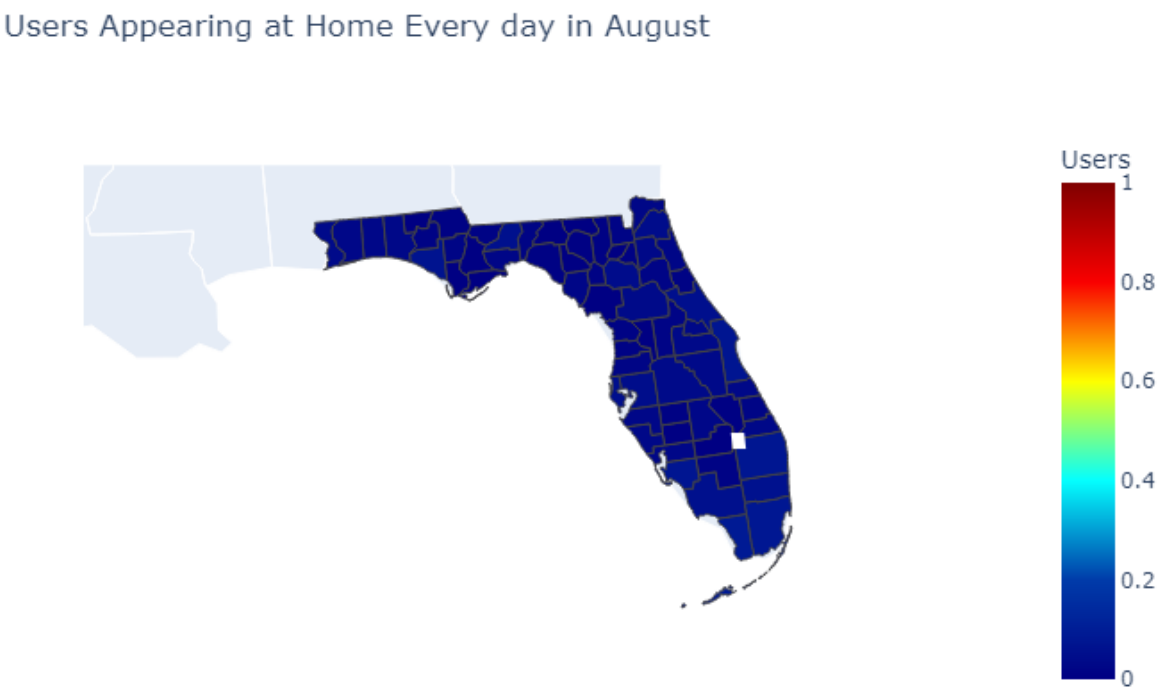


Figure C.5: Proportion of users with same daily home, home county, and Florida county appearances in August 2017

Work disruptions are significantly less impacted by the lag day variable. For work, only 1.5% of users appeared at their work location every day, and 0.2% of users appeared at work every day and showed up either every day or not at all in their home county or in another Florida county.

APPENDIX D

Recovery Period Anomaly Detection Survey Validation

In Sectioin 3.3.5, I show how the survey results from (Wong et al., 2018) compare to model results based on an anomaly threshold a_u set to 3 standard deviations from the mean of the August data for each user. In Table D.1, I share the results for every anomaly definition I tested.

Region	**Survey % Evacuated**	**LBS % with House Recovery**			
		$a_u = min$	$a_u = mean - 2 * SD$	$a_u = mean - 3 * SD$	$a_u = mean - 4 * SD$
Northeast/ Central-East	46.2%	47.1%	57.0%	46.3%	38.7%
Central-West	73.7%	52.6%	61.6%	51.4%	43.8%
Southwest	72.4%	64.9%	72.6%	64.0%	56.8%
Southeast	61.9%	62.3%	70.4%	60.1%	52.1%
Total	**57.1%**	**55.2%**	**64.1%**	**53.9%**	**46.2%**

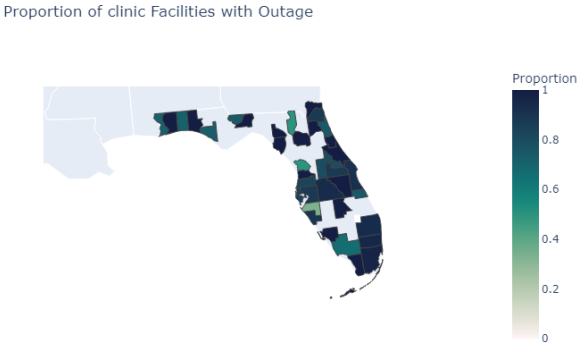
Table D.1: Survey evacuation versus LBS home recovery by anomaly definitions

Using $a_u = mean - 2 * SD$ consistently overestimates the survey results while $a_u = mean - 4 * SD$ consistently underestimates the survey results. Both $a_u = min$ and $a_u = mean - 3 * SD$ approximate the survey results well, though $a_u = mean - 3 * SD$ is slightly stricter. I elect to share results using the $a_u = mean - 3 * SD$ definition to be robust to users who may have just one extremely low point in their training data (for example, if they have a single night away from home). I expect the same general patterns would emerge if I used $a_u = min$.

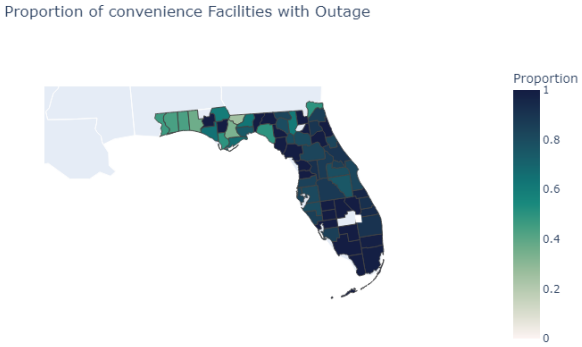
APPENDIX E

Functional Closures of Florida Essential Services Facilities

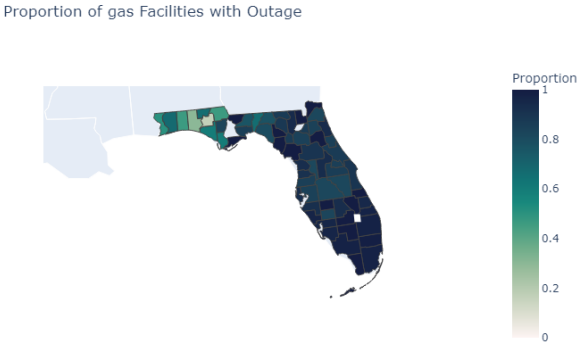
From the algorithm described in Chapter 2 I estimated functional closures for all Florida facilities data obtained from OpenStreetMaps as described in Chapter 4. Figure E.1 shows the total proportions of each facility type that experienced a detectable functional closure by county.



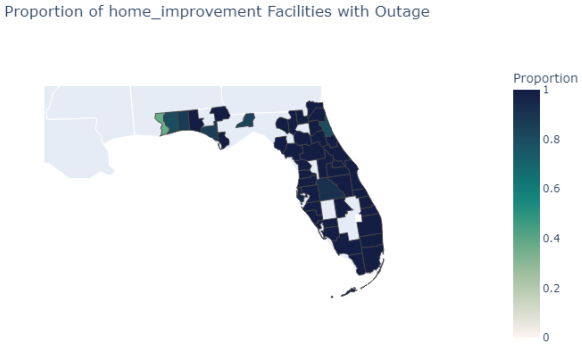
(a) Estimated proportion of clinics functional closures by county



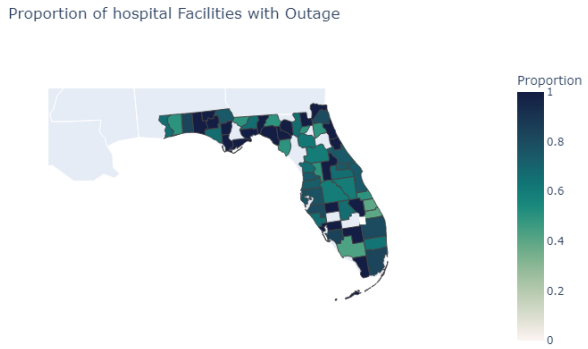
(b) Estimated proportion of convenience store functional closures by county



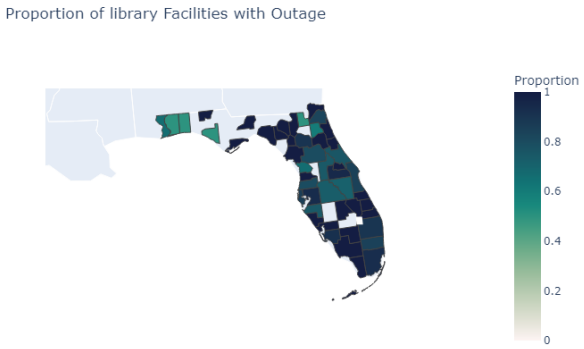
(c) Estimated proportion of gas station functional closures by county



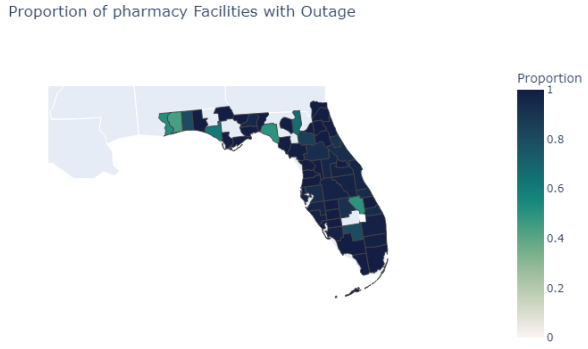
(d) Estimated proportion of home improvement functional closures by county



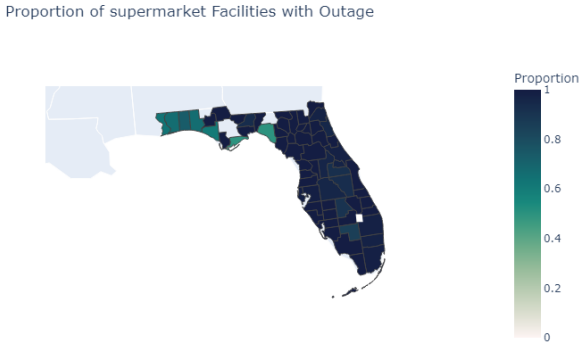
(a) Estimated proportion of hospital functional closures by county



(b) Estimated proportion of library functional closures by county



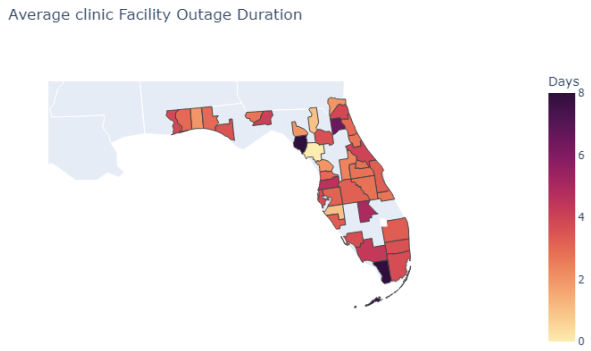
(c) Estimated proportion of pharmacy functional closures by county



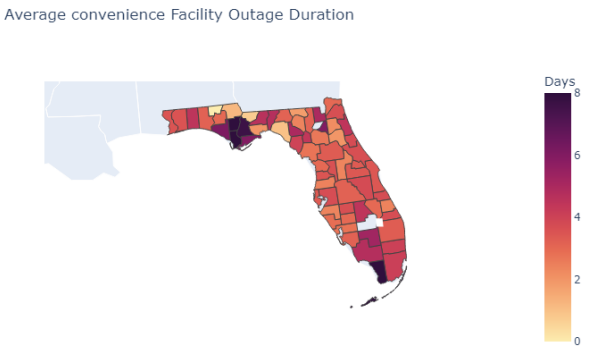
(d) Estimated proportion of supermarket functional closures by county

Figure E.1: County-level estimations for proportion of facilities facing some period of functional closure

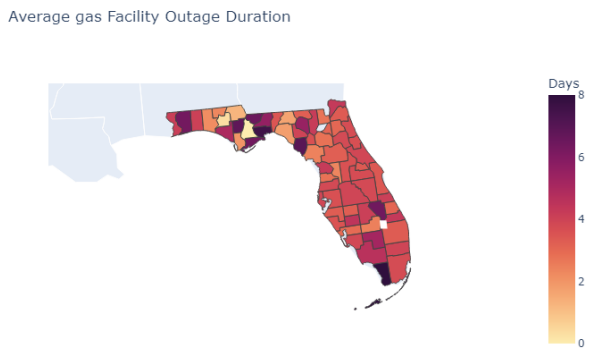
Figure E.2 shows the mean duration (in days) of the detected functional closure, including facilities where no functional closure was detected (i.e. duration equals 0).



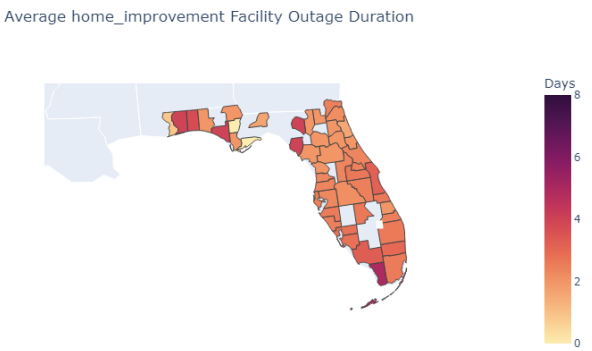
(a) Mean duration of clinics functional closures by county (days)



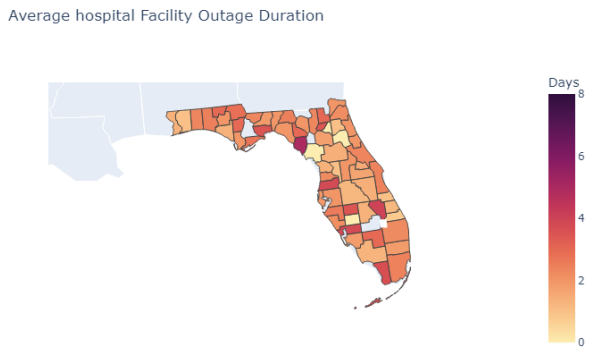
(b) Mean duration of convenience store functional closures by county (days)



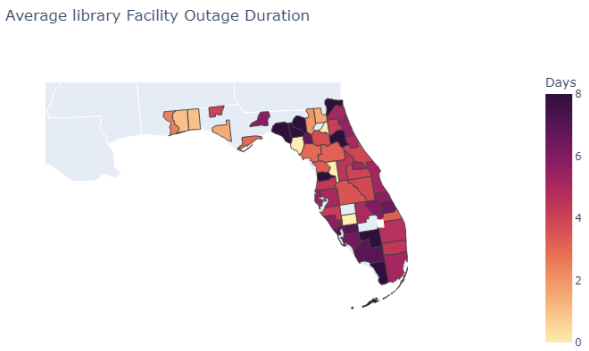
(c) Mean duration of gas station functional closures by county (days)



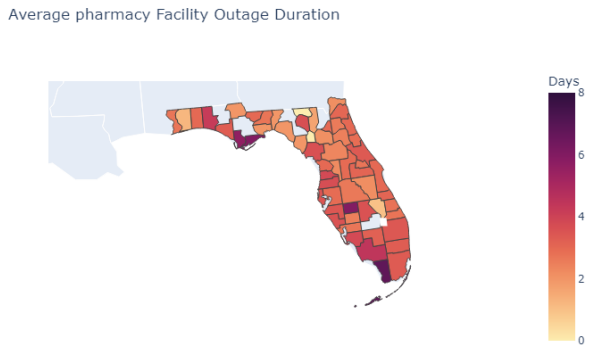
(d) Mean duration of home improvement functional closures by county (days)



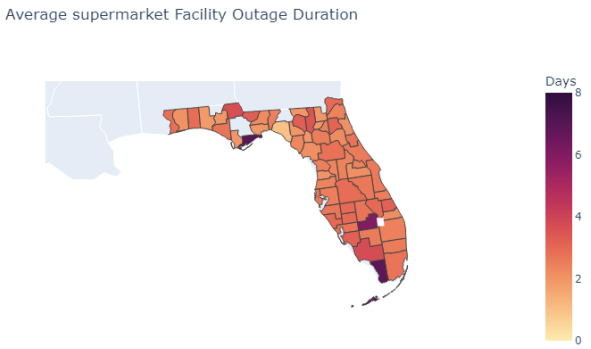
(a) Mean duration of hospital functional closures by county (days)



(b) Mean duration of library functional closures by county (days)



(c) Mean duration of pharmacy functional closures by county (days)



(d) Mean duration of supermarket functional closures by county (days)

Figure E.2: County-level estimations for mean duration of facility functional closures

APPENDIX F

Variables included in Statistical Modeling for Evaluating Access to Essential Services vs. Recovery

Table F.1 shows all variables used in estimating the recovery status of Floridians in the period surrounding Hurricane Irma. Data was obtained for 7,639 total users in Monroe, Collier, Lee and Hendry Counties for each day in September, 2017.

Category	Variable name	Variable	Units	Source
Recovery status	state	Recovery status	Binary	Recovery estimation algorithm
	prev_state	Users recovery status at same location on previous day	Binary	Calculated
	other_loc_state	Users recovery status at other location on same day	Binary	Calculated
	home	Location is home (versus work)	Binary	Home finding algorithm
Date	day	Days since landfall (9/10/2017)	Days	Date
	Friday	Day is a Friday	Binary	Date
	Monday	Day is a Monday	Binary	Date
	Saturday	Day is a Saturday	Binary	Date
	Sunday	Day is a Sunday	Binary	Date
	Thursday	Day is a Thursday	Binary	Date
	Tuesday	Day is a Tuesday	Binary	Date
	Wednesday	Day is a Wednesday	Binary	Date
School status	school_out	School closure status	Binary	County level school districts
Evacuation status	evac	Evacuation status by zone	Binary	Florida Division of Emergency Management public records request
Access	home_work_car_tt	Travel time between user's home and work by car	Minutes	OpenTripPlanner
	home_work_walk_tt	Travel time between user's home and work by walking	Minutes	OpenTripPlanner
	home_work_transit_tt	Travel time between user's home and work by transit	Minutes	OpenTripPlanner
	('time_(min)', 'clinic', 'CAR')	Minimum travel time to clinic by car	Minutes	OpenTripPlanner

('time_(min)', 'clinic', 'TRANSIT')	Minimum travel time to clinic by transit	Minutes	OpenTripPlanner
('time_(min)', 'clinic', 'WALK')	Minimum travel time to clinic by walking	Minutes	OpenTripPlanner
('time_(min)', 'convenience', 'CAR')	Minimum travel time to convenience store store by car	Minutes	OpenTripPlanner
('time_(min)', 'convenience', 'TRANSIT')	Minimum travel time to convenience store store by transit	Minutes	OpenTripPlanner
('time_(min)', 'convenience', 'WALK')	Minimum travel time to convenience store store by walking	Minutes	OpenTripPlanner
('time_(min)', 'gas', 'CAR')	Minimum travel time to gas station by car	Minutes	OpenTripPlanner
('time_(min)', 'gas', 'TRANSIT')	Minimum travel time to gas station by transit	Minutes	OpenTripPlanner
('time_(min)', 'gas', 'WALK')	Minimum travel time to gas station by walking	Minutes	OpenTripPlanner
('time_(min)', 'home', 'CAR')	Minimum travel time to home improvement improvement improvement by car	Minutes	OpenTripPlanner
('time_(min)', 'home', 'TRANSIT')	Minimum travel time to home improvement improvement by transit	Minutes	OpenTripPlanner
('time_(min)', 'home', 'WALK')	Minimum travel time to home improvement improvement by walking	Minutes	OpenTripPlanner
('time_(min)', 'hospital', 'CAR')	Minimum travel time to hospital by car	Minutes	OpenTripPlanner
('time_(min)', 'hospital', 'TRANSIT')	Minimum travel time to hospital by transit	Minutes	OpenTripPlanner
('time_(min)', 'hospital', 'WALK')	Minimum travel time to hospital by walking	Minutes	OpenTripPlanner
('time_(min)', 'library', 'CAR')	Minimum travel time to library by car	Minutes	OpenTripPlanner
('time_(min)', 'library', 'TRANSIT')	Minimum travel time to library by transit	Minutes	OpenTripPlanner
('time_(min)', 'library', 'WALK')	Minimum travel time to library by walking	Minutes	OpenTripPlanner
('time_(min)', 'pharmacy', 'CAR')	Minimum travel time to pharmacy by car	Minutes	OpenTripPlanner
('time_(min)', 'pharmacy', 'TRANSIT')	Minimum travel time to pharmacy by transit	Minutes	OpenTripPlanner
('time_(min)', 'pharmacy', 'WALK')	Minimum travel time to pharmacy by walking	Minutes	OpenTripPlanner
('time_(min)', 'school', 'CAR')	Minimum travel time to school by car	Minutes	OpenTripPlanner
('time_(min)', 'school', 'TRANSIT')	Minimum travel time to school by transit	Minutes	OpenTripPlanner
('time_(min)', 'school', 'WALK')	Minimum travel time to school by walking	Minutes	OpenTripPlanner
('time_(min)', 'supermarket', 'CAR')	Minimum travel time to supermarket by car	Minutes	OpenTripPlanner
('time_(min)', 'supermarket', 'TRANSIT')	Minimum travel time to supermarket by transit	Minutes	OpenTripPlanner
('time_(min)', 'supermarket', 'WALK')	Minimum travel time to supermarket by walking	Minutes	OpenTripPlanner
('time_(min)', 'clinic', 'CAR')_prelandfall	Minimum pre-landfall travel time to clinic by car	Minutes	OpenTripPlanner
('time_(min)', 'clinic', 'TRANSIT')_prelandfall	Minimum pre-landfall travel time to clinic by transit	Minutes	OpenTripPlanner

('time_(min)', 'clinic', 'WALK')_prelandfall	Minimum pre-landfall travel time to clinic by walking	Minutes	OpenTripPlanner
('time_(min)', 'convenience', 'CAR')_prelandfall	Minimum pre-landfall travel time to convenience store by car	Minutes	OpenTripPlanner
('time_(min)', 'convenience', 'TRANSIT')_prelandfall	Minimum pre-landfall travel time to convenience store by transit	Minutes	OpenTripPlanner
('time_(min)', 'convenience', 'WALK')_prelandfall	Minimum pre-landfall travel time to convenience store by walking	Minutes	OpenTripPlanner
('time_(min)', 'gas', 'CAR')_prelandfall	Minimum pre-landfall travel time to gas station by car	Minutes	OpenTripPlanner
('time_(min)', 'gas', 'TRANSIT')_prelandfall	Minimum pre-landfall travel time to gas station by transit	Minutes	OpenTripPlanner
('time_(min)', 'gas', 'WALK')_prelandfall	Minimum pre-landfall travel time to gas station by walking	Minutes	OpenTripPlanner
('time_(min)', 'home', 'CAR')_prelandfall	Minimum pre-landfall travel time to home improvement by car	Minutes	OpenTripPlanner
('time_(min)', 'home', 'TRANSIT')_prelandfall	Minimum pre-landfall travel time to home improvement by transit	Minutes	OpenTripPlanner
('time_(min)', 'home', 'WALK')_prelandfall	Minimum pre-landfall travel time to home improvement by walking	Minutes	OpenTripPlanner
('time_(min)', 'hospital', 'CAR')_prelandfall	Minimum pre-landfall travel time to hospital by car	Minutes	OpenTripPlanner
('time_(min)', 'hospital', 'TRANSIT')_prelandfall	Minimum pre-landfall travel time to hospital by transit	Minutes	OpenTripPlanner
('time_(min)', 'hospital', 'WALK')_prelandfall	Minimum pre-landfall travel time to hospital by walking	Minutes	OpenTripPlanner
('time_(min)', 'library', 'CAR')_prelandfall	Minimum pre-landfall travel time to library by car	Minutes	OpenTripPlanner
('time_(min)', 'library', 'TRANSIT')_prelandfall	Minimum pre-landfall travel time to library by transit	Minutes	OpenTripPlanner
('time_(min)', 'library', 'WALK')_prelandfall	Minimum pre-landfall travel time to library by walking	Minutes	OpenTripPlanner
('time_(min)', 'pharmacy', 'CAR')_prelandfall	Minimum pre-landfall travel time to pharmacy by car	Minutes	OpenTripPlanner
('time_(min)', 'pharmacy', 'TRANSIT')_prelandfall	Minimum pre-landfall travel time to pharmacy by transit	Minutes	OpenTripPlanner
('time_(min)', 'pharmacy', 'WALK')_prelandfall	Minimum pre-landfall travel time to pharmacy by walking	Minutes	OpenTripPlanner
('time_(min)', 'school', 'CAR')_prelandfall	Minimum pre-landfall travel time to school by car	Minutes	OpenTripPlanner
('time_(min)', 'school', 'TRANSIT')_prelandfall	Minimum pre-landfall travel time to school by transit	Minutes	OpenTripPlanner
('time_(min)', 'school', 'WALK')_prelandfall	Minimum pre-landfall travel time to school by walking	Minutes	OpenTripPlanner
('time_(min)', 'supermarket', 'CAR')_prelandfall	Minimum pre-landfall travel time to supermarket by car	Minutes	OpenTripPlanner

('time_(min)', 'supermarket', 'TRANSIT')_prelandfall	Minimum pre-landfall travel time to supermarket by transit	Minutes	OpenTripPlanner
('time_(min)', 'supermarket', 'WALK')_prelandfall	Minimum pre-landfall travel time to supermarket by walking	Minutes	OpenTripPlanner
('time_(min)', 'clinic', 'CAR')_diff	Difference from minimum pre-landfall travel time to clinic by car	Minutes	OpenTripPlanner
('time_(min)', 'clinic', 'TRANSIT')_diff	Difference from prelandfall travel time to clinic by transit	Minutes	OpenTripPlanner
('time_(min)', 'clinic', 'WALK')_diff	Difference from prelandfall travel time to clinic by walking	Minutes	OpenTripPlanner
('time_(min)', 'convenience', 'CAR')_diff	Difference from prelandfall travel time to convenience store by car	Minutes	OpenTripPlanner
('time_(min)', 'convenience', 'TRANSIT')_diff	Difference from prelandfall travel time to convenience store by transit	Minutes	OpenTripPlanner
('time_(min)', 'convenience', 'WALK')_diff	Difference from prelandfall travel time to convenience store by walking	Minutes	OpenTripPlanner
('time_(min)', 'gas', 'CAR')_diff	Difference from prelandfall travel time to gas station by car	Minutes	OpenTripPlanner
('time_(min)', 'gas', 'TRANSIT')_diff	Difference from prelandfall travel time to gas station by transit	Minutes	OpenTripPlanner
('time_(min)', 'gas', 'WALK')_diff	Difference from prelandfall travel time to gas station by walking	Minutes	OpenTripPlanner
('time_(min)', 'home', 'CAR')_diff	Difference from prelandfall travel time to home improvement by car	Minutes	OpenTripPlanner
('time_(min)', 'home', 'TRANSIT')_diff	Difference from prelandfall travel time to home improvement by transit	Minutes	OpenTripPlanner
('time_(min)', 'home', 'WALK')_diff	Difference from prelandfall travel time to home improvement by walking	Minutes	OpenTripPlanner
('time_(min)', 'hospital', 'CAR')_diff	Difference from prelandfall travel time to hospital by car	Minutes	OpenTripPlanner
('time_(min)', 'hospital', 'TRANSIT')_diff	Difference from prelandfall travel time to hospital by transit	Minutes	OpenTripPlanner
('time_(min)', 'hospital', 'WALK')_diff	Difference from prelandfall travel time to hospital by walking	Minutes	OpenTripPlanner
('time_(min)', 'library', 'CAR')_diff	Difference from prelandfall travel time to library by car	Minutes	OpenTripPlanner
('time_(min)', 'library', 'TRANSIT')_diff	Difference from prelandfall travel time to library by transit	Minutes	OpenTripPlanner
('time_(min)', 'library', 'WALK')_diff	Difference from prelandfall travel time to library by walking	Minutes	OpenTripPlanner
('time_(min)', 'pharmacy', 'CAR')_diff	Difference from prelandfall travel time to pharmacy by car	Minutes	OpenTripPlanner
('time_(min)', 'pharmacy', 'TRANSIT')_diff	Difference from prelandfall travel time to pharmacy by transit	Minutes	OpenTripPlanner

('time_(min)', 'pharmacy', 'WALK')_diff	Difference from pre-landfall travel time to pharmacy by walking	Minutes	OpenTripPlanner
('time_(min)', 'school', 'CAR')_diff	Difference from pre-landfall travel time to school by car	Minutes	OpenTripPlanner
('time_(min)', 'school', 'TRANSIT')_diff	Difference from pre-landfall travel time to school by transit	Minutes	OpenTripPlanner
('time_(min)', 'school', 'WALK')_diff	Difference from pre-landfall travel time to school by walking	Minutes	OpenTripPlanner
('time_(min)', 'supermarket', 'CAR')_diff	Difference from pre-landfall travel time to supermarket by car	Minutes	OpenTripPlanner
('time_(min)', 'supermarket', 'TRANSIT')_diff	Difference from pre-landfall travel time to supermarket by transit	Minutes	OpenTripPlanner
('time_(min)', 'supermarket', 'WALK')_diff	Difference from pre-landfall travel time to supermarket by walking	Minutes	OpenTripPlanner
('within_10_min', 'clinic', 'CAR')	Number of clinics within 10 minutes by car	Count	OpenTripPlanner
('within_10_min', 'clinic', 'TRANSIT')	Number of clinics within 10 minutes by transit	Count	OpenTripPlanner
('within_10_min', 'clinic', 'WALK')	Number of clinics within 10 minutes by walking	Count	OpenTripPlanner
('within_10_min', 'convenience', 'CAR')	Number of convenience stores within 10 minutes by car	Count	OpenTripPlanner
('within_10_min', 'convenience', 'TRANSIT')	Number of convenience stores within 10 minutes by transit	Count	OpenTripPlanner
('within_10_min', 'convenience', 'WALK')	Number of convenience stores within 10 minutes by walking	Count	OpenTripPlanner
('within_10_min', 'gas', 'CAR')	Number of gas stations within 10 minutes by car	Count	OpenTripPlanner
('within_10_min', 'gas', 'TRANSIT')	Number of gas stations within 10 minutes by transit	Count	OpenTripPlanner
('within_10_min', 'gas', 'WALK')	Number of gas stations within 10 minutes by walking	Count	OpenTripPlanner
('within_10_min', 'home', 'CAR')	Number of home improvements within 10 minutes by car	Count	OpenTripPlanner
('within_10_min', 'home', 'TRANSIT')	Number of home improvements within 10 minutes by transit	Count	OpenTripPlanner
('within_10_min', 'home', 'WALK')	Number of home improvements within 10 minutes by walking	Count	OpenTripPlanner
('within_10_min', 'hospital', 'CAR')	Number of hospitals within 10 minutes by car	Count	OpenTripPlanner
('within_10_min', 'hospital', 'TRANSIT')	Number of hospitals within 10 minutes by transit	Count	OpenTripPlanner
('within_10_min', 'hospital', 'WALK')	Number of hospitals within 10 minutes by walking	Count	OpenTripPlanner
('within_10_min', 'library', 'CAR')	Number of libraries within 10 minutes by car	Count	OpenTripPlanner
('within_10_min', 'library', 'TRANSIT')	Number of libraries within 10 minutes by transit	Count	OpenTripPlanner

('within_10_min', 'library', 'WALK')	Number of libraries within 10 minutes by walking	Count	OpenTripPlanner
('within_10_min', 'pharmacy', 'CAR')	Number of pharmacies within 10 minutes by car	Count	OpenTripPlanner
('within_10_min', 'pharmacy', 'TRANSIT')	Number of pharmacies within 10 minutes by transit	Count	OpenTripPlanner
('within_10_min', 'pharmacy', 'WALK')	Number of pharmacies within 10 minutes by walking	Count	OpenTripPlanner
('within_10_min', 'school', 'CAR')	Number of schools within 10 minutes by car	Count	OpenTripPlanner
('within_10_min', 'school', 'TRANSIT')	Number of schools within 10 minutes by transit	Count	OpenTripPlanner
('within_10_min', 'school', 'WALK')	Number of schools within 10 minutes by walking	Count	OpenTripPlanner
('within_10_min', 'supermarket', 'CAR')	Number of supermarkets within 10 minutes by car	Count	OpenTripPlanner
('within_10_min', 'supermarket', 'TRANSIT')	Number of supermarkets within 10 minutes by transit	Count	OpenTripPlanner
('within_10_min', 'supermarket', 'WALK')	Number of supermarkets within 10 minutes by walking	Count	OpenTripPlanner
('within_20_min', 'clinic', 'CAR')	Number of clinics within 20 minutes by car	Count	OpenTripPlanner
('within_20_min', 'clinic', 'TRANSIT')	Number of clinics within 20 minutes by transit	Count	OpenTripPlanner
('within_20_min', 'clinic', 'WALK')	Number of clinics within 20 minutes by walking	Count	OpenTripPlanner
('within_20_min', 'convenience', 'CAR')	Number of convenience stores within 20 minutes by car	Count	OpenTripPlanner
('within_20_min', 'convenience', 'TRANSIT')	Number of convenience stores within 20 minutes by transit	Count	OpenTripPlanner
('within_20_min', 'convenience', 'WALK')	Number of convenience stores within 20 minutes by walking	Count	OpenTripPlanner
('within_20_min', 'gas', 'CAR')	Number of gas stations within 20 minutes by car	Count	OpenTripPlanner
('within_20_min', 'gas', 'TRANSIT')	Number of gas stations within 20 minutes by transit	Count	OpenTripPlanner
('within_20_min', 'gas', 'WALK')	Number of gas stations within 20 minutes by walking	Count	OpenTripPlanner
('within_20_min', 'home', 'CAR')	Number of home improvements within 20 minutes by car	Count	OpenTripPlanner
('within_20_min', 'home', 'TRANSIT')	Number of home improvements within 20 minutes by transit	Count	OpenTripPlanner
('within_20_min', 'home', 'WALK')	Number of home improvements within 20 minutes by walking	Count	OpenTripPlanner
('within_20_min', 'hospital', 'CAR')	Number of hospitals within 20 minutes by car	Count	OpenTripPlanner
('within_20_min', 'hospital', 'TRANSIT')	Number of hospitals within 20 minutes by transit	Count	OpenTripPlanner

('within_20_min', 'hospital', 'WALK')	Number of hospitals within 20 minutes by walking	Count	OpenTripPlanner
('within_20_min', 'library', 'CAR')	Number of libraries within 20 minutes by car	Count	OpenTripPlanner
('within_20_min', 'library', 'TRANSIT')	Number of libraries within 20 minutes by transit	Count	OpenTripPlanner
('within_20_min', 'library', 'WALK')	Number of libraries within 20 minutes by walking	Count	OpenTripPlanner
('within_20_min', 'pharmacy', 'CAR')	Number of pharmacies within 20 minutes by car	Count	OpenTripPlanner
('within_20_min', 'pharmacy', 'TRANSIT')	Number of pharmacies within 20 minutes by transit	Count	OpenTripPlanner
('within_20_min', 'pharmacy', 'WALK')	Number of pharmacies within 20 minutes by walking	Count	OpenTripPlanner
('within_20_min', 'school', 'CAR')	Number of schools within 20 minutes by car	Count	OpenTripPlanner
('within_20_min', 'school', 'TRANSIT')	Number of schools within 20 minutes by transit	Count	OpenTripPlanner
('within_20_min', 'school', 'WALK')	Number of schools within 20 minutes by walking	Count	OpenTripPlanner
('within_20_min', 'supermarket', 'CAR')	Number of supermarkets within 20 minutes by car	Count	OpenTripPlanner
('within_20_min', 'supermarket', 'TRANSIT')	Number of supermarkets within 20 minutes by transit	Count	OpenTripPlanner
('within_20_min', 'supermarket', 'WALK')	Number of supermarkets within 20 minutes by walking	Count	OpenTripPlanner
('within_30_min', 'clinic', 'CAR')	Number of clinics within 30 minutes by car	Count	OpenTripPlanner
('within_30_min', 'clinic', 'TRANSIT')	Number of clinics within 30 minutes by transit	Count	OpenTripPlanner
('within_30_min', 'clinic', 'WALK')	Number of clinics within 30 minutes by walking	Count	OpenTripPlanner
('within_30_min', 'convenience', 'CAR')	Number of convenience stores within 30 minutes by car	Count	OpenTripPlanner
('within_30_min', 'convenience', 'TRANSIT')	Number of convenience stores within 30 minutes by transit	Count	OpenTripPlanner
('within_30_min', 'convenience', 'WALK')	Number of convenience stores within 30 minutes by walking	Count	OpenTripPlanner
('within_30_min', 'gas', 'CAR')	Number of gas stations within 30 minutes by car	Count	OpenTripPlanner
('within_30_min', 'gas', 'TRANSIT')	Number of gas stations within 30 minutes by transit	Count	OpenTripPlanner
('within_30_min', 'gas', 'WALK')	Number of gas stations within 30 minutes by walking	Count	OpenTripPlanner
('within_30_min', 'home', 'CAR')	Number of home improvements within 30 minutes by car	Count	OpenTripPlanner
('within_30_min', 'home', 'TRANSIT')	Number of home improvements within 30 minutes by transit	Count	OpenTripPlanner

('within_30_min', 'home', 'WALK')	Number of home improvements within 30 minutes by walking	Count	OpenTripPlanner
('within_30_min', 'hospital', 'CAR')	Number of hospitals within 30 minutes by car	Count	OpenTripPlanner
('within_30_min', 'hospital', 'TRANSIT')	Number of hospitals within 30 minutes by transit	Count	OpenTripPlanner
('within_30_min', 'hospital', 'WALK')	Number of hospitals within 30 minutes by walking	Count	OpenTripPlanner
('within_30_min', 'library', 'CAR')	Number of libraries within 30 minutes by car	Count	OpenTripPlanner
('within_30_min', 'library', 'TRANSIT')	Number of libraries within 30 minutes by transit	Count	OpenTripPlanner
('within_30_min', 'library', 'WALK')	Number of libraries within 30 minutes by walking	Count	OpenTripPlanner
('within_30_min', 'pharmacy', 'CAR')	Number of pharmacies within 30 minutes by car	Count	OpenTripPlanner
('within_30_min', 'pharmacy', 'TRANSIT')	Number of pharmacies within 30 minutes by transit	Count	OpenTripPlanner
('within_30_min', 'pharmacy', 'WALK')	Number of pharmacies within 30 minutes by walking	Count	OpenTripPlanner
('within_30_min', 'school', 'CAR')	Number of schools within 30 minutes by car	Count	OpenTripPlanner
('within_30_min', 'school', 'TRANSIT')	Number of schools within 30 minutes by transit	Count	OpenTripPlanner
('within_30_min', 'school', 'WALK')	Number of schools within 30 minutes by walking	Count	OpenTripPlanner
('within_30_min', 'supermarket', 'CAR')	Number of supermarkets within 30 minutes by car	Count	OpenTripPlanner
('within_30_min', 'supermarket', 'TRANSIT')	Number of supermarkets within 30 minutes by transit	Count	OpenTripPlanner
('within_30_min', 'supermarket', 'WALK')	Number of supermarkets within 30 minutes by walking	Count	OpenTripPlanner
('within_40_min', 'clinic', 'CAR')	Number of clinics within 40 minutes by car	Count	OpenTripPlanner
('within_40_min', 'clinic', 'TRANSIT')	Number of clinics within 40 minutes by transit	Count	OpenTripPlanner
('within_40_min', 'clinic', 'WALK')	Number of clinics within 40 minutes by walking	Count	OpenTripPlanner
('within_40_min', 'convenience', 'CAR')	Number of convenience stores within 40 minutes by car	Count	OpenTripPlanner
('within_40_min', 'convenience', 'TRANSIT')	Number of convenience stores within 40 minutes by transit	Count	OpenTripPlanner
('within_40_min', 'convenience', 'WALK')	Number of convenience stores within 40 minutes by walking	Count	OpenTripPlanner
('within_40_min', 'gas', 'CAR')	Number of gas stations within 40 minutes by car	Count	OpenTripPlanner
('within_40_min', 'gas', 'TRANSIT')	Number of gas stations within 40 minutes by transit	Count	OpenTripPlanner

('within_40_min', 'gas', 'WALK')	Number of gas stations within 40 minutes by walking	Count	OpenTripPlanner
('within_40_min', 'home', 'CAR')	Number of home improvements within 40 minutes by car	Count	OpenTripPlanner
('within_40_min', 'home', 'TRANSIT')	Number of home improvements within 40 minutes by transit	Count	OpenTripPlanner
('within_40_min', 'home', 'WALK')	Number of home improvements within 40 minutes by walking	Count	OpenTripPlanner
('within_40_min', 'hospital', 'CAR')	Number of hospitals within 40 minutes by car	Count	OpenTripPlanner
('within_40_min', 'hospital', 'TRANSIT')	Number of hospitals within 40 minutes by transit	Count	OpenTripPlanner
('within_40_min', 'hospital', 'WALK')	Number of hospitals within 40 minutes by walking	Count	OpenTripPlanner
('within_40_min', 'library', 'CAR')	Number of libraries within 40 minutes by car	Count	OpenTripPlanner
('within_40_min', 'library', 'TRANSIT')	Number of libraries within 40 minutes by transit	Count	OpenTripPlanner
('within_40_min', 'library', 'WALK')	Number of libraries within 40 minutes by walking	Count	OpenTripPlanner
('within_40_min', 'pharmacy', 'CAR')	Number of pharmacies within 40 minutes by car	Count	OpenTripPlanner
('within_40_min', 'pharmacy', 'TRANSIT')	Number of pharmacies within 40 minutes by transit	Count	OpenTripPlanner
('within_40_min', 'pharmacy', 'WALK')	Number of pharmacies within 40 minutes by walking	Count	OpenTripPlanner
('within_40_min', 'school', 'CAR')	Number of schools within 40 minutes by car	Count	OpenTripPlanner
('within_40_min', 'school', 'TRANSIT')	Number of schools within 40 minutes by transit	Count	OpenTripPlanner
('within_40_min', 'school', 'WALK')	Number of schools within 40 minutes by walking	Count	OpenTripPlanner
('within_40_min', 'supermarket', 'CAR')	Number of supermarkets within 40 minutes by car	Count	OpenTripPlanner
('within_40_min', 'supermarket', 'TRANSIT')	Number of supermarkets within 40 minutes by transit	Count	OpenTripPlanner
('within_40_min', 'supermarket', 'WALK')	Number of supermarkets within 40 minutes by walking	Count	OpenTripPlanner
('within_50_min', 'clinic', 'CAR')	Number of clinics within 50 minutes by car	Count	OpenTripPlanner
('within_50_min', 'clinic', 'TRANSIT')	Number of clinics within 50 minutes by transit	Count	OpenTripPlanner
('within_50_min', 'clinic', 'WALK')	Number of clinics within 50 minutes by walking	Count	OpenTripPlanner
('within_50_min', 'convenience', 'CAR')	Number of convenience stores within 50 minutes by car	Count	OpenTripPlanner
('within_50_min', 'convenience', 'TRANSIT')	Number of convenience stores within 50 minutes by transit	Count	OpenTripPlanner

('within_50_min', 'convenience', 'WALK')	Number of convenience stores within 50 minutes by walking	Count	OpenTripPlanner
('within_50_min', 'gas', 'CAR')	Number of gas stations within 50 minutes by car	Count	OpenTripPlanner
('within_50_min', 'gas', 'TRANSIT')	Number of gas stations within 50 minutes by transit	Count	OpenTripPlanner
('within_50_min', 'gas', 'WALK')	Number of gas stations within 50 minutes by walking	Count	OpenTripPlanner
('within_50_min', 'home', 'CAR')	Number of home improvements within 50 minutes by car	Count	OpenTripPlanner
('within_50_min', 'home', 'TRANSIT')	Number of home improvements within 50 minutes by transit	Count	OpenTripPlanner
('within_50_min', 'home', 'WALK')	Number of home improvements within 50 minutes by walking	Count	OpenTripPlanner
('within_50_min', 'hospital', 'CAR')	Number of hospitals within 50 minutes by car	Count	OpenTripPlanner
('within_50_min', 'hospital', 'TRANSIT')	Number of hospitals within 50 minutes by transit	Count	OpenTripPlanner
('within_50_min', 'hospital', 'WALK')	Number of hospitals within 50 minutes by walking	Count	OpenTripPlanner
('within_50_min', 'library', 'CAR')	Number of libraries within 50 minutes by car	Count	OpenTripPlanner
('within_50_min', 'library', 'TRANSIT')	Number of libraries within 50 minutes by transit	Count	OpenTripPlanner
('within_50_min', 'library', 'WALK')	Number of libraries within 50 minutes by walking	Count	OpenTripPlanner
('within_50_min', 'pharmacy', 'CAR')	Number of pharmacies within 50 minutes by car	Count	OpenTripPlanner
('within_50_min', 'pharmacy', 'TRANSIT')	Number of pharmacies within 50 minutes by transit	Count	OpenTripPlanner
('within_50_min', 'pharmacy', 'WALK')	Number of pharmacies within 50 minutes by walking	Count	OpenTripPlanner
('within_50_min', 'school', 'CAR')	Number of schools within 50 minutes by car	Count	OpenTripPlanner
('within_50_min', 'school', 'TRANSIT')	Number of schools within 50 minutes by transit	Count	OpenTripPlanner
('within_50_min', 'school', 'WALK')	Number of schools within 50 minutes by walking	Count	OpenTripPlanner
('within_50_min', 'supermarket', 'CAR')	Number of supermarkets within 50 minutes by car	Count	OpenTripPlanner
('within_50_min', 'supermarket', 'TRANSIT')	Number of supermarkets within 50 minutes by transit	Count	OpenTripPlanner
('within_50_min', 'supermarket', 'WALK')	Number of supermarkets within 50 minutes by walking	Count	OpenTripPlanner
('within_60_min', 'clinic', 'CAR')	Number of clinics within 60 minutes by car	Count	OpenTripPlanner
('within_60_min', 'clinic', 'TRANSIT')	Number of clinics within 60 minutes by transit	Count	OpenTripPlanner

('within_60_min', 'clinic', 'WALK')	Number of clinics within 60 minutes by walking	Count	OpenTripPlanner
('within_60_min', 'convenience', 'CAR')	Number of convenience stores within 60 minutes by car	Count	OpenTripPlanner
('within_60_min', 'convenience', 'TRANSIT')	Number of convenience stores within 60 minutes by transit	Count	OpenTripPlanner
('within_60_min', 'convenience', 'WALK')	Number of convenience stores within 60 minutes by walking	Count	OpenTripPlanner
('within_60_min', 'gas', 'CAR')	Number of gas stations within 60 minutes by car	Count	OpenTripPlanner
('within_60_min', 'gas', 'TRANSIT')	Number of gas stations within 60 minutes by transit	Count	OpenTripPlanner
('within_60_min', 'gas', 'WALK')	Number of gas stations within 60 minutes by walking	Count	OpenTripPlanner
('within_60_min', 'home', 'CAR')	Number of home improvements within 60 minutes by car	Count	OpenTripPlanner
('within_60_min', 'home', 'TRANSIT')	Number of home improvements within 60 minutes by transit	Count	OpenTripPlanner
('within_60_min', 'home', 'WALK')	Number of home improvements within 60 minutes by walking	Count	OpenTripPlanner
('within_60_min', 'hospital', 'CAR')	Number of hospitals within 60 minutes by car	Count	OpenTripPlanner
('within_60_min', 'hospital', 'TRANSIT')	Number of hospitals within 60 minutes by transit	Count	OpenTripPlanner
('within_60_min', 'hospital', 'WALK')	Number of hospitals within 60 minutes by walking	Count	OpenTripPlanner
('within_60_min', 'library', 'CAR')	Number of libraries within 60 minutes by car	Count	OpenTripPlanner
('within_60_min', 'library', 'TRANSIT')	Number of libraries within 60 minutes by transit	Count	OpenTripPlanner
('within_60_min', 'library', 'WALK')	Number of libraries within 60 minutes by walking	Count	OpenTripPlanner
('within_60_min', 'pharmacy', 'CAR')	Number of pharmacies within 60 minutes by car	Count	OpenTripPlanner
('within_60_min', 'pharmacy', 'TRANSIT')	Number of pharmacies within 60 minutes by transit	Count	OpenTripPlanner
('within_60_min', 'pharmacy', 'WALK')	Number of pharmacies within 60 minutes by walking	Count	OpenTripPlanner
('within_60_min', 'school', 'CAR')	Number of schools within 60 minutes by car	Count	OpenTripPlanner
('within_60_min', 'school', 'TRANSIT')	Number of schools within 60 minutes by transit	Count	OpenTripPlanner
('within_60_min', 'school', 'WALK')	Number of schools within 60 minutes by walking	Count	OpenTripPlanner
('within_60_min', 'supermarket', 'CAR')	Number of supermarkets within 60 minutes by car	Count	OpenTripPlanner
('within_60_min', 'supermarket', 'TRANSIT')	Number of supermarkets within 60 minutes by transit	Count	OpenTripPlanner

	('within_60_min', 'supermarket', 'WALK')	Number of supermarkets within 60 minutes by walking	Count	OpenTripPlanner
Geography	AREA_SQMI	Area	Square miles (census tract)	CDC 2018 SoVI
	urban	Location is in urban boundary	Binary	TDA Functional Classification and Urban Boundary
	monroe	Location is in Monroe County	Binary	TIGER/Line Florida 2017
	collier	Location is in Collier County	Binary	TIGER/Line Florida 2018
	lee	Location is in Lee County	Binary	TIGER/Line Florida 2019
	hendry	Location is in Hendry County	Binary	TIGER/Line Florida 2020
Social Vulnerability	E_TOTPOP	Population estimate	Count (census tract)	CDC 2018 SoVI
	E_HU	Housing units	Count (census tract)	CDC 2018 SoVI
	E_HH	Households	Count (census tract)	CDC 2018 SoVI
	E_POV	Persons below poverty	Count (census tract)	CDC 2018 SoVI
	E_UNEMP	Civilians(age 16+) unemployed	Count (census tract)	CDC 2018 SoVI
	E_PCI	Per capita income	$ (census tract)	CDC 2018 SoVI
	E_NOHSDP	Persons (age 25+) with no high school diploma	Count (census tract)	CDC 2018 SoVI
	E_AGE65	Persons aged 65 and older	Count (census tract)	CDC 2018 SoVI
	E_AGE17	Persons aged 17 and younger	Count (census tract)	CDC 2018 SoVI
	E_DISABL	Civilian non-institutionalized population with a disability	Count (census tract)	CDC 2018 SoVI
	E_SNGPNT	Single parent household with children under 18	Count (census tract)	CDC 2018 SoVI
	E_MINRTY	Minority (all persons except white, non-Hispanic)	Count (census tract)	CDC 2018 SoVI
	E_LIMENG	Persons (age 5+) who speak English "less than well"	Count (census tract)	CDC 2018 SoVI
	E_MUNIT	Housing in structures with 10 or more units	Count (census tract)	CDC 2018 SoVI
	E_MOBILE	Mobile homes	Count (census tract)	CDC 2018 SoVI
	E_CROWD	At household level (occupied housing units), more people than rooms	Count (census tract)	CDC 2018 SoVI
	E_NOVEH	Households with no vehicle available	Count (census tract)	CDC 2018 SoVI
	E_GROUPQ	Persons in group quarters	Count (census tract)	CDC 2018 SoVI
	EP_POV	Percentage of persons below poverty	% (of census tract)	CDC 2018 SoVI
	EP_UNEMP	Unemployment Rate	% (of census tract)	CDC 2018 SoVI
	EP_NOHSDP	Percentage of persons with no high school diploma (age 25+)	% (of census tract)	CDC 2018 SoVI
	EP_AGE65	Percentage of persons aged 65 and older	% (of census tract)	CDC 2018 SoVI
	EP_AGE17	Percentage of persons aged 17 and younger	% (of census tract)	CDC 2018 SoVI
	EP_DISABL	Percentage of civilian non-institutionalized population with a disability	% (of census tract)	CDC 2018 SoVI
	EP_SNGPNT	Percentage of single parent households with children under 18	% (of census tract)	CDC 2018 SoVI

EP_MINRTY	Percentage minority (all persons except white, non-Hispanic)	% (of census tract)	CDC 2018 SoVI
EP_LIMENG	Percentage of persons (age 5+) who speak English "less than well"	% (of census tract)	CDC 2018 SoVI
EP_MUNIT	Percentage of housing in structures with 10 or more units	% (of census tract)	CDC 2018 SoVI
EP_MOBILE	Percentage of mobile homes	% (of census tract)	CDC 2018 SoVI
EP_CROWD	Percentage of occupied housing units with more people than rooms	% (of census tract)	CDC 2018 SoVI
EP_NOVEH	Percentage of households with no vehicle available	% (of census tract)	CDC 2018 SoVI
EP_GROUPQ	Percentage of persons in group quarters	% (of census tract)	CDC 2018 SoVI
EPL_POV	Percentile percentage of persons below poverty	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_UNEMP	Percentile percentage of civilian (age 16+) unemployed	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_PCI	Percentile per capita income	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_NOHSDP	Percentile percentage of persons with no high school diploma (age 25+)	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
SPL_THEME1	Socioeconomic theme: EPL_POV + EPL_UNEMP + EPL_PCI + EPL_NOHSDP	Sum of % (census tract)	CDC 2018 SoVI
RPL_THEME1	Percentile ranking for socioeconomic theme	Percent rank (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_AGE65	Percentile percentage of persons aged 65 and older	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_AGE17	Percentile percentage of persons aged 17 and younger	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_DISABL	Percentile percentage of civilian non-institutionalized population with a disability	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_SNGPNT	Percentile percentage of single parent households with children under 18	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
SPL_THEME2	Household composition theme: EPL_AGE65 + EPL_AGE17 + EPL_DISABL + EPL_SNGPNT	Sum of % (census tract)	CDC 2018 SoVI
RPL_THEME2	Percentile ranking for household composition theme	Percent rank (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_MINRTY	Percentile percentage minority (all persons except white, non-Hispanic)	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_LIMENG	Percentile percentage of persons (age 5+) who speak English "less than well"	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI

SPL_THEME3	Minority status/language theme: EPL_MINRTY + EPL_LIMENG	Sum of % (census tract)	CDC 2018 SoVI
RPL_THEME3	Percentile ranking for minority status/language theme	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_MUNIT	Percentile percentage housing in structures with 10 or more units	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_MOBILE	Percentile percentage mobile homes	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_CROWD	Percentile percentage households with more people than rooms	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_NOVEH	Percentile percentage households with no vehicle available	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
EPL_GROUPQ	Percentile percentage of persons in group quarters estimate	Percentile (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
SPL_THEME4	Housing type/transportation theme: EPL_MUNIT + EPL_MOBIL + EPL_CROWD + EPL_NOVEH + EPL_GROUPQ	Sum of % (census tract)	CDC 2018 SoVI
RPL_THEME4	Percentile ranking for housing type / transportation theme	Percent rank (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
SPL_THEMES	Sum of series themes: SPL_THEME1 + SPL_THEME2 + SPL_THEME3 + SPL_THEME4	Sum of % (census tract)	CDC 2018 SoVI
RPL_THEMES	Percentile ranking for sum of series themes	Percent rank (census tract relative to other U.S. census tracts)	CDC 2018 SoVI
F_POV	Flag- percentage of persons in poverty is in the 90th percentile	Binary	CDC 2018 SoVI
F_UNEMP	Flag - the percentage of civilian unemployed is in the 90th percentile	Binary	CDC 2018 SoVI
F_PCI	Flag - per capita income is in the 90th percentile	Binary	CDC 2018 SoVI
F_NOHSDP	Flag - the percentage of persons with no high school diploma is in the 90th percentile	Binary	CDC 2018 SoVI
F_THEME1	Sum of flags for socioeconomic status theme: F_POV + F_UNEMP + F_PCI + F_NOHSDP	Count	CDC 2018 SoVI
F_AGE65	Flag - the percentage of persons aged 65 and older is in the 90th percentile	Binary	CDC 2018 SoVI
F_AGE17	Flag - the percentage of persons aged 17 and younger is in the 90th percentile	Binary	CDC 2018 SoVI
F_DISABL	Flag - the percentage of persons with a disability is in the 90th percentile	Binary	CDC 2018 SoVI

	F_SNGPNT	Flag - the percentage of single parent households is in the 90th percentile	Binary	CDC 2018 SoVI
	F_THEME2	Sum of flags for household composition theme: F_AGE65 + F_AGE17 + F_DISABL + F_SNGPNT	Count	CDC 2018 SoVI
	F_MINRTY	Flag - the percentage of minority is in the 90th percentile	Binary	CDC 2018 SoVI
	F_LIMENG	Flag - the percentage those with limited English is in the 90th percentile	Binary	CDC 2018 SoVI
	F_THEME3	Sum of flags for Minority Status/Language theme	Count	CDC 2018 SoVI
	F_MUNIT	Flag - the percentage of households in multi-unit housing is in the 90th percentile	Binary	CDC 2018 SoVI
	F_MOBILE	Flag - the percentage of mobile homes is in the 90th percentile	Binary	CDC 2018 SoVI
	F_CROWD	Flag - the percentage of crowded households is in the 90th percentile	Binary	CDC 2018 SoVI
	F_NOVEH	Flag - the percentage of households with no vehicles is in the 90th percentile	Binary	CDC 2018 SoVI
	F_GROUPQ	Flag - the percentage of persons in institutionalized group quarters is in the 90th percentile	Binary	CDC 2018 SoVI
	F_THEME4	Sum of flags for housing type / transportation theme	Count	CDC 2018 SoVI
	F_TOTAL	Sum of flags for the four themes	Count	CDC 2018 SoVI
	E_UNINSUR	Uninsured in the total civilian non-institutionalized population	Count (census tract)	CDC 2018 SoVI
	EP_UNINSUR	Percentage uninsured in the total civilian non-institutionalized population	% (of census tract)	CDC 2018 SoVI
	E_DAYPOP	Estimated daytime population	Count (census tract)	CDC 2018 SoVI
	density	Density	Population/square mile (census tract)	Calculated from SoVI variables
Power	noon_percent_outage	% of county without power at noon	% (of county)	Florida Division of Emergency Management as reported through Florida Today
	power_data_days_to_25	Days until power restored for 25% of county	Days	Florida Division of Emergency Management as reported through Florida Today
	power_data_days_to_50	Days until power restored for 50% of county	Days	Florida Division of Emergency Management as reported through Florida Today
	power_data_days_to_75	Days until power restored for 75% of county	Days	Florida Division of Emergency Management as reported through Florida Today

	power_data_days_to_90	Days until power restored for 90% of county	Days	Florida Division of Emergency Management as reported through Florida Today
	power_data_days_to_95	Days until power restored for 95% of county	Days	Florida Division of Emergency Management as reported through Florida Today
	power_data_days_to_99	Days until power restored for 99% of county	Days	Florida Division of Emergency Management as reported through Florida Today
	power_data_hours_to_25	Hours until power restored for 25% of county	Hours	Florida Division of Emergency Management as reported through Florida Today
	power_data_hours_to_50	Hours until power restored for 50% of county	Hours	Florida Division of Emergency Management as reported through Florida Today
	power_data_hours_to_75	Hours until power restored for 75% of county	Hours	Florida Division of Emergency Management as reported through Florida Today
	power_data_hours_to_90	Hours until power restored for 90% of county	Hours	Florida Division of Emergency Management as reported through Florida Today
	power_data_hours_to_95	Hours until power restored for 95% of county	Hours	Florida Division of Emergency Management as reported through Florida Today
	power_data_hours_to_99	Hours until power restored for 99% of county	Hours	Florida Division of Emergency Management as reported through Florida Today
Cell Network	cell_service_outage	% of cell sites out	% (of county)	FCC Hurricane Irma Communications Status Reports
Wind	vmax_gust	Maximum 10-m 1-minute gust wind experienced at the grid point during the storm	Meters/second	[167]
	vmax_sust	Maximum sustained wind speed	Meters/second	[167]
	gust_dur	Duration gust wind was at or above a specified speed (default is 20 m/s), in minutes	Minutes	[167]
	sust_dur	Duration sustained wind was at or above a specified speed (default is 20 m/s)	Minutes	[167]
Storm Surge	maxDepth	Maximum surge depth	Meters	[168]
	maxFlowSpeed	Maximum surge flow speed	Meters/second	[168]
	maxUnitDischarge	Maximum unit discharge	Square-meters/second	[168]
	maxForcePerLength	Maximum force per length	Kilograms/square-second	[168]
	maxSignificantWaveHeight	Maximum significant wave height	Meters	[168]

Table F.1: Table of variables used in random forest model estimating recovery state

Figure F.1 below shows the correlation between the 374 included features. For readability,

labels are included to describe groups of features.

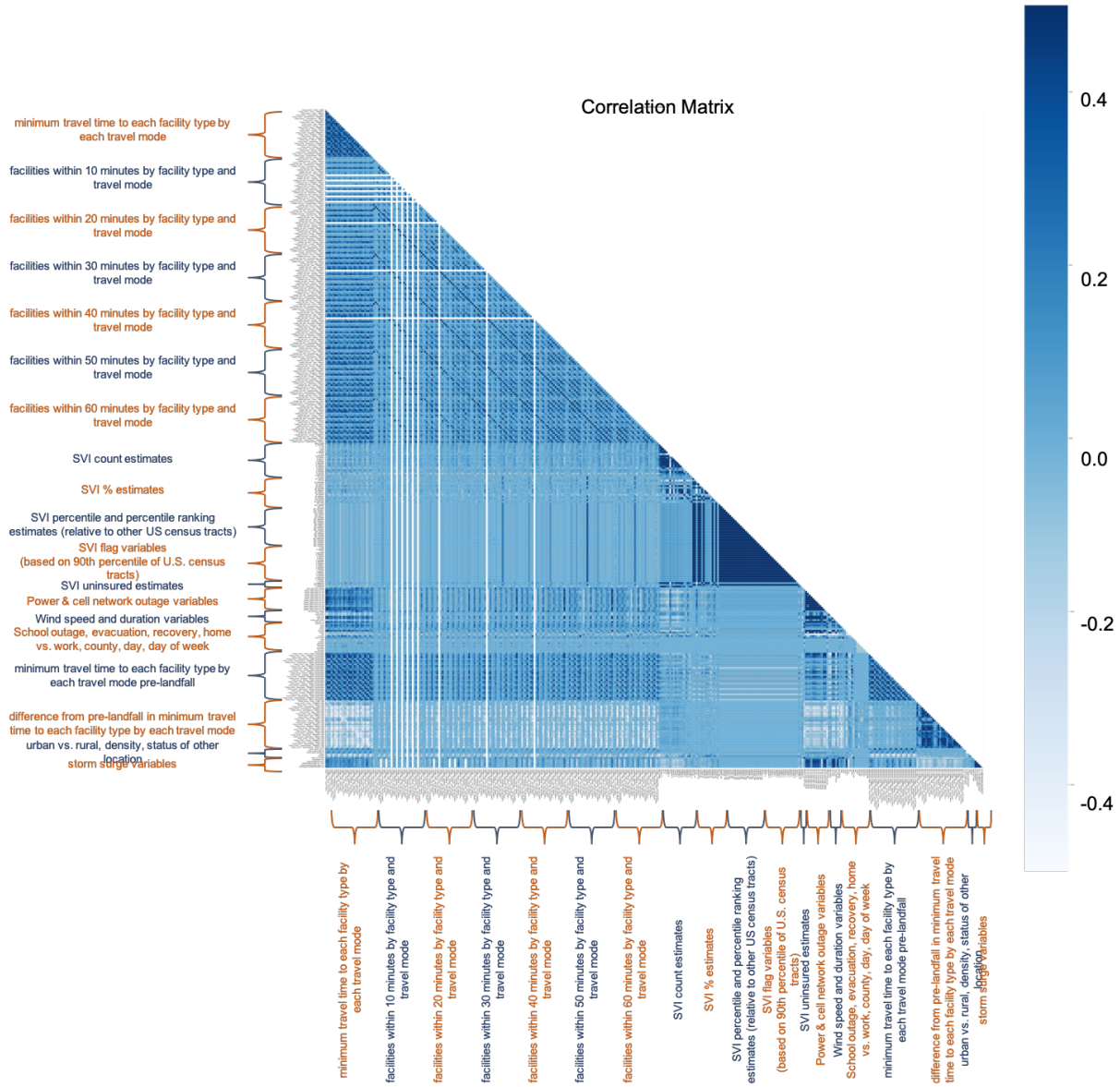


Figure F.1: Correlation between model features

In general, correlation between feature pairs is low except for between variables within the same group. The pre-landfall access values are deduced from the minimum travel time access values, and so are highly correlated. The social vulnerability index percentile and flag features are highly correlated, in part because they do not contain much variability as they are determined relative to other census tract values across the United States. The storm surge variables appear correlated with the wind and power outage variables.

APPENDIX G

Random Forest Classifier Cross Validation Scenarios

The table below in Figure G.1 shows each model scenario that was evaluated with random sample cross validation. Cross validation was iterated 10 times with a test set size of 10%.

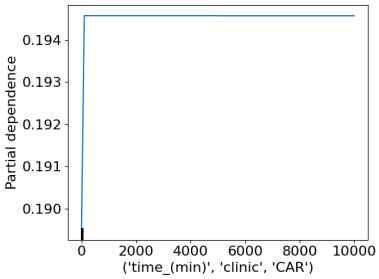
Counties	Class Weighting	Dropped Features	Test accuracy	Precision	Recall	ROC AUC	F1	Brier Score
Monroe and Collier	None	None	0.942	0.878	0.888	0.982	0.883	0.044
Monroe and Collier	None	Group A: access features containing "within"	0.943	0.882	0.890	0.983	0.886	0.043
Monroe and Collier	None	Group B: flag and percentile SVI features	0.941	0.875	0.886	0.982	0.880	0.045
Monroe and Collier	None	Group C: transit features	0.942	0.878	0.889	0.982	0.883	0.044
Monroe and Collier	None	Group D: power features exept noon % outage	0.942	0.877	0.888	0.982	0.883	0.045
Monroe and Collier	balanced subsample	None	0.942	0.886	0.879	0.983	0.882	0.044
Monroe and Collier	balanced subsample	Group A: access features containing "within"	0.944	0.889	0.886	0.984	0.888	0.042
Monroe and Collier	balanced subsample	Group B: flag and percentile SVI features	0.943	0.888	0.881	0.983	0.885	0.043
Monroe and Collier	balanced subsample	Group C: transit features	0.943	0.889	0.883	0.983	0.886	0.043
Monroe and Collier	balanced subsample	Group D: power features exept noon % outage	0.942	0.886	0.878	0.982	0.882	0.044
Monroe and Collier	Sample weighting by 25	None	0.943	0.889	0.879	0.983	0.884	0.043
Monroe and Collier	Sample weighting by 25	Group A: access features containing "within"	0.945	0.892	0.883	0.984	0.888	0.041
Monroe and Collier	Sample weighting by 25	Group B: flag and percentile SVI features	0.943	0.891	0.878	0.983	0.884	0.043
Monroe and Collier	Sample weighting by 25	Group C: transit features	0.944	0.890	0.880	0.983	0.885	0.043
Monroe and Collier	Sample weighting by 25	Group D: power features exept noon % outage	0.942	0.889	0.875	0.983	0.882	0.044
Monroe and Collier	Sample weighting by 25	None	0.944	0.890	0.881	0.983	0.885	0.042
Monroe and Collier	Sample weighting by 25	Group A: access features containing "within"	0.944	0.889	0.885	0.983	0.887	0.042
Monroe and Collier	Sample weighting by 25	Group B: flag and percentile SVI features	0.943	0.890	0.881	0.983	0.885	0.043
Monroe and Collier	Sample weighting by 25	Group C: transit features	0.944	0.890	0.883	0.983	0.886	0.042
Monroe and Collier	Sample weighting by 25	Group D: power features exept noon % outage	0.944	0.890	0.881	0.983	0.885	0.043
Monroe, Collier, Lee, and Hendry	None	None	0.943	0.857	0.859	0.980	0.858	0.044
Monroe, Collier, Lee, and Hendry	None	Group A: access features containing "within"	0.946	0.861	0.867	0.981	0.864	0.041
Monroe, Collier, Lee, and Hendry	None	Group B: flag and percentile SVI features	0.944	0.858	0.859	0.980	0.859	0.043
Monroe, Collier, Lee, and Hendry	None	Group C: transit features	0.943	0.855	0.858	0.979	0.857	0.043
Monroe, Collier, Lee, and Hendry	None	Group D: power features exept noon % outage	0.943	0.855	0.858	0.980	0.856	0.044
Monroe, Collier, Lee, and Hendry	balanced subsample	None	0.944	0.868	0.848	0.980	0.858	0.043
Monroe, Collier, Lee, and Hendry	balanced subsample	Group A: access features containing "within"	0.945	0.866	0.859	0.981	0.863	0.041
Monroe, Collier, Lee, and Hendry	balanced subsample	Group B: flag and percentile SVI features	0.945	0.869	0.850	0.980	0.859	0.042
Monroe, Collier, Lee, and Hendry	balanced subsample	Group C: transit features	0.944	0.866	0.851	0.980	0.858	0.043
Monroe, Collier, Lee, and Hendry	balanced subsample	Group D: power features exept noon % outage	0.943	0.866	0.845	0.980	0.855	0.043

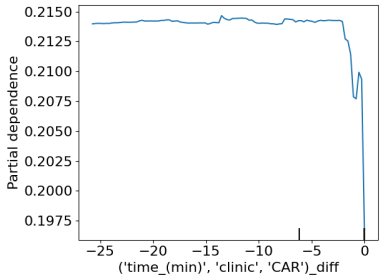
Monroe, Collier, Lee, and Hendry	Sample weighting by 25	None	0.944	0.867	0.846	0.980	0.857	0.043
Monroe, Collier, Lee, and Hendry	Sample weighting by 25	Group A: access features containing "within"	0.946	0.868	0.858	0.981	0.863	0.041
Monroe, Collier, Lee, and Hendry	Sample weighting by 25	Group B: flag and percentile SVI features	0.945	0.869	0.851	0.980	0.860	0.042
Monroe, Collier, Lee, and Hendry	Sample weighting by 25	Group C: transit features	0.944	0.868	0.849	0.980	0.858	0.043
Monroe, Collier, Lee, and Hendry	Sample weighting by 25	Group D: power features exept noon % outage	0.944	0.869	0.847	0.980	0.858	0.042
Monroe, Collier, Lee, and Hendry	Sample weighting by 25	None	0.945	0.869	0.852	0.980	0.860	0.042
Monroe, Collier, Lee, and Hendry	Sample weighting by 25	Group A: access features containing "within"	0.946	0.869	0.860	0.981	0.864	0.041
Monroe, Collier, Lee, and Hendry	Sample weighting by 25	Group B: flag and percentile SVI features	0.946	0.872	0.854	0.980	0.863	0.041
Monroe, Collier, Lee, and Hendry	Sample weighting by 25	Group C: transit features	0.945	0.868	0.854	0.980	0.861	0.042
Monroe, Collier, Lee, and Hendry	Sample weighting by 25	Group D: power features exept noon % outage	0.946	0.871	0.854	0.981	0.862	0.041
Monroe and Collier	balanced subsample	Groups A, B, C,and D	0.945	0.886	0.891	0.983	0.889	0.042
Monroe and Collier	balanced subsample	Groups A, B, and C	0.944	0.887	0.886	0.983	0.886	0.042
Monroe and Collier	balanced subsample	Groups A, B, and D	0.944	0.888	0.882	0.983	0.885	0.043
Monroe and Collier	balanced subsample	Groups A, C, and D	0.943	0.883	0.884	0.983	0.884	0.043
Monroe and Collier	balanced subsample	Groups B, C, and D	0.944	0.886	0.886	0.983	0.886	0.042
Monroe, Collier, Lee, and Hendry	balanced subsample	Groups A, B, C,and D	0.947	0.867	0.865	0.981	0.866	0.041
Monroe, Collier, Lee, and Hendry	balanced subsample	Groups A, B, and C	0.946	0.865	0.863	0.981	0.864	0.041
Monroe, Collier, Lee, and Hendry	balanced subsample	Groups A, B, and D	0.945	0.867	0.853	0.980	0.860	0.042
Monroe, Collier, Lee, and Hendry	balanced subsample	Groups A, C, and D	0.945	0.866	0.859	0.981	0.863	0.041
Monroe, Collier, Lee, and Hendry	balanced subsample	Groups B, C, and D	0.946	0.868	0.861	0.981	0.864	0.041
Monroe and Collier	balanced subsample	Sensitivity test removing previous state	0.926	0.869	0.824	0.973	0.846	0.057
Monroe, Collier, Lee, and Hendry	balanced subsample	Sensitivity test removing previous state	0.903	0.826	0.769	0.96	0.796	0.071
Monroe and Collier	balanced subsample	Sensitivity test removing previous state and state of user's other location	0.915	0.825	0.723	0.962	0.771	0.062
Monroe, Collier, Lee, and Hendry	balanced subsample	Sensitivity test removing previous state and state of user's other location	0.889	0.768	0.633	0.944	0.694	0.076
Monroe and Collier	balanced subsample	variables, excluding SVI count and percentage features	0.945	0.886	0.89	0.983	0.888	0.042
Monroe, Collier, Lee, and Hendry	balanced subsample	variables, excluding SVI count and percentage features	0.946	0.867	0.863	0.981	0.865	0.041

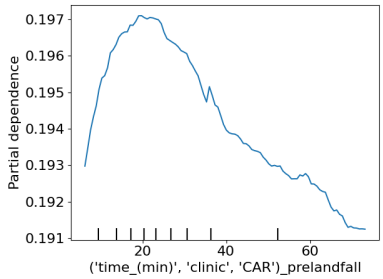
Figure G.1: Random forest scenarios with mean of cross validation evaluation metrics

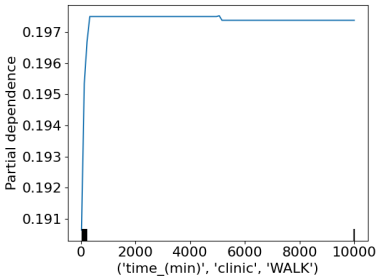
APPENDIX H

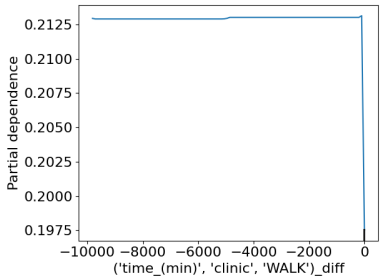
Partial Dependence Plots of Random Forest Model Variables

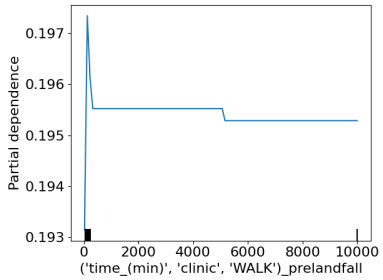


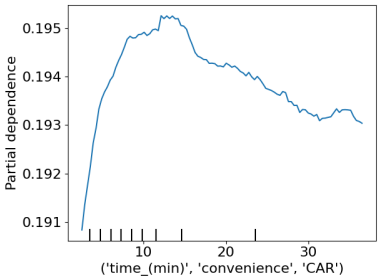


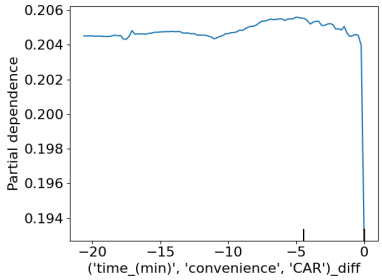


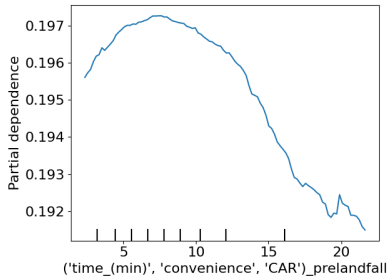


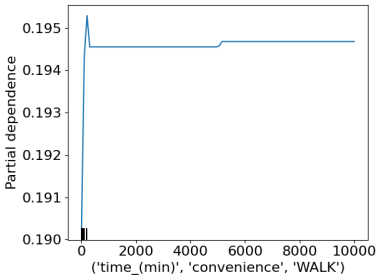


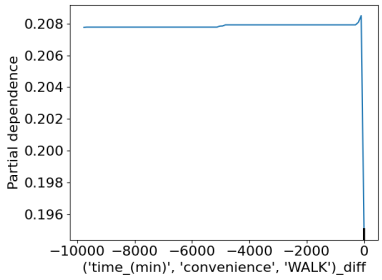


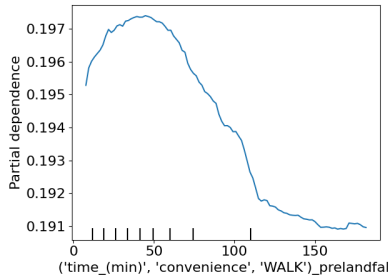


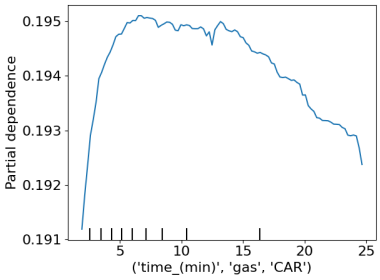


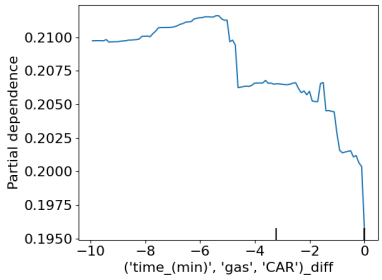


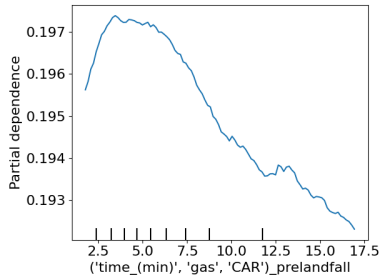

Partial dependence
('time_(min)', 'convenience', 'WALK')

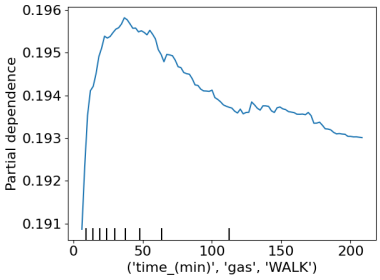

Partial dependence
('time_(min)', 'convenience', 'WALK')_diff

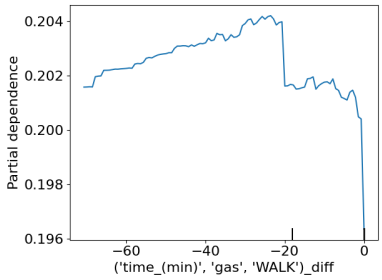

Partial dependence
('time_(min)', 'convenience', 'WALK')_prelandfall

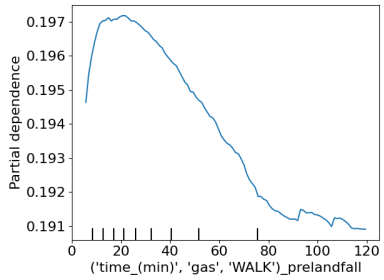

Partial dependence
('time_(min)', 'gas', 'CAR')

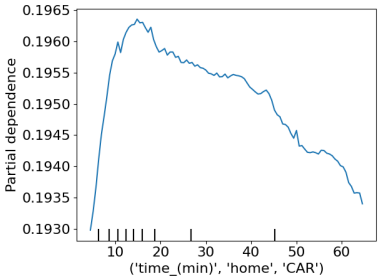

Partial dependence
('time_(min)', 'gas', 'CAR')_diff

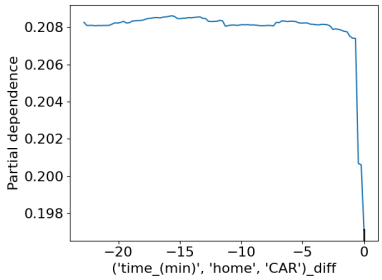

Partial dependence
('time_(min)', 'gas', 'CAR')_prelandfall

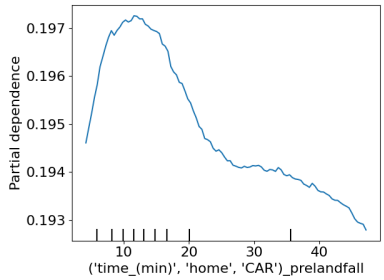

Partial dependence
('time_(min)', 'gas', 'WALK')

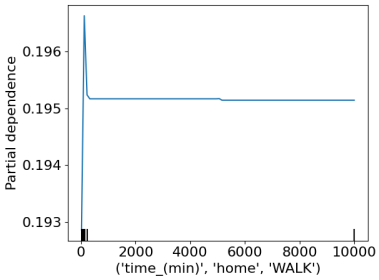

Partial dependence
('time_(min)', 'gas', 'WALK')_diff

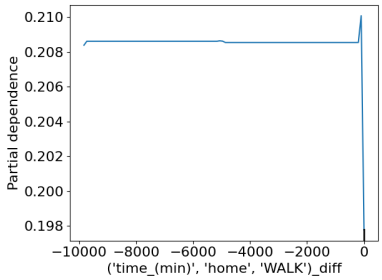

Partial dependence
('time_(min)', 'gas', 'WALK')_prelandfall

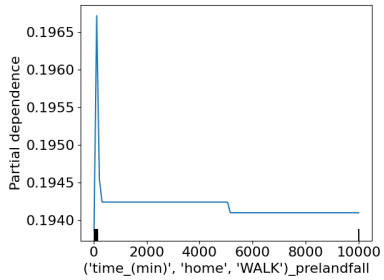

Partial dependence
('time_(min)', 'home', 'CAR')

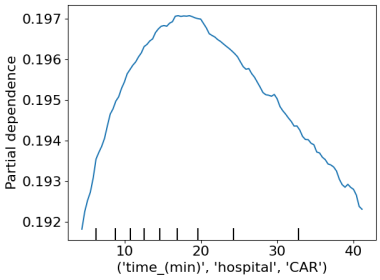

Partial dependence
('time_(min)', 'home', 'CAR')_diff

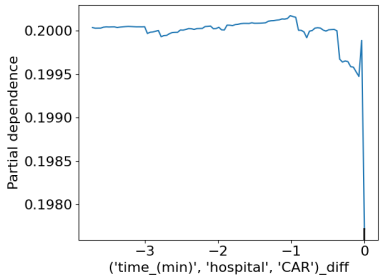

Partial dependence
('time_(min)', 'home', 'CAR')_prelandfall

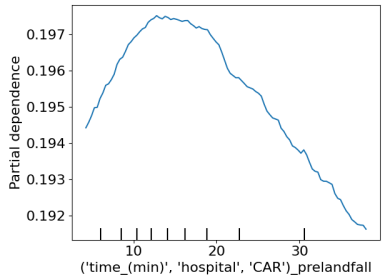

Partial dependence
('time_(min)', 'home', 'WALK')

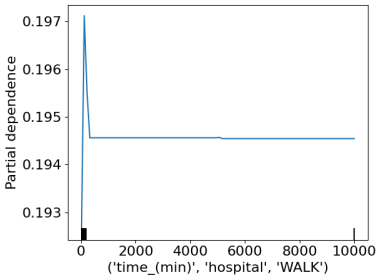

Partial dependence
('time_(min)', 'home', 'WALK')_diff

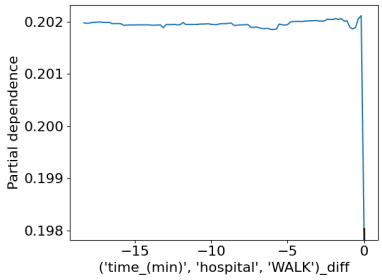

Partial dependence
('time_(min)', 'home', 'WALK')_prelandfall

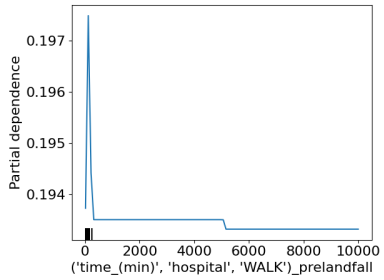

Partial dependence
0.197
0.196
0.195
0.194
0.193
0.192
10
20
30
40
('time_(min)', 'hospital', 'CAR')

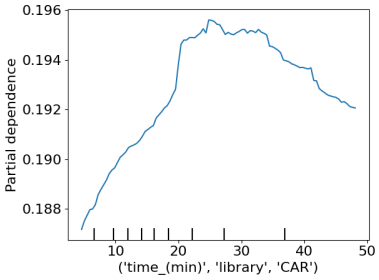

Partial dependence
0.2000
0.1995
0.1990
0.1985
0.1980
−3
−2
−1
0
('time_(min)', 'hospital', 'CAR')_diff

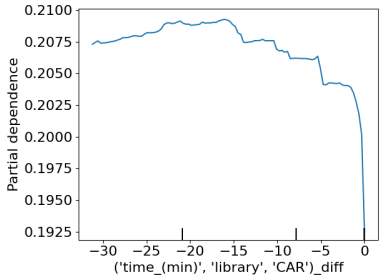

Partial dependence
0.197
0.196
0.195
0.194
0.193
0.192
10
20
30
('time_(min)', 'hospital', 'CAR')_prelandfall

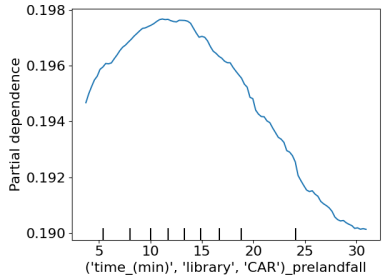

Partial dependence
0.197
0.196
0.195
0.194
0.193
0
2000
4000
6000
8000
10000
('time_(min)', 'hospital', 'WALK')

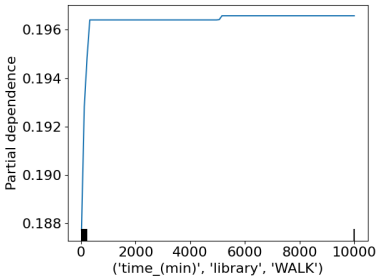

Partial dependence
0.202
0.201
0.200
0.199
0.198
−15
−10
−5
0
('time_(min)', 'hospital', 'WALK')_diff

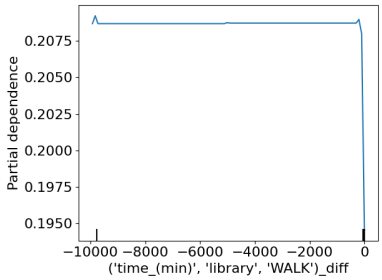

Partial dependence
0.197
0.196
0.195
0.194
0
2000
4000
6000
8000
10000
('time_(min)', 'hospital', 'WALK')_prelandfall

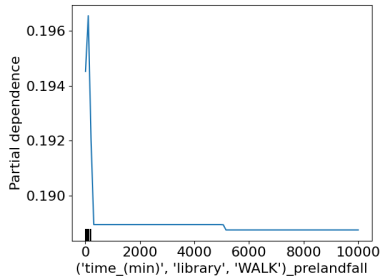

Partial dependence
0.196
0.194
0.192
0.190
0.188
10
20
30
40
50
('time_(min)', 'library', 'CAR')

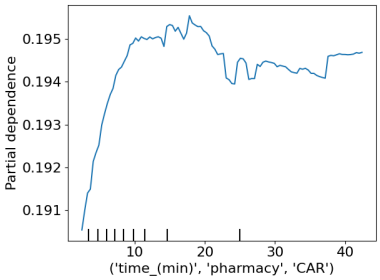

Partial dependence
0.2100
0.2075
0.2050
0.2025
0.2000
0.1975
0.1950
0.1925
−30
−25
−20
−15
−10
−5
0
('time_(min)', 'library', 'CAR')_diff

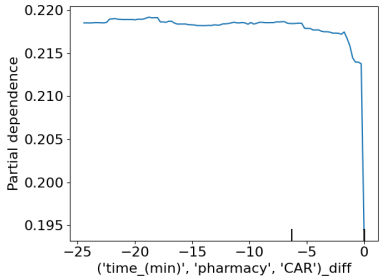

Partial dependence
0.198
0.196
0.194
0.192
0.190
5
10
15
20
25
30
('time_(min)', 'library', 'CAR')_prelandfall

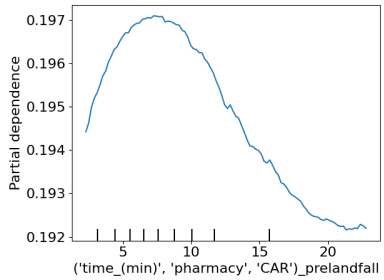

Partial dependence
0.196
0.194
0.192
0.190
0.188
0
2000
4000
6000
8000
10000
('time_(min)', 'library', 'WALK')

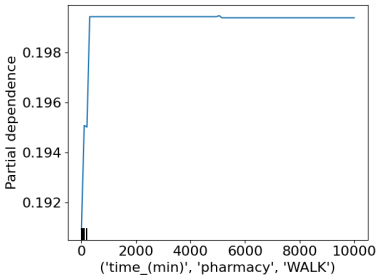

Partial dependence
0.2075
0.2050
0.2025
0.2000
0.1975
0.1950
−10000
−8000
−6000
−4000
−2000
0
('time_(min)', 'library', 'WALK')_diff

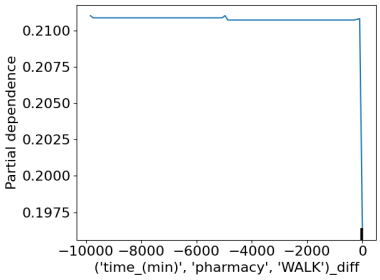

Partial dependence
0.196
0.194
0.192
0.190
0
2000
4000
6000
8000
10000
('time_(min)', 'library', 'WALK')_prelandfall

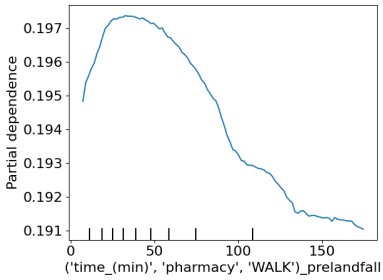

Partial dependence
0.195
0.194
0.193
0.192
0.191
10
20
30
40
('time_(min)', 'pharmacy', 'CAR')

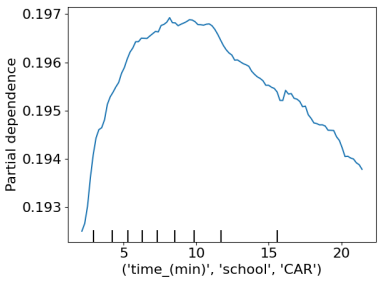

Partial dependence
0.220
0.215
0.210
0.205
0.200
0.195
−25
−20
−15
−10
−5
0
('time_(min)', 'pharmacy', 'CAR')_diff

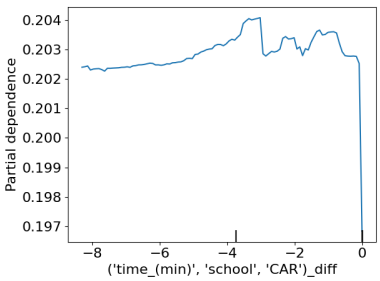

Partial dependence
0.197
0.196
0.195
0.194
0.193
0.192
5
10
15
20
('time_(min)', 'pharmacy', 'CAR')_prelandfall

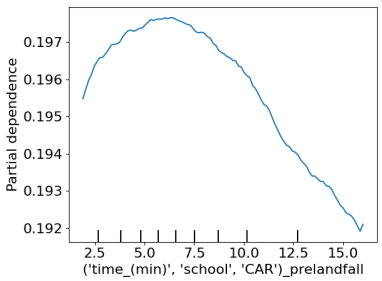

Partial dependence
0.198
0.196
0.194
0.192
0
2000
4000
6000
8000
10000
('time_(min)', 'pharmacy', 'WALK')

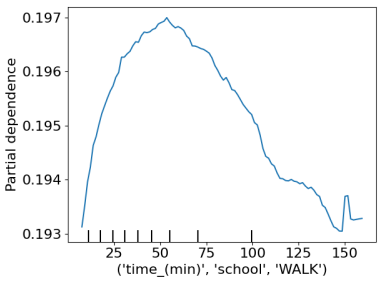

Partial dependence
0.2100
0.2075
0.2050
0.2025
0.2000
0.1975
−10000
−8000
−6000
−4000
−2000
0
('time_(min)', 'pharmacy', 'WALK')_diff

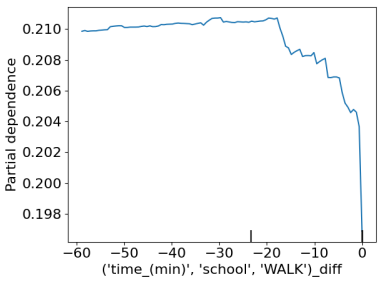

Partial dependence
0.197
0.196
0.195
0.194
0.193
0.192
0.191
0
50
100
150
('time_(min)', 'pharmacy', 'WALK')_prelandfall

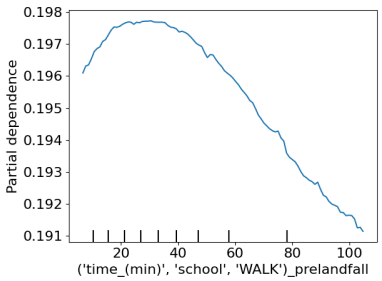

Partial dependence
0.197
0.196
0.195
0.194
0.193
5
10
15
20
('time_(min)', 'school', 'CAR')

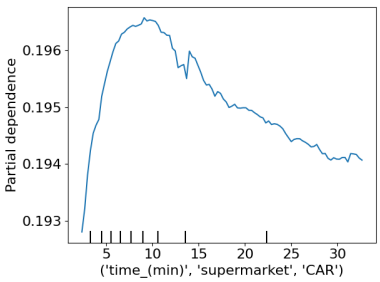

Partial dependence
0.204
0.203
0.202
0.201
0.200
0.199
0.198
0.197
−8
−6
−4
−2
0
('time_(min)', 'school', 'CAR')_diff

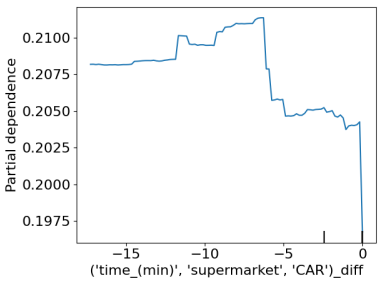

Partial dependence
0.197
0.196
0.195
0.194
0.193
0.192
2.5
5.0
7.5
10.0
12.5
15.0
('time_(min)', 'school', 'CAR')_prelandfall

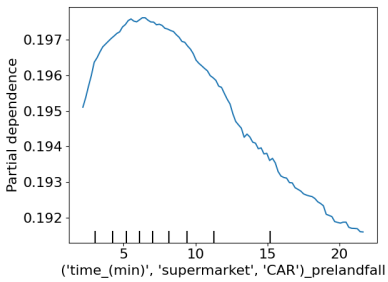

Partial dependence
0.197
0.196
0.195
0.194
0.193
25
50
75
100
125
150
('time_(min)', 'school', 'WALK')

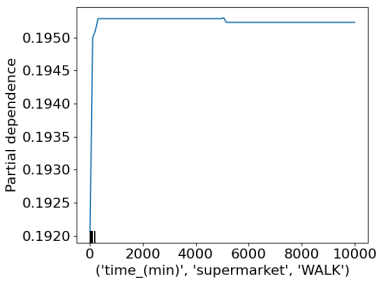

Partial dependence
0.210
0.208
0.206
0.204
0.202
0.200
0.198
−60
−50
−40
−30
−20
−10
0
('time_(min)', 'school', 'WALK')_diff

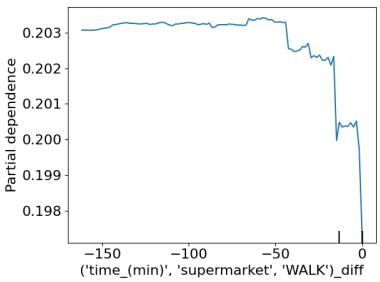

Partial dependence
0.198
0.197
0.196
0.195
0.194
0.193
0.192
0.191
20
40
60
80
100
('time_(min)', 'school', 'WALK')_prelandfall

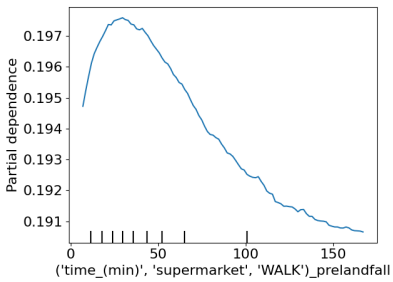

Partial dependence
0.196
0.195
0.194
0.193
5
10
15
20
25
30
('time_(min)', 'supermarket', 'CAR')

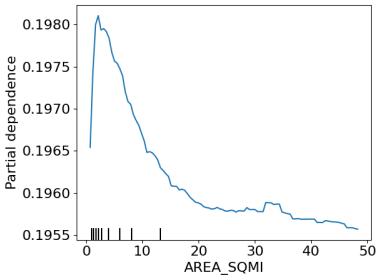

Partial dependence
0.2100
0.2075
0.2050
0.2025
0.2000
0.1975
−15
−10
−5
0
('time_(min)', 'supermarket', 'CAR')_diff

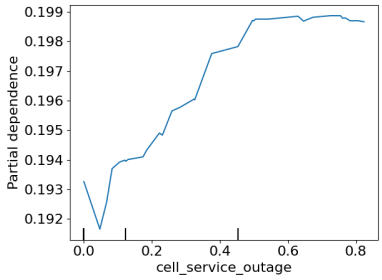

Partial dependence
0.197
0.196
0.195
0.194
0.193
0.192
5
10
15
20
('time_(min)', 'supermarket', 'CAR')_prelandfall

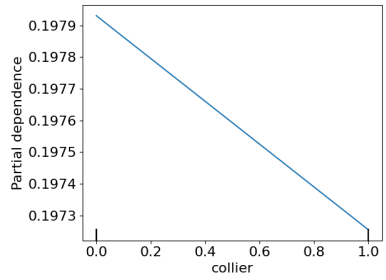

Partial dependence
0.1950
0.1945
0.1940
0.1935
0.1930
0.1925
0.1920
0
2000
4000
6000
8000
10000
('time_(min)', 'supermarket', 'WALK')

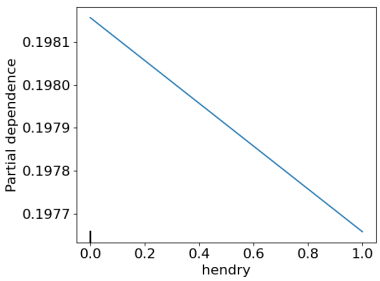

Partial dependence
0.203
0.202
0.201
0.200
0.199
0.198
−150
−100
−50
0
('time_(min)', 'supermarket', 'WALK')_diff

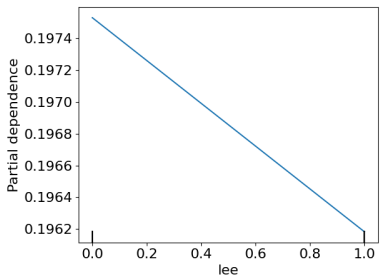

Partial dependence
0.197
0.196
0.195
0.194
0.193
0.192
0.191
0
50
100
150
('time_(min)', 'supermarket', 'WALK')_prelandfall

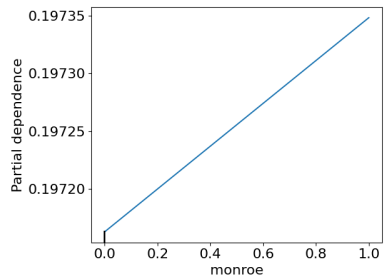

Partial dependence
0.1980
0.1975
0.1970
0.1965
0.1960
0.1955
0 10 20 30 40 50
AREA_SQMI

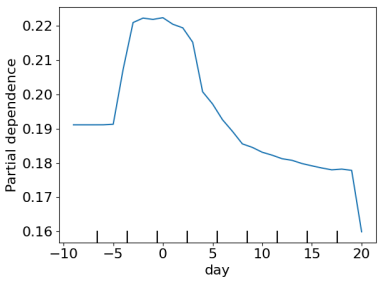

Partial dependence
0.199
0.198
0.197
0.196
0.195
0.194
0.193
0.192
0.0 0.2 0.4 0.6 0.8
cell_service_outage

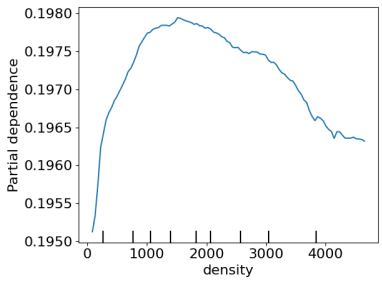

Partial dependence
0.1979
0.1978
0.1977
0.1976
0.1975
0.1974
0.1973
0.0 0.2 0.4 0.6 0.8 1.0
collier

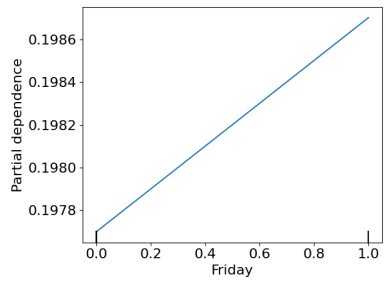

Partial dependence
0.1981
0.1980
0.1979
0.1978
0.1977
0.0 0.2 0.4 0.6 0.8 1.0
hendry

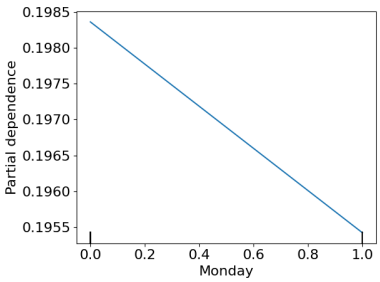

Partial dependence
0.1974
0.1972
0.1970
0.1968
0.1966
0.1964
0.1962
0.0 0.2 0.4 0.6 0.8 1.0
lee

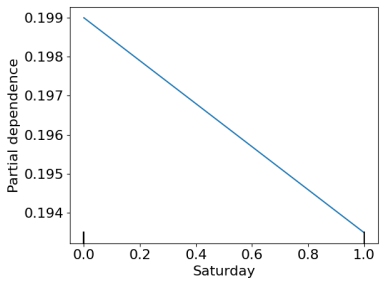

Partial dependence
0.19735
0.19730
0.19725
0.19720
0.0 0.2 0.4 0.6 0.8 1.0
monroe

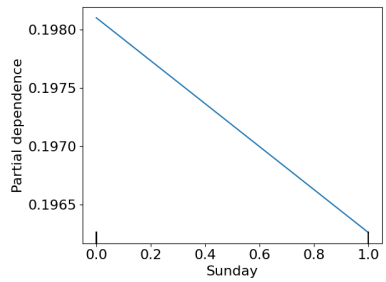

Partial dependence
0.22
0.21
0.20
0.19
0.18
0.17
0.16
−10 −5 0 5 10 15 20
day

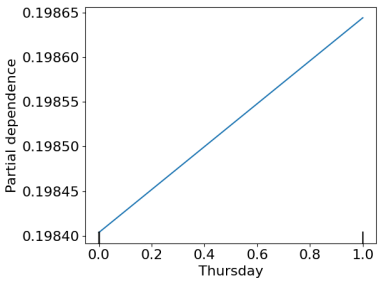

Partial dependence
0.1980
0.1975
0.1970
0.1965
0.1960
0.1955
0.1950
0 1000 2000 3000 4000
density

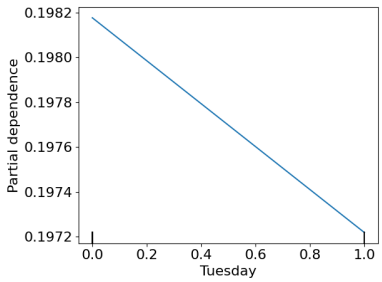

Partial dependence
0.1986
0.1984
0.1982
0.1980
0.1978
0.0 0.2 0.4 0.6 0.8 1.0
Friday

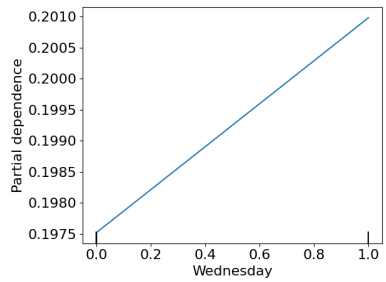

Partial dependence
0.1985
0.1980
0.1975
0.1970
0.1965
0.1960
0.1955
0.0 0.2 0.4 0.6 0.8 1.0
Monday

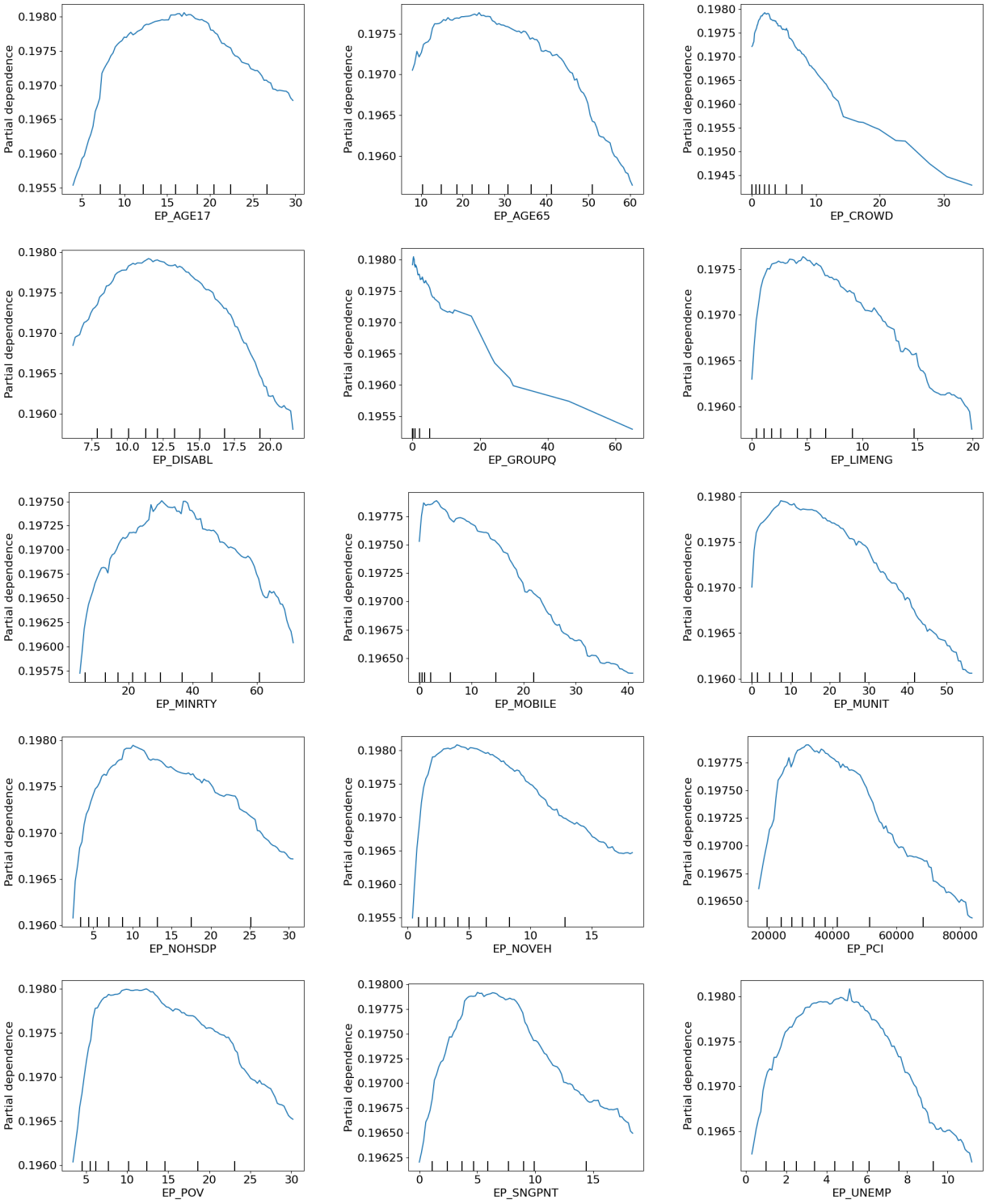

Partial dependence
0.199
0.198
0.197
0.196
0.195
0.194
0.0 0.2 0.4 0.6 0.8 1.0
Saturday

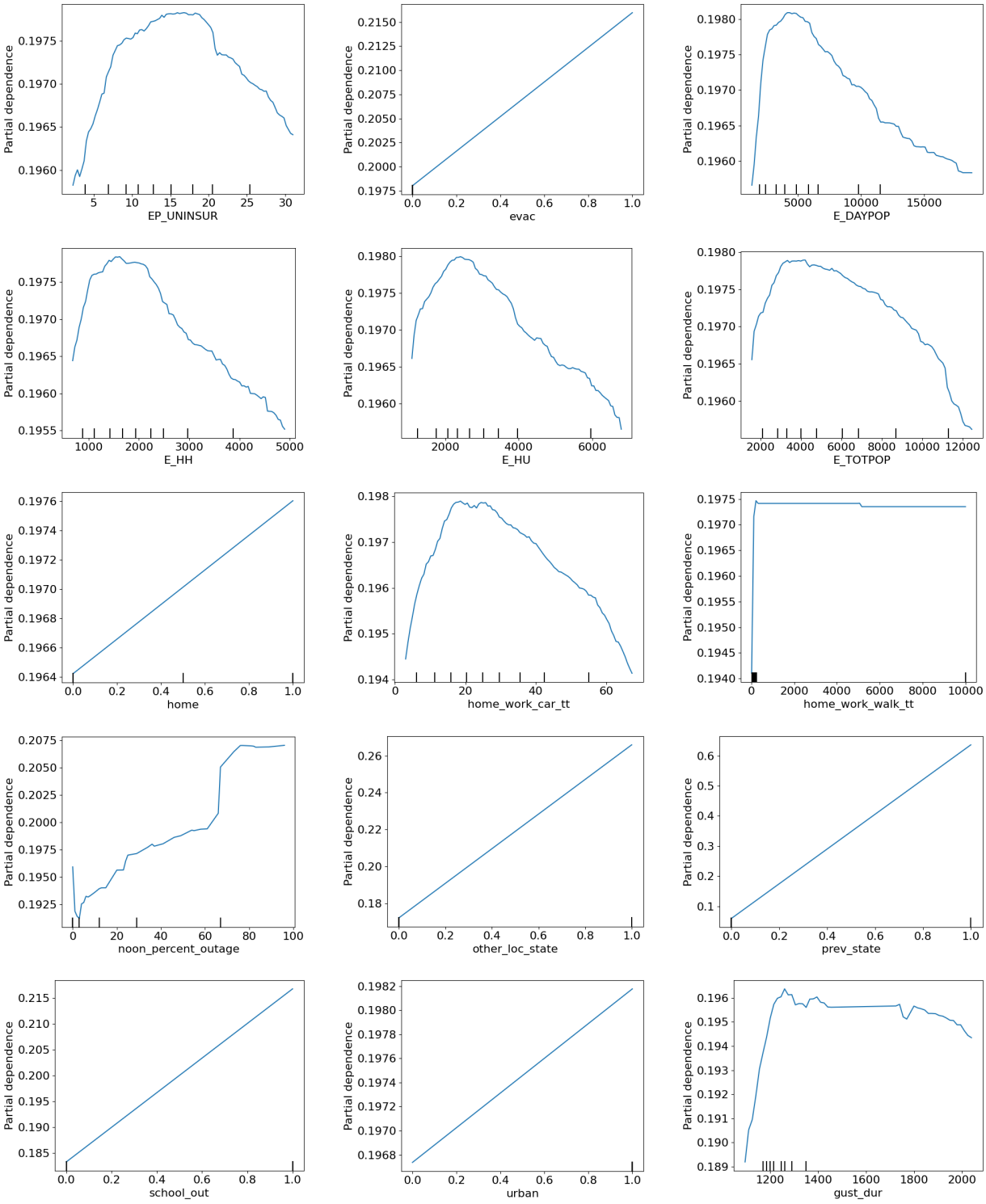

Partial dependence
0.1980
0.1975
0.1970
0.1965
0.0 0.2 0.4 0.6 0.8 1.0
Sunday


Partial dependence
0.19865
0.19860
0.19855
0.19850
0.19845
0.19840
0.0 0.2 0.4 0.6 0.8 1.0
Thursday


Partial dependence
0.1982
0.1980
0.1978
0.1976
0.1974
0.1972
0.0 0.2 0.4 0.6 0.8 1.0
Tuesday


Partial dependence
0.2010
0.2005
0.2000
0.1995
0.1990
0.1985
0.1980
0.1975
0.0 0.2 0.4 0.6 0.8 1.0
Wednesday


Partial dependence
EP_AGE17
EP_AGE65
EP_CROWD
EP_DISABL
EP_GROUPQ
EP_LIMENG
EP_MINRTY
EP_MOBILE
EP_MUNIT
EP_NOHSDP
EP_NOVEH
EP_PCI
EP_POV
EP_SNGPNT
EP_UNEMP


Partial dependence
EP_UNINSUR
evac
E_DAYPOP
E_HH
E_HU
E_TOTPOP
home
home_work_car_tt
home_work_walk_tt
noon_percent_outage
other_loc_state
prev_state
school_out
urban
gust_dur

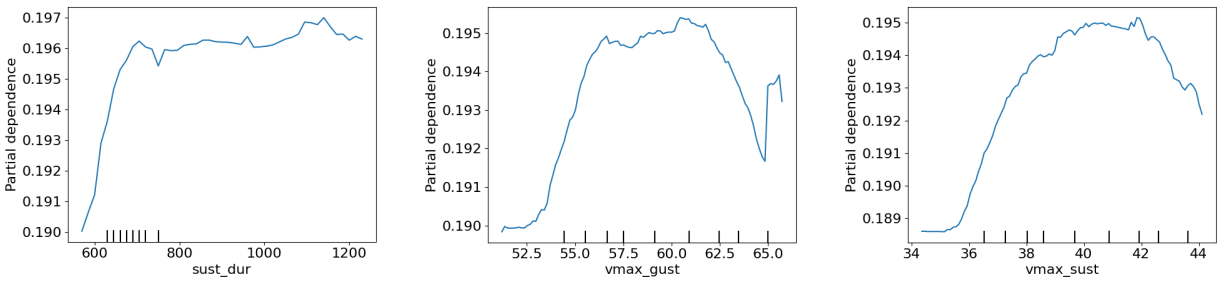


Figure H.1: Partial dependence plots for random forest model estimating recovery for Collier, Monroe, Hendry, and Lee Counties

BIBLIOGRAPHY

[1] United Nations Publications. *Global Assessment Report on Disaster Risk Reduction 2022: Our World at Risk: Transforming Governance for a Resilient Future*. UN, June 2022.

[2] Leah Platt Boustan, Matthew E Kahn, Paul W Rhode, and Maria Lucia Yanguas. The effect of natural disasters on economic activity in us counties: A century of data. *Journal of Urban Economics*, 118:103257, 2020.

[3] Stanley A Changnon, Roger A Pielke, David Changnon, Richard T Sylves, and Roger Pulwarty. Human factors explain the increased losses from weather and climate extremes. *Bulletin of the American Meteorological Society*, 81(3):437–442, March 2000.

[4] Rebekah Green, Lisa K Bates, and Andrew Smyth. Impediments to recovery in new orleans' upper and lower ninth ward: one year after hurricane katrina. *Disasters*, 31(4):311–335, December 2007.

[5] Diana Mitsova, Ann-Margaret Esnard, Alka Sapat, and Betty S Lai. Socioeconomic vulnerability and electric power restoration timelines in florida: the case of hurricane irma. *Natural Hazards*, 94(2):689–709, November 2018.

[6] Cheng-Chun Lee, Mikel Maron, and Ali Mostafavi. Community-scale big data reveals disparate impacts of the texas winter storm of 2021 and its managed power outage. *Humanities and Social Sciences Communications*, 9(1):335, September 2022.

[7] Takahiro Yabe, Kota Tsubouchi, Naoya Fujiwara, Yoshihide Sekimoto, and Satish V Ukkusuri. Understanding post-disaster population recovery patterns. *Journal of The Royal Society Interface*, 17(163):20190532, February 2020.

[8] Brett F Sanders, Jochen E Schubert, Daniel T Kahl, Katharine J Mach, David Brady, Amir AghaKouchak, Fonna Forman, Richard A Matthew, Nicola Ulibarri, and Steven J Davis. Large and inequitable flood risks in Los Angeles, California. *Nature Sustainability*, 6(1):47–57, 2023.

[9] Marccus D Hendricks and Shannon Van Zandt. Unequal protection revisited: Planning for environmental justice, hazard vulnerability, and critical infrastructure in communities of color. *Environmental Justice*, 14(2):87–97, April 2021.

[10] Paolo Gardoni and Colleen Murphy. Society-based design: promoting societal well-being by designing sustainable and resilient infrastructure. *Sustainable and Resilient Infrastructure*, 5(1-2):4–19, January 2020.

[11] Colleen Murphy and Paolo Gardoni. The role of society in engineering risk analysis: a capabilities-based approach. *Risk Analysis*, 26(4):1073–1083, August 2006.

[12] David R Godschalk. Urban hazard mitigation: Creating resilient cities. *Natural Hazards Review*, 4(3):136–143, 2003.

[13] Robert Bolin and Lois Stanford. The northridge earthquake: community-based approaches to unmet recovery needs. *Disasters*, 22(1):21–38, 1998.

[14] Melissa L Finucane, Linnea Warren May, and Joan Chang. *A Scoping Literature Review on Indicators and Metrics for Assessing Racial Equity in Disaster Preparation, Response, and Recovery*. RAND, 2021.

[15] Billie L Turner, Roger E Kasperson, Pamela A Matson, James J McCarthy, Robert W Corell, Lindsey Christensen, Noelle Eckley, Jeanne X Kasperson, Amy Luers, Marybeth L Martello, et al. A framework for vulnerability analysis in sustainability science. *Proceedings of the National Academy of Sciences*, 100(14):8074–8079, 2003.

[16] Deniz Berfin Karakoc, Kash Barker, Christopher W Zobel, and Yasser Almoghathawi. Social vulnerability and equity perspectives on interdependent infrastructure network component importance. *Sustainable Cities and Society*, 57:102072, June 2020.

[17] Melissa L Finucane, Joie Acosta, Amanda Wicker, and Katie Whipkey. Short-Term solutions to a Long-Term challenge: Rethinking disaster recovery planning to reduce vulnerabilities and inequities. *International Journal of Environmental Research and Public Health*, 17(2), January 2020.

[18] Amy McDermott. Herd immunity is an important—and often misunderstood—public health phenomenon. *Proceedings of the National Academy of Sciences*, 118(21):e2107692118, 2021.

[19] Anna G White, Seth D Guikema, and Tom M Logan. Urban population characteristics and their correlation with historic discriminatory housing practices. *Applied Geography*, 132:102445, July 2021.

[20] Benjamin Rachunok and Roshanak Nateghi. Overemphasis on recovery inhibits community transformation and creates resilience traps. *Nature Communications*, 12(1):7331, December 2021.

[21] Tom M Logan, Mitchell J Anderson, Timothy G Williams, and Lindsey Conrow. Measuring inequalities in urban systems: An approach for evaluating the distribution of amenities and burdens. *Computers, Environment and Urban Systems*, 86:101590, 2021.

[22] Thushara Gunda, Amanda Wachtel, Shruti Khadka Mishra, and Emily Moog. Quantitative approaches for including equity in risk and resilience infrastructure planning analyses. *Risk Analysis*, September 2023.

[23] Terje Aven. On some recent definitions and analysis frameworks for risk, vulnerability, and resilience. *Risk Analysis*, 31(4):515–522, April 2011.

[24] Thomas Elmqvist, Erik Andersson, Niki Frantzeskaki, Timon McPhearson, Per Olsson, Owen Gaffney, Kazuhiko Takeuchi, and Carl Folke. Sustainability and resilience for transformation in the urban century. *Nature Sustainability*, 2(4):267–273, April 2019.

[25] Sara Meerow, Joshua P Newell, and Melissa Stults. Defining urban resilience: A review. *Landscape and Urban Planning*, 147:38–49, March 2016.

[26] Jeryang Park, Thomas P Seager, Palakurth Suresh Chandra Rao, Matteo Convertino, and Igor Linkov. Integrating risk and resilience approaches to catastrophe management in engineering systems. *Risk Analysis*, 33(3):356–367, 2013.

[27] Maria Koliou, John W van de Lindt, Therese P McAllister, Bruce R Ellingwood, Maria Dillard, and Harvey Cutler. State of the research in community resilience: Progress and challenges. *Sustainable and Resilient Infrastructure*, 5(3):131–151, 2020.

[28] Tom M Logan and Seth D Guikema. Reframing resilience: Equitable access to essential services. *Risk Analysis*, 40(8):1538–1553, August 2020.

[29] Mai Thi Nguyen and Emma Boundy. Big data and smart (equitable) cities. In Piyushimita (vonu) Thakuriah, Nebiyou Tilahun, and Moira Zellner, editors, *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics*, pages 517–542. Springer International Publishing, Cham, 2017.

[30] M Niyazi and J Behnamian. Application of emerging digital technologies in disaster relief operations: A systematic review. *Archives of Computational Methods in Engineering*, 30(3):1579–1599, April 2023.

[31] Moumita Basu, Anurag Shandilya, Prannay Khosla, Kripabandhu Ghosh, and Saptarshi Ghosh. Extracting resource needs and availabilities from microblogs for aiding Post-Disaster relief operations. *IEEE Transactions on Computational Social Systems*, 6(3):604–618, June 2019.

[32] Takahiro Yabe, Nicholas KW Jones, P Suresh C Rao, Marta C Gonzalez, and Satish V Ukkusuri. Mobile phone location data for disasters: A review from natural hazards and epidemics. *Computers, Environment and Urban Systems*, 94:101777, 2022.

[33] T M Logan, T G Williams, A J Nisbet, K D Liberman, C T Zuo, and S D Guikema. Evaluating urban accessibility: leveraging open-source data and analytics to overcome existing limitations. *Environment and Planning B: Urban Analytics and City Science*, 46(5):897–913, June 2019.

[34] Alberto Vanolo. Is there anybody out there? the place and role of citizens in tomorrow's smart cities. *Futures*, 82:26–36, September 2016.

[35] Marta C González, César A Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.

[36] Francesco Calabrese, Mi Diao, Giusy Di Lorenzo, Joseph Ferreira, and Carlo Ratti. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26:301–313, January 2013.

[37] Pu Wang, Timothy Hunter, Alexandre M Bayen, Katja Schechtner, and Marta C González. Understanding road usage patterns in urban areas. *Scientific Reports*, 2:1001, December 2012.

[38] Cynthia Chen, Jingtao Ma, Yusak Susilo, Yu Liu, and Menglin Wang. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, 68:285–299, July 2016.

[39] Robert Goodspeed, Meixin Yuan, Aaron Krusniak, and Tierra Bills. Assessing the value of new big data sources for transportation planning: Benton harbor, michigan case study. In S C M Geertman, Christopher Pettit, Robert Goodspeed, and Aija Staffans, editors, *Urban Informatics and Future Cities*, pages 127–150. Springer International Publishing, Cham, 2021.

[40] M Keith Chen and Ryne Rohla. The effect of partisanship and political advertising on close family ties. *Science*, 360(6392):1020–1024, 2018.

[41] Shan Jiang, Joseph Ferreira, and Marta C Gonzalez. Activity-Based human mobility patterns inferred from mobile phone data: A case study of singapore. *IEEE Transactions on Big Data*, 3(2):208–219, 2017.

[42] Stephanie E Chang. Transportation planning for disasters: An accessibility approach. *Environment and Planning A*, 35(6):1051–1072, June 2003.

[43] Shanjiang Zhu and David M Levinson. Disruptions to transportation networks: A review. In *Network Reliability in Practice*, pages 5–20. Springer New York, 2012.

[44] Genevieve Giuliano and Jacqueline Golob. Impacts of the northridge earthquake on transit and highway use. *Journal of Transportation and Statistics*, 1(2):1–20, 1998.

[45] Kontou Eleftheria, Murray-Tuite Pamela, and Wernstedt Kris. Commuter adaptation in response to hurricane sandy's damage. *Natural Hazards Review*, 18(2):04016010, May 2017.

[46] Susan L Cutter. The landscape of disaster resilience indicators in the USA. *Natural Hazards*, 80(2):741–758, January 2016.

[47] Tim G Frazier, Courtney M Thompson, and Raymond J Dezzani. A framework for the development of the SERV model: A spatially explicit Resilience-Vulnerability model. *Applied Geography*, 51:158–172, July 2014.

[48] Luca Pappalardo, Dino Pedreschi, Zbigniew Smoreda, and Fosca Giannotti. Using big data to study the link between human mobility and socio-economic development. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 871–878, 2015.

[49] Roy Penchansky and J William Thomas. The concept of access: definition and relationship to consumer satisfaction. *Medical Care*, pages 127–140, 1981.

[50] Emily Saurman. Improving access: modifying penchansky and thomas's theory of access. *Journal of Health Services Research & Policy*, 21(1):36–39, January 2016.

[51] Veraset. https://www.veraset.com/, 2023. Accessed: 2023-6-15.

[52] Stephen Wong, Susan Shaheen, and Joan Walker. Understanding evacuee behavior: A case study of hurricane irma. *Institute of Transportation Studies, Research Reports, Working Papers, Proceedings*, December 2018.

[53] Mitsova Diana, Esnard Ann-Margaret, Sapat Alka, Lamadrid Alberto, Escaleras Monica, and Velarde-Perez Catherine. Effects of infrastructure service disruptions following hurricane irma: Multilevel analysis of postdisaster recovery outcomes. *Natural Hazards Review*, 22(1):04020055, February 2021.

[54] Amanda Coston, Neel Guha, Derek Ouyang, Lisa Lu, Alexandra Chouldechova, and Daniel E Ho. Leveraging administrative data for bias audits: Assessing disparate coverage with mobility data for COVID-19 policy. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pages 173–184, New York, NY, USA, March 2021. Association for Computing Machinery.

[55] Kerina Helen Jones, Helen Daniels, Sharon Heys, and David Vincent Ford. Challenges and potential opportunities of mobile phone call detail records in health research: Review. *JMIR Mhealth Uhealth*, 6(7):e161, July 2018.

[56] Nuria Oliver, Aleksandar Matic, and Enrique Frias-Martinez. Mobile network data for public health: Opportunities and challenges. *Frontiers in Public Health*, 3:189, August 2015.

[57] Benjamin C M Fung, Ke Wang, Rui Chen, and Philip S Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, 42(4):1–53, June 2010.

[58] National Institute of Standards and Technology. NIST PRIVACY FRAMEWORK: A TOOL FOR IMPROVING PRIVACY THROUGH ENTERPRISE RISK MANAGEMENT. Technical report, National Institute of Standards and Technology, Gaithersburg, MD, January 2020.

[59] Jonathan Cinnamon, Sarah K Jones, and W Neil Adger. Evidence and future potential of mobile phone data for disease disaster management. *Geoforum*, 75:253–264, October 2016.

[60] Bart Custers and Helena Uršič. Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection. *International Data Privacy Law*, 6(1):4–15, January 2016.

[61] David Lyon. Surveillance. *Internet Policy Review*, 11(4), November 2022.

[62] Linnet Taylor. The price of certainty: How the politics of pandemic data demand an ethics of care. *Big Data & Society*, 7(2):2053951720942539, July 2020.

[63] Valerie Washington, Seth Guikema, Joi-Lynn Mondisa, and Aditi Misra. A data-driven method for identifying the locations of hurricane evacuations from mobile phone location data. *Risk Analysis*, August 2023.

[64] Tim G Frazier, Courtney M Thompson, Ray J Dezzani, and Danielle Butsick. Spatial and temporal quantification of resilience at the community scale. *Applied Geography*, 42:95–107, August 2013.

[65] T M Logan, M J Anderson, and A C Reilly. Risk of isolation increases the expected burden from sea-level rise. *Nature Climate Change*, March 2023.

[66] Misty L Heggeness. Estimating the immediate impact of the COVID-19 shock on parental attachment to the labor market and the double bind of mothers. *Review of Economics of the Household*, 18(4):1053–1078, October 2020.

[67] Monica Gentili, Kim Isett, Nicoleta Serban, and Julie Swann. Small-Area estimation of spatial access to care and its implications for policy. *Journal of Urban Health*, 92(5):864–909, October 2015.

[68] Michael J Widener, Leia Minaker, Steven Farber, Jeff Allen, Brigitte Vitali, Paul C Coleman, and Brian Cook. How do changes in the daily food and transportation environments affect grocery store accessibility? *Applied Geography*, 83:46–62, June 2017.

[69] United Nations Educational, Scientific and Cultural Organization and World Bank. *Culture in City Reconstruction and Recovery*. Paris: UNESCO. © UNESCO and World Bank., 2018.

[70] Zhenzhen Wang, Sylvia Y He, and Yee Leung. Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, 11:141–155, April 2018.

[71] Feilong Wang and Cynthia Chen. On data processing required to derive mobility patterns from passively-generated mobile phone data. *Transportation Research Part C: Emerging Technologies*, 87:58–74, February 2018.

[72] Serdar Çolak, Lauren P Alexander, Bernardo G Alvim, Shomik R Mehndiretta, and Marta C González. Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record: Journal of the Transportation Research Board*, 2526:126–135, 2015.

[73] Gregory Bucci and Tom Morton. Cell phone data and travel behavior research. Technical report, FHWA, July 2014.

[74] Kyra H Grantz, Hannah R Meredith, Derek A T Cummings, C Jessica E Metcalf, Bryan T Grenfell, John R Giles, Shruti Mehta, Sunil Solomon, Alain Labrique, Nishant Kishore, Caroline O Buckee, and Amy Wesolowski. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature Communications*, 11(1):4961, September 2020.

[75] Amy Wesolowski, Caroline O Buckee, Kenth Engø-Monsen, and C J E Metcalf. Connecting mobility to infectious diseases: The promise and limits of mobile phone data. *The Journal of Infectious Diseases*, 214(suppl_4):S414–S420, December 2016.

[76] Aude Marzuoli and Fengmei Liu. Monitoring of natural disasters through anomaly detection on mobile phone data. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 4089–4098. IEEE, 2019.

[77] Ling Yin, Jie Chen, Hao Zhang, Zhile Yang, Qiao Wan, Li Ning, Jinxing Hu, and Qi Yu. Improving emergency evacuation planning with mobile phone location data. *Environment and Planning B: Urban Analytics and City Science*, 47(6):964–980, July 2020.

[78] Manzhu Yu, Chaowei Yang, and Yun Li. Big data in natural disaster management: A review. *Geosciences Journal*, 8(5):165, May 2018.

[79] María Martínez-Rojas, María del Carmen Pardo-Ferreira, and Juan Carlos Rubio-Romero. Twitter as a tool for the management and analysis of emergency situations: A systematic literature review. *International Journal of Information Management*, 43:196–208, December 2018.

[80] Maja Kucharczyk and Chris H Hugenholtz. Remote sensing of natural hazard-related disasters with small drones: Global trends, biases, and research opportunities. *Remote Sensing of Environment*, 264:112577, October 2021.

[81] Yanxin Wang, Jian Li, Xi Zhao, Gengzhong Feng, and Xin Robert Luo. Using mobile phone data for emergency management: a systematic literature review. *Information Systems Frontiers*, pages 1–21, September 2020.

[82] Andrew A Cook, Göksel Mısırlı, and Zhong Fan. Anomaly detection for IoT Time-Series data: A survey. *IEEE Internet of Things Journal*, 7(7):6481–6494, 2019.

[83] Jan G De Gooijer and Rob J Hyndman. 25 years of time series forecasting. *International Journal of Forecasting*, 22(3):443–473, January 2006.

[84] Markus Goldstein and Seiichi Uchida. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLoS One*, 11(4):e0152173, April 2016.

[85] Salima Omar, Asri Ngadi, and Hamid H. Jebur. Machine learning techniques for anomaly detection: An overview. *International Journal of Computer Applications*, 79(2):33–41, October 2013.

[86] National Hurricane Center. NATIONAL HURRICANE CENTER TROPICAL CYCLONE REPORT: Hurricane irma. Technical report, National Oceanic and Atmospheric Administration, June 2018.

[87] IRMA graphics archive. https://www.nhc.noaa.gov/archive/2017/IRMA_graphics.php?product=3day_cone_with_line_and_wind, 2017. Accessed: 2021-4-8.

[88] Hurricane irma local Report/Summary. https://www.weather.gov/mfl/hurricaneirma, July 2018. Accessed: 2021-3-1.

[89] Hurricane irma. https://www.monroecounty-fl.gov/982/Hurricane-Irma, 2017. Accessed: 2021-3-1.

[90] Hagerty Consulting, Inc. Collier county hurricane irma response After-Action report. Technical report, Collier County, October 2017.

[91] Collier County Public [@collierschools]. FINAL: 2017-18 academic calendar approved at 9/26 school board meeting. https://twitter.com/collierschools/status/912869132839899136, September 2017. Accessed: 2021-5-18.

[92] Kyra Gurney and Miami. Some florida keys schools to open monday. for others hit by irma, it could be much longer. *Miami Herald*, September 2017.

[93] Mike Snider. Disney world dodges major hurricane irma damage. *USA Today*, September 2017.

[94] OpenStreetMap contributors. Florida retrieved from https://download.geofabrik.de/north-america/us/florida.html. https://www.openstreetmap.org, 2023.

[95] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297, December 2021.

[96] Bang Xiang Yong and Alexandra Brintrup. Bayesian autoencoders with uncertainty quantification: Towards trustworthy anomaly detection. *Expert Systems with Applications*, 209:118196, December 2022.

[97] A J Scott and M Knott. A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 30(3):507–512, 1974.

[98] Rebecca Killick and Idris Eckley. changepoint: An r package for changepoint analysis. *Journal of Statistical Software*, 58(3):1–19, 2014.

[99] Rebecca Killick and Idris A Eckley. changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, 58:1–19, June 2014.

[100] SNAP store locations. https://usda-fns.hub.arcgis.com/datasets/USDA-FNS::snap-store-locations/explore?location=7.499992%2C-159.172902%2C1.92. Accessed: 2023-9-25.

[101] Florida Department of Health. Health care facilities in florida. https://hub.arcgis.com/datasets/geoplatform::health-care-facilities-in-florida-2014?geometry=-104.490%2C24.338%2C-62.632%2C31.139, 2014. Accessed: 2021-4-8.

[102] Timothy K M Beatty, Jay P Shimshack, and Richard J Volpe. Disaster preparedness and disaster response: Evidence from sales of emergency supplies before and after hurricanes. *Journal of the Association of Environmental and Resource Economists*, 6(4):633–668, July 2019.

[103] Howard Cohen. Four shelters open in the keys as hurricane irma's track shifts toward island chain. *Miami Herald*, September 2017.

[104] Anna Camille Svirsko, Tom Logan, Christina Domanowski, and Daphne Skipper. Developing robust facility reopening processes following natural disasters. *Operations Research Forum*, 3(3), August 2022.

[105] Armita Kar, Yasuyuki Motoyama, Andre L Carrel, Harvey J Miller, and Huyen T K Le. COVID-19 exacerbates unequal food access. *Applied Geography*, 134:102517, September 2021.

[106] Susan L Cutter, Bryan J Boruff, and W Lynn Shirley. Social vulnerability to environmental hazards. *Social Science Quarterly*, 84(2):242–261, June 2003.

[107] Sapam Ranabir Singh, Mohammad Reza Eghdami, and Sarbjeet Singh. The concept of social vulnerability: A review from disasters perspectives. *International Journal of Interdisciplinary and Multidisciplinary Studies*, 1(6):71–82, 2014.

[108] Junia Howell and James R Elliott. Damages done: The longitudinal impacts of natural hazards on wealth inequality in the united states. *Social Problems*, 66(3):448–467, August 2018.

[109] Kelsea Best, Siobhan Kerr, Allison Reilly, Anand Patwardhan, Deb Niemeier, and Seth Guikema. Spatial regression identifies socioeconomic inequality in multi-stage power outage recovery after hurricane isaac. *Natural Hazards*, February 2023.

[110] Dapeng Yu, Jie Yin, Robert L Wilby, Stuart N Lane, Jeroen C J H Aerts, Ning Lin, Min Liu, Hongyong Yuan, Jianguo Chen, Christel Prudhomme, Mingfu Guan, Avinoam Baruch, Charlie W D Johnson, Xi Tang, Lizhong Yu, and Shiyuan Xu. Disruption of emergency response to vulnerable populations during floods. *Nature Sustainability*, 3(9):728–736, May 2020.

[111] Francesco Calabrese, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. Estimating Origin-Destination flows using mobile phone location data. *IEEE Pervasive Computing*, 10(4):36–44, 2011.

[112] Elisa F Long, M Keith Chen, and Ryne Rohla. Political storms: Emergent partisan skepticism of hurricane risks. *Science Advances*, 6(37), September 2020.

[113] Takahiro Yabe, Yoshihide Sekimoto, Kota Tsubouchi, and Satoshi Ikemoto. Cross-comparative analysis of evacuation behavior after earthquakes using mobile phone data. *PLoS One*, 14(2):e0211375, February 2019.

[114] Barbara Furletti, Lorenzo Gabrielli, Chiara Renso, and Salvatore Rinzivillo. Analysis of gsm calls data for understanding user mobility behavior. In *2013 IEEE International Conference on Big Data*, pages 550–555. IEEE, 2013.

[115] Kevin S Kung, Kael Greco, Stanislav Sobolevsky, and Carlo Ratti. Exploring universal patterns in human home-work commuting from mobile phone data. *PLoS One*, 9(6):e96180, June 2014.

[116] Santi Phithakkitnukoon, Titipat Sukhvibul, Merkebe Demissie, Zbigniew Smoreda, Juggapong Natwichai, and Carlos Bento. Inferring social influence in transport mode choice using mobile phone data. *EPJ Data Science*, 6(1):11, June 2017.

[117] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[118] Takahiro Yabe, Yoshihide Sekimoto, Akihito Sudo, and Kota Tsubouchi. Predicting delay of commuting activities following frequently occurring disasters using location data from smartphones. *Journal of Disaster Research*, 12(2):287–295, 2017.

[119] Muhammad Fahim and Alberto Sillitti. Anomaly detection, analysis and prediction techniques in IoT environment: A systematic literature review. *IEEE Access*, 7:81664–81681, 2019.

[120] Douglas M Hawkins. *Identification of Outliers*. Springer, Dordrecht, 1980.

[121] Ayan Chatterjee and Bestoun S Ahmed. IoT anomaly detection methods and applications: A survey. *Internet of Things*, 19:100568, August 2022.

[122] Marek Novák, František Jakab, and Luis Lain. Anomaly detection in user daily patterns in smart-home environment. *Journal of Selected Areas in Health Informatics*, 3(6):1–11, 2013.

[123] Di Wu, Zhongkai Jiang, Xiaofeng Xie, Xuetao Wei, Weiren Yu, and Renfa Li. LSTM learning with bayesian and gaussian processing for anomaly detection in industrial IoT. *IEEE Transactions on Industrial Informatics*, 16(8):5244–5253, August 2020.

[124] Mahsa Mozaffari and Yasin Yilmaz. Online anomaly detection in multivariate settings. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, October 2019.

[125] John S Eberhardt III, Todd A Radano, and Benjamin E Peterson. Application of machine learned bayesian networks to detection of anomalies in complex systems, May 24 2016. US Patent 9,349,103.

[126] Dharanipragada Janakiram, AVUP Kumar, et al. Outlier detection in wireless sensor networks using bayesian belief networks. In *2006 1st International conference on communication systems software & middleware*, pages 1–6. IEEE, 2006.

[127] Murray Steven, McBean Edward, and Ghazali Mirnader. Real-Time water quality assessment with bayesian belief networks. *Water Distribution Systems Analysis 2010*, pages 388–393, April 2012.

[128] Jingyu Han, Guangpeng Sun, Xinhai Song, Jing Zhao, Jin Zhang, and Yi Mao. Detecting ECG abnormalities using an ensemble framework enhanced by bayesian belief network. *Biomedical Signal Processing and Control*, 72:103320, February 2022.

[129] Milos Hauskrecht, Michal Valko, Branislav Kveton, Shyam Visweswaran, and Gregory F Cooper. Evidence-based anomaly detection in clinical domains. *AMIA Annual Symposium Proceedings*, 2007:319–323, October 2007.

[130] Xueying Lu, Wenyan Xie, Jiankun Hu, and Haiji Wang. Anomaly recognition method for massive data of power internet of things based on bayesian belief network. In *2023 2nd International Conference on Big Data, Information and Computer Network (BDICN)*, pages 17–20, January 2023.

[131] Loss avoidance assessment hurricane irma. Technical Report DR-4337, Florida Division of Emergency Management, 2018.

[132] Kelli Kennedy and Adriana Gomez Licon. Officials: Returning keys residents must be self-sustaining. https://apnews.com/article/836e1c09db724409a8e8a18ccebfe35f, September 2017. Accessed: 2023-6-30.

[133] Chao Chen, Xiao Lin, and Gabriel Terejanu. An approximate bayesian long short-term memory algorithm for outlier detection. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 201–206. IEEE, 2018.

[134] Southeast florida origin destination travel survey. Technical report, Miami-Dade Transportation Planning Organization, September 2017.

[135] Amanda Crawford. Florida keys leaders fear population decrease after irma. https://www.wesh.com/article/florida-keys-leaders-fear-population-decrease-after-irma/12820510#, October 2017. Accessed: 2023-6-1.

[136] Collier county announces new mandatory evacuations. https://www.colliercountyfl.gov/Home/Components/News/News/34636/, September 2017. Accessed: 2023-6-5.

[137] Hurricane irma evacuation directives. https://www.monroecounty-fl.gov/982/Hurricane-Irma, September 2017. Accessed: 2023-6-5.

[138] Reuters. Florida keys, airports partially re-open after irma rips through state. https://www.cnbc.com/2017/09/12/florida-keys-airports-partially-re-open-after-irma-rips-through-state.html, September 2017. Accessed: 2023-6-7.

[139] Devanandham Henry and Jose Emmanuel Ramirez-Marquez. Generic metrics and quantitative approaches for system resilience as a function of time. *Reliability Engineering and System Safety*, 99:114–122, March 2012.

[140] Paul O'Hare, Iain White, and Angela Connelly. Insurance as maladaptation: Resilience and the 'business as usual' paradox. *Environment and Planning C: Government and Policy*, 34(6):1175–1193, September 2016.

[141] Seyedmohsen Hosseini, Kash Barker, and Jose E Ramirez-Marquez. A review of definitions and measures of system resilience. *Reliability Engineering and System Safety*, 145:47–61, January 2016.

[142] Sarah E Anderson, Ryan R Bart, Maureen C Kennedy, Andrew J MacDonald, Max A Moritz, Andrew J Plantinga, Christina L Tague, and Matthew Wibbenmeyer. The dangers of disaster-driven responses to climate change. *Nature Climate Change*, 8(8):651–653, August 2018.

[143] Susan L Cutter, Kevin D Ash, and Christopher T Emrich. The geographies of community disaster resilience. *Global Environmental Change*, 29:65–77, November 2014.

[144] FEMA. National risk index. https://hazards.fema.gov/nri/learn-more. Accessed: 2023-10-1.

[145] CDC/ATSDR social vulnerability index. https://www.atsdr.cdc.gov/placeandhealth/svi/index.html, July 2023. Accessed: 2023-9-17.

[146] Laura A Bakkensen, Cate Fox-Lent, Laura K Read, and Igor Linkov. Validating resilience and vulnerability indices in the context of natural disasters. *Risk Analysis*, 37(5):982–1004, May 2017.

[147] Samuel Rufat, Eric Tate, Christopher T Emrich, and Federico Antolini. How valid are social vulnerability models? *Annals of the American Association of Geographers*, 109(4):1131–1153, July 2019.

[148] Seth E Spielman, Joseph Tuccillo, David C Folch, Amy Schweikert, Rebecca Davies, Nathan Wood, and Eric Tate. Evaluating social vulnerability indicators: criteria and their application to the social vulnerability index. *Natural Hazards*, 100(1):417–436, January 2020.

[149] Michel Bruneau, Stephanie E Chang, Ronald T Eguchi, George C Lee, Thomas D O'Rourke, Andrei M Reinhorn, Masanobu Shinozuka, Kathleen Tierney, William A Wallace, and Detlof Von Winterfeldt. A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake Spectra*, 19(4):733–752, 2003.

[150] Alexander A Ganin, Maksim Kitsak, Dayton Marchese, Jeffrey M Keisler, Thomas Seager, and Igor Linkov. Resilience and efficiency in transportation networks. *Science Advances*, 3(12):e1701079, December 2017.

[151] Jeffrey T Kullgren, Catherine G McLaughlin, Nandita Mitra, and Katrina Armstrong. Nonfinancial barriers and access to care for U.S. adults. *Health Services Research*, 47(1 Pt 2):462–485, February 2012.

[152] Sanju Maharjan, Nebiyou Tilahun, and Alireza Ermagun. Spatial equity of modal access gap to multiple destination types across chicago. *Journal of Transport Geography*, 104:103437, October 2022.

[153] Karel Martens. Justice in transport as justice in accessibility: applying walzer's 'spheres of justice' to the transport sector. *Transportation*, 39(6):1035–1053, November 2012.

[154] Michael J Widener. Spatial access to food: Retiring the food desert metaphor. *Physiology & Behavior*, 193(Pt B):257–260, September 2018.

[155] Tim G Williams, Tom M Logan, Connie T Zuo, Kevin D Liberman, and Seth D Guikema. Parks and safety: a comparative study of green space access and inequity in five US cities. *Landscape and Urban Planning*, 201:103841, September 2020.

[156] Tierra S Bills and Joan L Walker. Looking beyond the mean for equity analysis: Examining distributional impacts of transportation improvements. *Transport Policy*, 54:61–69, February 2017.

[157] Tessa Swanson and Seth Guikema. Using mobile phone data to evaluate access to essential services following natural hazards. *Risk Analysis*, July 2023.

[158] Malcolm Morgan, Marcus Young, Robin Lovelace, and Layik Hama. Opentripplanner for r. *Journal of Open Source Software*, 4(44):1926, 2019.

[159] Geoff Boeing. OSMnx: A python package to work with graph-theoretic OpenStreetMap street networks. *Journal Open Source Software*, 2(12):215, April 2017.

[160] Florida transit data exchange - FTIS: Florida transit information system. https://ftis.org/Posts.aspx. Accessed: 2023-9-15.

[161] Transitland . https://www.transit.land/terms, 2023.

[162] OpenMobilityData. https://transitfeeds.com/, 2023.

[163] FDOT open data hub. https://gis-fdot.opendata.arcgis.com/datasets/7ad4473a939d473ab8c17f83155aedbe/explore. Accessed: 2023-10-1.

[164] Florida division of emergency management. https://www.floridadisaster.org/. Accessed: 2023-10-1.

[165] Treasure Coast Newspapers. Map: Hurricane irma power outages. https://data.tcpalm.com/storm-power-outages/. Accessed: 2023-10-1.

[166] Hurricane irma. https://www.fcc.gov/irma. Accessed: 2023-9-17.

[167] G Brooke Anderson, Andrea Schumacher, Seth Guikema, Joshua M Ferreri, and Steven Quiring. [no title]. https://cran.r-hub.io/web/packages/stormwindmodel/vignettes/Details.html, October 2018. Accessed: 2023-9-19.

[168] Ben Nelson-Mercer. High-Resolution computational modeling of hurricane irma to develop building damage functions based on Location-Based data. 3rd International Workshop on Waves, Storm Surges, and Coastal Hazards, October 2023.

[169] Joern Birkmann. Risk and vulnerability indicators at different scales: Applicability, usefulness and policy implications. *Environmental Hazards*, 7(1):20–31, January 2007.

[170] Jeanette Eckert and Sujata Shetty. Food systems, planning and quantifying access: Using GIS to plan for food retail. *Applied Geography*, 31(4):1216–1223, October 2011.

[171] T M Logan, M H Hobbs, L C Conrow, N L Reid, R A Young, and M J Anderson. The x-minute city: Measuring the 10, 15, 20-minute city and an evaluation of its use for sustainable urban design. *Cities*, 131:103924, December 2022.

[172] Alexander Fekete. Spatial disaster vulnerability and risk assessments: challenges in their quality and acceptance. *Natural Hazards*, 61(3):1161–1178, April 2012.

[173] Benjamin Beccari. A comparative analysis of disaster risk, vulnerability and resilience composite indicators. *PLoS Currents*, 8, March 2016.

[174] Zachary T Goodman, Caitlin A Stamatis, Justin Stoler, Christopher T Emrich, and Maria M Llabre. Methodological challenges to confirmatory latent variable models of social vulnerability. *Natural Hazards*, 106(3):2731–2749, February 2021.