Examining Travel to Non-work Destinations: Integrating Geosocial Media and Smartphone-based GPS Traces

by

Zhiqiu Jiang

A dissertation submitted to University of Virginia in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Constructed Environment

Department of Urban and Environmental Planning
University of Virginia School of Architecture
Summer 2021

ABSTRACT

Urban commercial districts and centers are places that provide concentrated opportunities for non-work activities. Rapid development in these areas has made them critical for local economic development as well as exerting significant influence on urban society and culture. Traveling to these non-work destinations, such as shopping centers, restaurants, bars, grocery stores, movie theaters, etc., is an important part of urban life. For a long time, survey-based data is often used to examine non-work trips and travel patterns. These data always have limited sample sizes that impede temporally and spatially fine-grained analysis. Recent advances in information and communication technology (ICT) and mobile devices create new opportunities for today's transportation planners to understand travel behavior using non-survey sources of data. These data are user-generated, geo-located, and contain contextual information (e.g., text, images, videos). The emergence of such "transportation big data" has resulted in a large quantity of information documenting people's everyday movements, travel events, attitudes, perceptions, and emotions, all connected with the location and time.

This dissertation develops a data fusion framework that integrates geosocial media, fine-grained individual GPS trace data, land use and built environment data, and demographic data from the U.S. census to quantify people's travel experiences and mobility patterns to commercial and mixed-use districts, taking the Phoenix Metropolitan Area as a study case. Specifically, the geosocial media data used in this dissertation is collected from Yelp reviews and the GPS trajectory data is collected from smartphone apps with GPS-enabled location services. This dissertation research first examines the experience of travel (travel attitude) in major commercial and mixed-use districts using transportation texts embedded in Yelp reviews. Then, it analyzes travel behavior to these destinations using GPS trajectory data with a fine scale in space and time. Following on from the prior two analyses, it develops a data fusion framework by integrating geosocial media and GPS traces to further examine 1) the relationship between attitude and built environment, and 2) the impacts of attitude and built environment on travel behavior.

Given the prospect of the big data era for transportation research, this dissertation research shows the promises of emerging data and analytics in providing useful information about travelers' attitudes and behaviors. It also enhances our understanding of non-work travel and has implications for transportation planning and management. Therefore, this dissertation makes two major contributions to urban transportation planning research, one regarding the travel to non-work destinations, and second regarding the methods developed to integrate multiple types of big data for transportation planning informatics.

Keywords: Geosocial Media, GPS Data, Transportation Planning Informatics, Non-work Travel

Examining Travel to Non-work Destinations:
Integrating Geosocial Media and Smartphone-based GPS Traces

A Dissertation Presented

by

Zhiqiu Jiang

Approved as to style and content by:

____
Andrew Mondschein, Chair

____
Ellen Bassett, Member

____
Yangfeng Ji, Member

____
Guoping Huang, Member

____
Natasha Foutz, Member

ACKNOWLEDGMENTS

This dissertation would not have been possible without the support and assistance of many people.

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Andrew Mondschein, for his advice and support with full encouragement and enthusiasm in the past five years. His expertise was invaluable in formulating the research questions and methodology. Without his constant guidance and insightful feedback this dissertation would not have come to be. I greatly appreciate his sincere mentorship in research, scholarship, and teaching during my PhD study.

I would also like to thank my committee members. Thank you to Dr. Ellen Bassett for participating in guiding, kind support, and encouragement through my doctoral process and helping me to sharpen my thinking and bringing my work to a higher level. To Dr. Yangfeng Ji for contributing constructive feedback and advice and supporting my work wholeheartedly. To Dr. Guoping Huang for constantly supporting and providing valuable guidance throughout my studies. To Dr. Natasha Foutz for encouraging me to think beyond the limits of my field.

I greatly acknowledge all of my professors at UVA. Specifically, I would like to thank Dr. Nana Last and Dr. Shiqiao Li for their guidance of my dissertation research in colloquiums that broadened my horizons with different perspectives and encouraged me to conduct interdisciplinary research.

Thank you to all my research collaborators. I am very grateful that I worked with academic scholars from diverse departments and institutions during my doctoral time. Especially to Max Zheng (UVA Department of Computer Science) for his participation in social media research.

I also want to thank the UVA Scholars' Lab. Especially to Chris Gist and Drew MacQueen for their treasured support and kind help that deepened my knowledge in spatial analytics. To Brandon Walsh and Shane Lin for expanding my text mining skills to digital humanities.

I am more than lucky to know my peers at the Constructed Environment PhD Program. I would like to thank you for all the shares of ideas, inspiring discussions, and spiritual encouragement: all of you collectively created a supportive and positive research environment. My sincere thanks also go out to visiting scholars. Memories of our work time at the PhD office were valuable to me.

In addition, I gratefully acknowledge the funding received towards my doctoral study from School of Data Science, Scholars' Lab, and Center for Undergraduate Excellence Double Hoos

at University of Virginia - all of which supported me to learn analytical methods and conduct data science research.

I am blessed with such a supportive family that has always been there for me. My grandparents Yuzhong and Jiying have sustained me throughout my journey with endless love. I am extremely grateful for my mom and my dad - Jie and Mingzheng - for their support and encouragement in every possible way. Without them I would never have enjoyed so many opportunities. I also want to thank my relatives and friends who have helped and supported me.

And to Xiao, my husband, best friend, and life partner who has been always supportive in every way. Your contribution was more than moral support. I am grateful to your engagement in research discussions, share of ideas, and wise feedback and your listening to my random and whimsical thoughts about research, life, and anything else. Your understanding, patience, company, humor, and love mean everything. I am blessed to have you along.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure | Page

CHAPTER 1

INTRODUCTION

In the United States, it is estimated that 83% of the population lives in urban areas in 2019 (World Bank, 2020). By 2050, 89% of the U.S. population is projected to live in urban areas (United Nations, 2018). Urban planning and design, as applied disciplines, are permeated with best practices that are seeking to promote sustainability, equity, and justice for all throughout the whole city system. Transportation, as an essential facet of people's daily lives, facilitates the participation in daily activities and opportunities, e.g., work, recreation, shopping, dining, and businesses. One fundamental purpose of transportation systems in urban areas is to make infrastructure, resources, and urban amenities more accessible to individuals (Carmon & Fainstein, 2013; Tumlin, 2012). A lack of individual mobility or access to urban opportunities will hinder personal economic and social development (Woldeamanuel, 2016). In this sense, urban planners seek to better understand people's travel needs and measure their accessibility to urban opportunities to ensure that they can approach to various destinations and resources as desired.

With the rapid development of information and communication technologies (ICT), the emergence of novel data and methodologies provides tremendous opportunities for transportation planners and researchers to examine everyday activity-travel experience, travel behavior, and transportation accessibility (Rashidi et al., 2017). Transportation big data[1] has the five V's characteristics of big data: volume, velocity, variety, veracity, and value (M. Chen et al., 2014; Tsai et al., 2015). Recent studies have shown a promising future for making use

[1] A detailed description of transportation big data can be found in Chapter 2.

of a range of multi-source geo-spatial big data in transportation planning informatics and urban data science (Manovich, 2012; X. Wu et al., 2014; C. Zhou et al., 2017). Rich semantics of space and place associated with activity-travel are embedded in the large-scale “user-generated data”, such as geotagged social media posts, online reviews, mobile phone GPS data, crowdsourced traffic and street data, and so forth.

Among these new data sources, social media data, for example, with the platforms’ widespread use, has become a popular data source in more recent transportation studies. Specifically, with the assistance of big data analytics and data mining techniques, researchers can use social media data to extract a variety of information about travel, including the public opinions about transportation environment and built infrastructure, responses and attitudes towards new transportation technologies, such as electric vehicles and ride sharing services (e.g., bike-sharing and car-sharing), and travel experiences during the trip. (Efthymiou & Antoniou, 2012; Rahim Taleqani et al., 2019; Sabab Zulfiker et al., 2020).

Moreover, crowdsourced GPS trajectory data is another emerging transportation big data source. People’s movements can be retrieved via mobile locations recorded from a variety of GPS-enabled apps. Traditionally, planners collect travel behavior data based on conventional survey methods, such as travel diaries, questionnaires, or interviews. For a long time, these methods have been the primary means to obtain travel behavior information, e.g., mode of transportation, distance traveled, frequency of trips, travel purpose, trip duration (Bohte & Maat, 2009; C. Chen et al., 2016). These survey-based methods, however, impose a significant burden on participants, as they have to remember their travel behavior, recognize their trip attributes and report them often without any supplementary support. In addition, these traditional methods always requires longer data collection time and usually covers a smaller sample size (T. Jones et al., 2013). Thus, a more efficient and accurate travel behavior data collection means is needed. Coming with the wide use of mobile phones, such mobile phone GPS traces have provided researchers with alternatives to improve data quality, by capturing more details of individuals' travel behaviors to give insight into fine-grained travel patterns (Gonzalez et al., 2010; P. Stopher et al., 2008).

In this dissertation, I seek to utilize and integrate large-scale user-generated data and analytical methods to examine travel to non-work destinations. My research includes an investigation of the experience of travelers (attitude), travel behaviors and accessibility, and the relationship between travel attitude, behaviors and the built environment. One major challenge in travel studies lies in the lack of sufficient data about non-commuting trips. Thus, new transportation datasets have the potential to fill this gap. I process large-scale transportation big data and effectively extract useful information about the trip characteristics, in order to collectively better understand the interactions between transportation systems, built environment, and human factors (travel behavior, attitudes, and perceptions), which serves as the theoretical framework of the dissertation. In particular, taking the Phoenix Metropolitan Area, Arizona as a case study, I focus on fusing geosocial media and smartphone-based GPS trajectory data to examine individuals' travel to non-work destinations integrated with textual analysis and statistical methods in data processing and analysis.

Through a series of empirical analyses of big datasets collected for the study area, I explore both the potential and limitations of geosocial media and mobile phone GPS trajectory data in travel information retrieval and behavioral studies. The dissertation analysis progress is designed as follows: 1) The first analysis makes use of geosocial media data - Yelp reviews in Phoenix - to examine how the public perceives the parking environment of their frequently visited commercial areas based on a textual analysis of the extracted parking reviews. Since each Yelp review is linked with a real-world travel destination, the parking experience is embedded in the reviews to represent how ease to find a place to park when driving to the destination, as an indicator of travel attitude; 2) The second analysis investigates trips to six major commercial districts using smartphone-based GPS data collected for one month with an average of 20 million trajectory points daily in Phoenix. Each participant's home location (at census block groups level), time of travel, and mode of travel are estimated; and 3) The third analysis continues an exploration of how travel attitudes, the characteristics of where people live, and other built environment characteristics affect destination choices to commercial districts by integrating the results from the first two analyses.

Overall, these analyses support a framework of using large-scale user-generated datasets in transportation planning research, highlighting in particular that user-generated content in geosocial media and human mobility GPS data, and, thus, contributing to the field of transportation planning informatics. I find that the common characteristics for these emerging data sources are their large size (up to terabytes of data), incomplete demographics, explicit or implicit geographic information, and rapid and potential real-time updating. I employ a variety of analytical methods and show that these new data resources do provide unprecedented opportunities to advance transportation planners' capabilities in the analysis of travel experience, detection of emotions and attitudes towards transportation supplies, recognition of travel patterns, examination of transportation accessibility, and prediction for future behaviors. It is worth noting that these data are often burdened with questions of representativeness and ethical use that must be carefully addressed, particularly when used for public purposes. Therefore, there is a pressing need for further research on the specific contexts and prerequisites that these data require so that they can be appropriately used in human mobility and transportation planning research. In future, a number of multi-source data fusion research practices will call for attention in establishing a robust field of transportation planning informatics.

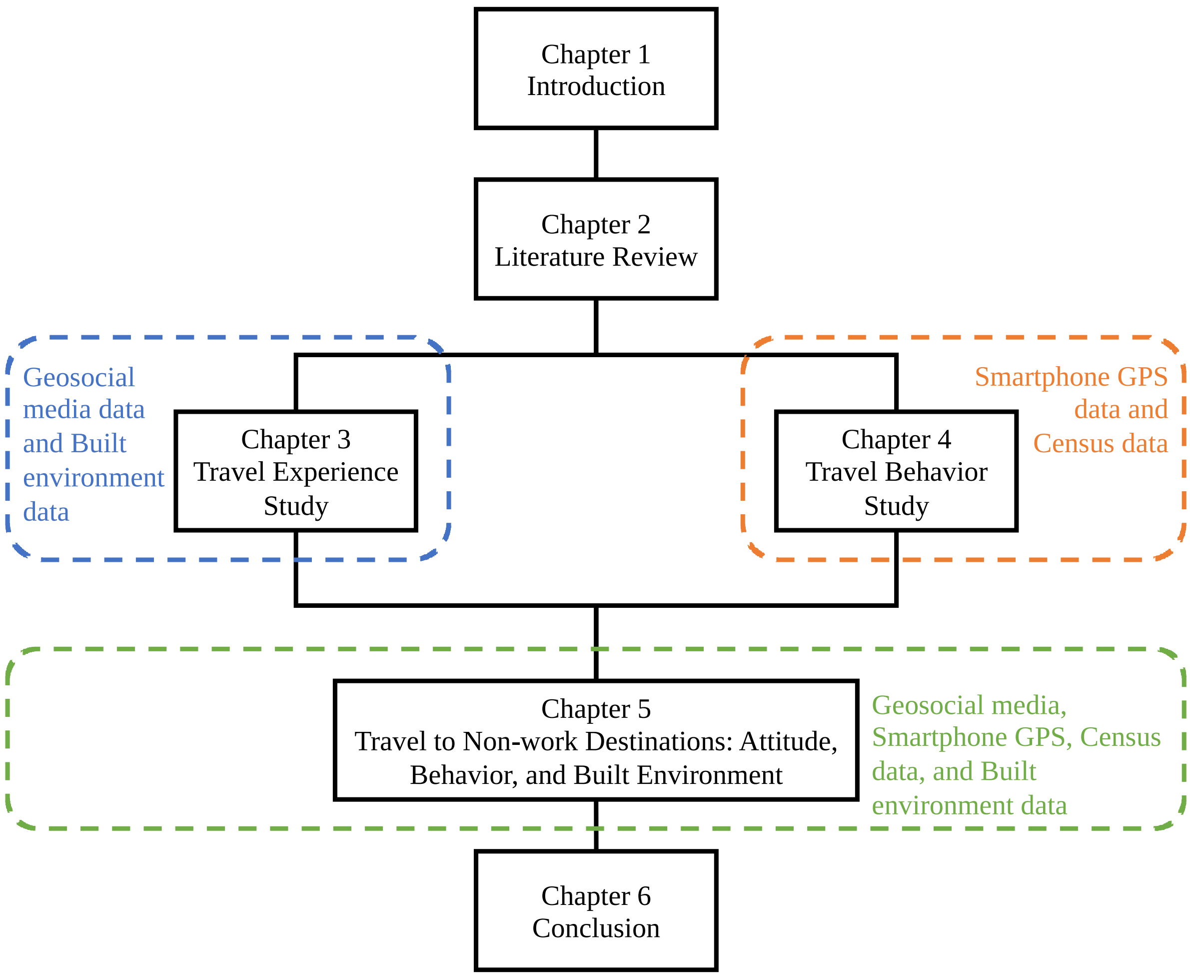

Figure 1. Dissertation organization

This introductory chapter provides an overview of some of the key components of the dissertation. It addresses the research significance of studying the travel to non-work destinations, motivations in using new transportation data and data descriptions, as well as a conceptual framework of the dissertation. As Figure 1 shows, following Chapter 1, I conduct a literature review of current studies, particularly focusing on the need of expanding the literature on travel to non-work destinations and the need of using new data and analytics. Then I detail my study by first separating an analysis of geosocial media data and GPS data, and then combining these two. Specifically, in Chapter 3, I focus on using Yelp to analyze its embedded transportation content and parking sentiments in some of the major commercial and mixed-use districts. In Chapter 4, I focus on deriving the knowledge of non-work trips and travel patterns from smartphone-based GPS traces. In Chapter 5, I combine geosocial media, GPS data, census data, and built environment data to examine 1) the relationship between attitudes and built environment, and 2) the impacts of attitude and built environment on travel behavior. The final chapter summarizes the contributions of the dissertation and discusses the implications for transportation planning and urban data science.

1.1 The Access to Non-work Destinations

Traveling to non-work destinations, such as shopping centers, restaurants, bars, supermarkets, grocery stores, movie theaters etc., is an important part of urban life. The following two subsections describe the research background and motivations in using new data for examining the access to non-work destinations.

1.1.1 Research Background

Urban commercial districts and centers provide concentrated opportunities for shopping, dining, leisure, and other services. Rapid development in these areas has made them critical for local economic development as well as exerting significant influence on urban society and culture (Adams, 2012). The commercial districts in the city are, for a resident, among the most important vital elements in the urban environment. The elements of the urban environment are integrated around and across them to create a place filled with activities (K. G. Jones & Simmons, 1993). Constructing a diverse commercial environment, attracting more customers, and keeping these districts and centers vibrant are significant for urban economic development.

(1) Divergent Travel Behavior and Mobility Patterns to Non-Work Destinations

Individuals go for such non-work activities and exhibit divergent travel behavioral patterns when going to these places. Previous studies demonstrate that residents of disadvantaged neighborhoods don’t have the same accessibility to such destinations as the residents of advantaged neighborhoods (Stanley et al., 2011; Wee et al., 2011). Researchers attribute this inequality to a variety of factors: for example, although residents of disadvantaged neighborhoods may travel wherever they want, they may have different destination preferences and may be constrained by available resources they have (e.g., money, time, auto access) (Neutens et al., 2010; Ren et al., 2014). This inequitable accessibility also reveals that urban isolation and segregation extend beyond individuals’ residences. Understanding individuals’ mobility patterns to non-work destinations is crucial for urban planning, land and facility management, and business strategies (Crane, 2000).

(2) Factors Affecting the Choice to Non-work Destinations

As mentioned above, a variety of factors might contribute to the such divergent behavior patterns. First of all, the choice to go to a non-work destination for urban activity is a type of travel behavior. Scholars have begun to explore the potential effects of *travel attitude* on traveler behavioral choices and thereby cities, society, and the environment. Broadly, the perceptions, attitudes, and experiences of travel are associated with travelers’ behaviors.

Besides the impacts of travel attitudes and perceptions, existing literature suggests that transportation and infrastructure plans are largely associated with behavioral choices. The transportation road network, infrastructure, and transportation systems, serve communities and neighborhoods with inherent imbalance (Q. Wang et al., 2018). The imbalanced distribution of transportation infrastructure can exacerbate existing inequitable accessibility and mobility patterns of different communities (Graham, 2002, 2010; J. Lin & Mele, 2013). With noting this relationship between transportation infrastructure and urban accessibility, local land use and transportation plans often seek to provide a balanced design which can bring business benefits in urban non-work activity districts as well as mitigate accessibility inequality.

(3) Accessibility as A Measure for Travel Patterns

Following the observations of divergent non-work travel patterns, planners establish some methods to measure these travel behavioral differences by introducing the "accessibility" concept. Basically, accessibility is used as a measure of the ease of travel when reaching a destination, which generally represents a relative level. For a long time, the lack of fine-grained data impedes this measurement. Traditionally, data for travel behavior analysis has been derived from surveys with limited sample sizes that impede temporally and spatially fine-grained analysis (C. Chen et al., 2016). Small sample survey data often cannot supply enough information (e.g., time of travel) to validly assess local patterns, or to do so in a way that obscures the behavior of specific individuals. Recent advances in information and communication technologies have enabled researchers to collect mobility data based on ubiquitous and location-aware smartphones, with massive GPS space-time data at a fine scale (Dabiri & Heaslip, 2018).

1.1.2 Motivations in Using New Data

The concept of *Smart Cities* describes a more connected and intelligent constructed environment for embedded devices, such as smartphones, wearable devices, and sensors, leading to new ways of interactions between people and cities (Monzon, 2015). The development of digital technologies allows us to interconnect and communicate with each other more effectively and conveniently. The city system implementation begins with data generation and move to collection, aggregation, filtration, classification, preprocessing, computing and decision making (Rathore et al., 2018). In such contexts, the explosion of new transportation data sources in recent years has aroused increasing attentions for examining travelers' behaviors and their experiences of everyday activity-travel in cities.

To examine travel to non-work destinations, I identify two major new data sources that offer the opportunity to obtain useful information: geosocial media data and human GPS trajectories. I also include a description of other useful data sources such as census data and land use and built environment data in the following section.

1.2 Data Description

1.2.1 Travel Experience from LBSN Data - the Geosocial Media Data

The rapid growth of location-based social networks (LBSNs) has attracted billions of users, elevating our urban experience to a new stage. In this new context, rapid developments in the field of spatial data mining and geo-simulation provide highly valuable and promising tools that enhance our understanding of how cities function and how people interact with urban environments at a fine-grained space-time scale.

In recent years, transportation researchers have experienced the benefits of widespread geotagged social media platforms in capturing experience of travel, especially in the large metropolitan areas. A variety of LBSNs (Yelp, Twitter, TripAdvisor, Foursquare, Facebook, etc.) have been used to study travel information from a review of the literature, suggesting that the promises of using LBSNs for transportation information retrieval, analysis, and applications.

Yelp reviews, a typical LBSN dataset, are attracting more and more people to write "tips" and "reviews" when they have visited a destination. Using textual analysis, I find that a significant number of these reviews mention travel experiences. Since each review points to a real-world travel destination, I further develop the idea of using such relevant texts from Yelp reviews to understand how the public perceives destinations' transportation environment. The exploration and empirical analyses are conducted in Chapter 3 with my advisor Dr. Andrew Mondschein. We first extract transportation content from Yelp reviews, and then analyze individuals' sentiments towards parking[2].

1.2.2 Human Mobility Information from GPS Trajectory Data

Inferring an individual's activity and trip purposes is critical for transportation planning and travel behavior analysis. Detailed trip information including time, trajectories, and origin and destination locations reveal objectives behind these trips and preferences of the commuters. However, using the paper travel diaries or phone call surveys to record the trip information have apparent drawbacks such as the difficulties faced by respondents to exactly recall the trip-related information including departure time, origins and destinations, etc., and the non-response and postponing (Bohte & Maat, 2009).

---

[2] Chapter 3 describes the reason why parking sentiment is selected as an indicator of travel attitude in this study in detail.

One major challenge lies in the lack of high-quality data to examine non-work travel. In recent years, the knowledgeable findings and data mining from transportation big data have become more and more popular. Crowdsourced phone-based GPS data, such as mobile phone GPS traces, can provide fine-grained insight into travel behavior patterns to fill the data gap. Thus, as part of this dissertation research, I use phone-based GPS data collected for one month with an average of 20 million trajectory points daily to examine people's non-work travel behaviors. Data from over 90,000 individuals in Phoenix metropolitan area are used for the empirical analysis. I first study the problem of how to process large trajectory data to extract useful information about individual trip characteristics. In Chapter 4, I develop an analytical framework to detect the home location, time of travel, and travel mode of each respondent when going to commercial districts. I explore and quantify individuals' travel to six major commercial districts with a specific focus on the mode use by cars, which is the main mode that used by Phoenix individuals or households for non-work travel. According to the travel behavior information obtained from the GPS data, I also can measure the non-work travel accessibility and compare the travel time burdens that are suffered by residents from vulnerable neighborhoods.

1.2.3 Census Data

The census data - American Community Survey - was acquired online from the US Census Bureau (2018). Census data covers all the census block groups in Phoenix Metropolitan Area. It can give a comprehensive list of demographic and social-economic characteristics. By using census official Application Programming Interface (API), I also obtain the geographical boundaries of the study area. The API is a protocol that US Census Bureau makes available to allow the public to send and request information from them. The spatial information is useful for spatial analysis.

1.2.4 Land Use and Built Environment Data

The land use and built environment data was acquired via US EPA (Environmental Protection Agency) smart location database (SLD). The information in the SLD database is available nationwide in the form of ArcGIS online mapping datasets. I use the SLD data at the census block group level, a geographical unit used by the US Census Bureau which may have a population of 600-3,000 people (U.S. Census Bureau, 2014).

1.3 Conceptual Framework and Empirical Analyses

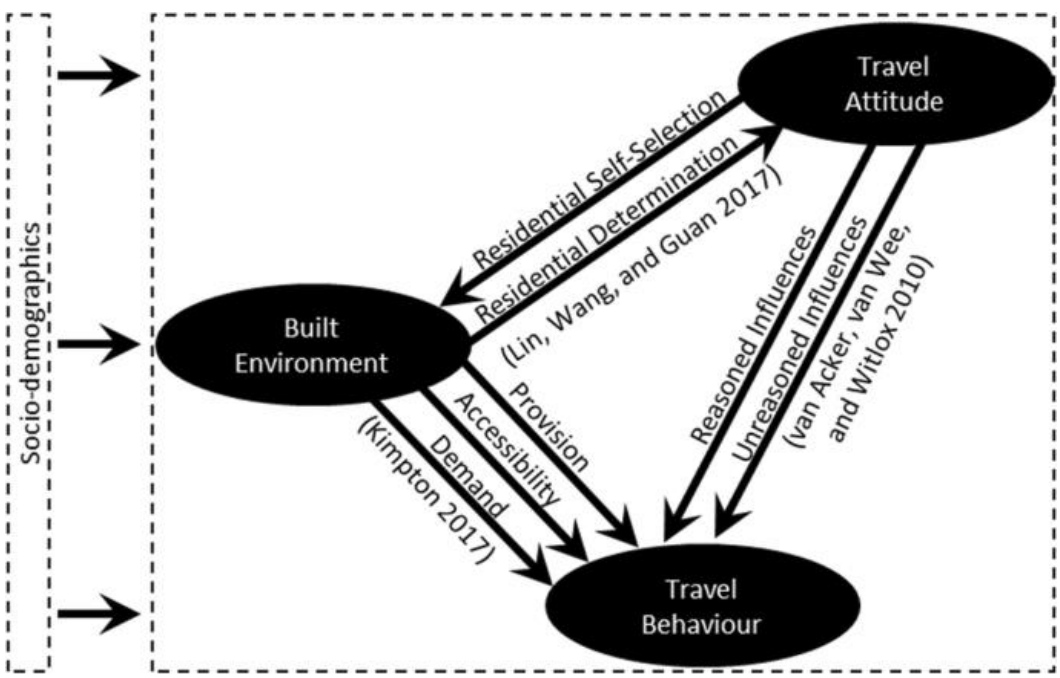

Figure 2. A conceptual framework of travel behavior

source: Kimpton, 2020

I utilized the conceptual framework (see Figure 2) that described the relationship between attitude, behavior, and built environment to design this study (Kimpton, 2017, 2020;

T. Lin et al., 2017; Van Acker et al., 2010). As shown in Figure 2, existing literature majorly focused on *residential self-selection* based on this conceptual framework. For example, the theoretical framework in Lin et al. (2017) emphasized that these influences are directional whereby *residential determination* occurs when the built environment influences travel attitude and behavior, and *residential self-selection* occurs when travel attitude influences the built environment and travel behavior.

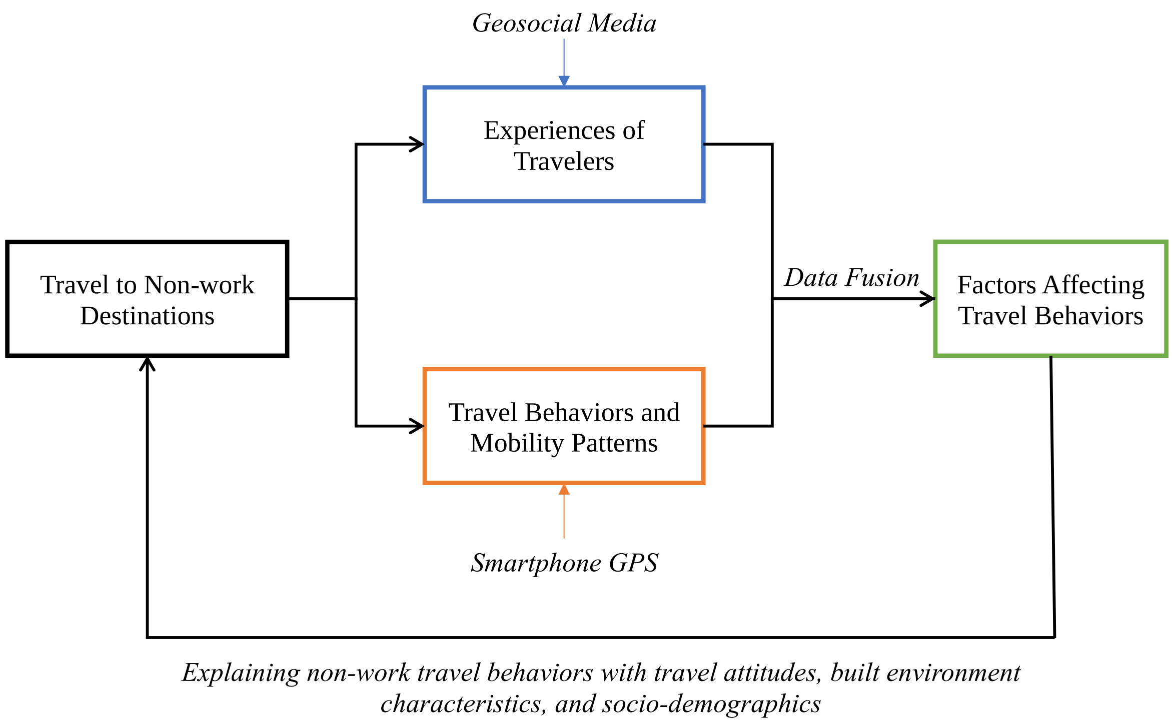

Figure 3. Data fusion framework

This dissertation's multi-source transportation big data fusion framework is shown in Figure 3. Through a series of empirical analyses of the novel datasets of the study area, I explore both the potential and limitations of geosocial media and mobile phone GPS trajectory data in travel information retrieval and behavioral studies. As Figure 3 shows, I first make use of geosocial media data - Yelp reviews in Phoenix - to examine how the public perceives the parking environment of their frequently visited commercial areas based on a textual analysis of the extracted parking reviews. Since each review is linked with a non-work travel destination, parking experiences embedded in the reviews can represent how ease to find a place to park when driving to the destination, as an indicator of *travel attitude*. Then I investigate non-work travel behaviors and mobility patterns using smartphone-based GPS data. Finally, I further design two data fusion frameworks to examine the intertwined relationship between attitude, built environment characteristics, and non-work destination choices.

1.4 Implications and Contributions

Implications for cities and planning follow from the conceptual framework and empirical findings in this dissertation. In summary, this dissertation makes two major contributions to urban transportation planning research, one regarding the travel to non-work destinations, and second regarding the methods developed to integrate multiple types of big data for transportation planning informatics.

The applications of employing *big data* for transportation analysis are promising and may include trip pattern identification, travel time evaluation, travel demand estimation, etc. The *big data* feature can possibly rule out the noises inherent in the traffic operations and even compensate deficiencies of traditional methods. Thus, this dissertation's contributions to transportation planning research reside primarily in employing multiple types of crowdsourcing data for advanced non-work travel behavior information retrieval and analysis.

At the end of this chapter, I only present a list of main contributions of each chapter. A detailed description of contributions of this dissertation in terms of research approach and planning implications can be found in the Chapter 6.

(1) Identifying Experiences of Travelers with Geosocial Media Data

In Chapter 3, using the features of the Yelp dataset, I extract transportation content embedded in the review texts. Particularly, I analyze the sentiment towards parking by assessing satisfaction or frustration with parking at different types of businesses in six major commercial districts across the region.

(2) Decoding Travel to Non-work Destinations

In Chapter 4, I unveil human mobility patterns based on the smartphone GPS trajectory data. I analyze trips to major commercial and mixed-use districts and compare trip characteristics (e.g., the distance of travel, time of travel) and accessibility of individuals from different types of neighborhoods using precise spatial and temporal information recorded in the phone-based GPS data.

(3) Integration of Geosocial Media and Phone-based GPS Data

In Chapter 5, I first examine associations between parking supply and parking sentiment and find that commercial districts with shared parking supplies have overall more positive parking sentiment. Then I study impacts of attitude and built environment on behavior by integrating geosocial media, GPS, census data, land use and built environment data.

Overall, the findings of this dissertation collectively establish a data integration framework that utilize and integrate large-scale user-generated datasets and analytics methods to understand interactions between transportation systems, built environment, and people. However, despite the quantity of available new data, each type of transportation big data has its specific advantages, limitations, and application scopes for planning informatics. Gaps are often present due to specific limitations of the instruments or their carrier platforms. One data type alone is not likely to be capable of providing multi-dimensional information about travel. Therefore, the investigation throughout this dissertation also demonstrates the potential of an integration approach which can combine multimodal types of new transportation data with different features, structures, resolutions, and precision for understanding, imagining and shaping the future of data-smart transportation planning.

CHAPTER 2

LITERATURE REVIEW

2.1 Need for Research on Travel to Non-work Destinations

Non-work trips, including a spectrum of trip purposes: social, recreational, shop, dining, service, and other non-work activities, is an important part of all trips. As retail and consumer services increased, there are ever more variety and opportunity in urban commercial areas with increasing variety of shopping, eating out, and other non-work activities and consequently induce more trips per capita. According to the most recent nationwide survey of travel, the 2017 National Household Travel Survey (NHTS) data, the growing amount of non-work trips[3] comprises more than 80% of all trips (McGuckin & Fucci, 2018).

Nevertheless, for a long time, travel behavior studies, despite the different data and analytical methods used in case studies, primarily focus on work trips (trips for commuting purposes). Modeling of home-based non-work trips and non-home-based trips has received less attention in the urban travel literature. As a result, a significant amount of research works focusing on commuting behaviors can be found in previous literature. However, the increasing number of the non-work trips and their consequences to traffic congestion have recently caused attention, but there is a limited number of studies examining transport to non-work destinations, and almost all recent empirical work on non-work travel has lacked a clear behavioral framework, which limits both the credibility of the analyses and the persuasiveness of the policy recommendations.

[3] The non-work trips include trips for social, recreational, shop, family, personal, school, and church activities in 2017 NHTS dataset.

Compared with commuting trips, travelers have much more flexibility to choose where they go for non-work activities. Non-work destination choices can be affected by a variety of factors. As the conceptual framework in Chapter 1 shows, attitude, built environment, and socio-demographics may affect the generation of non-work trips. Thus, in the following subsection, I describe the findings about the 1) non-work travel dimensions; and 2) factors affecting non-work trip generation respectively based on a review of previous literature.

2.2.1 Dimensions of Non-work Travel: Patterns and Measurement

To provide a review of papers on non-work travel dimensions, two related bodies of work are reviewed here: literature on non-work travel patterns, and literature on measurement of non-work travel.

(1) Non-work Travel Patterns

Travel patterns for non-work activities, such as meals, shopping, recreation, and socializing, are less routinized than commuting. Existing academic literature examined non-work travel patterns in the context of United States can be traced back to 1990s. At that time, non-work trips were found to be linked into trip chains or tours involving several stops (Nelson & Niles, 1999). Lockwood and Demetsky (1994) conducted a travel diary survey of 118 households in Northern Virginia in 1992[4]. They found that daily non-work trip making vary widely by individuals, by gender, employment, and marital status. As Table 1 shows, the average number of daily non-work trip per person was 2.32, but varied across gender, employment, and marital status. For example, single, unemployed women made four times

[4] The travel diary surveys were mailed to each household and conducted on a weekday in April, 1992 (Lockwood & Demetsky, 1994).

more trips than single, employed men. They also made more trips than married, unemployed women. A possible reason for this could be unemployed single mothers made more trips than unemployed single women without children[5]. Also, the average number of non-work trip made by women was overall higher than that by men, regardless of employment or marital status.

Table 1. Average daily non-work trips per person, by personal characteristics

Personal Characteristics		**Average Trips Per Person, By Gender**	
Employment	**Status Marital Status**	**Male**	**Female**
Employed	Married	1.71	2.96
	Single	0.96	2.55
Unemployed	Married	2.00	3.83
	Single	2.29	4.00
All persons		2.32	

Source: data is from Lockwood & Demetsky, 1994.

Additionally, Lockwood and Demetsky (1994) summarized the average daily number of non-work trips made by various household characteristics. The average number of non-work trips per household per day was 5.03, but varied upon neighborhood types (urban / suburban), income levels, and household structure (have children / no children). Suburban and lower income households averaged more non-work trips than urban and higher income households. The average trip rate appeared to increase if the household had children. Based on these collected data samples, their results suggested that a large portion of non-work travel in suburban communities was for children activities.

[5] A supplemental explanation for this can be found in the following paragraph.

For mode choices, Polzin et al. (1999) investigated differences in the choice of transportation mode for non-work activities by race and ethnicity using the 1995 Nationwide Personal Transportation Survey (NPTS) dataset from the US Department of Transportation (USDOT, 1995), which consists of non-work travel aggregated at the national level. They found that the dominant mode for all groups conducting non-work trips is the use of a privately-owned vehicle, either as the driver or as a passenger. African American had lower percentage of using cars for non-work travel than Whites. The authors believed that these differences were largely from differences in the rate of vehicle ownership and driver's license (Polzin et al., 1999).

Later on, Nelson et al. (2001) use the NPTS data (USDOT, 1995), with integrated research methods, to examine non-work trip purposes and travel demand. Table 2 shows the percentage distribution of trip purposes in the U.S. in 1995. "Shopping", "other family and personal business", "other social and recreational", and "eating out" accounted for more than half (54%) of all trips. The destination choice of these non-work trips was reported with larger flexibility than work trips. Furthermore, they investigated the mode of transportation for both trips to work and to other non-work activities. As Figure 4 shows, private vehicle dominates across all trip purposes. For non-work trip purposes, such as "shopping" or "out to eat", the automobile mode accounts for more than 85% trips. Walking is the second mode of choice especially for "out to eat" and "other social / recreation" activities, suggesting that a preference for taking a convenient walking from home to a nearby dining place.

Table 2. Trip purpose as percentage of all person trips

Trip Purpose	**Percentage**	**Destination Flexibility**
Work and Work Related	18	Somewhat inflexible
Shopping	21	Flexible
Other Family and Personal Business[1]	15	Somewhat flexible
Out to Eat	8	Flexible
Other Social / Recreation[2]	10	Flexible
Other	28	Somewhat inflexible

Source for columns 1-2: USDOT, 1995; Source for column 3: Integrated Transport Research (Nelson et al., 2001).
Note: [1] "Other family and personal business" includes the purchase of services such dry cleaning, auto repair, personal care, banking, and legal services.
[2] "Other social and recreational" includes entertainment, recreation, and cultural events.

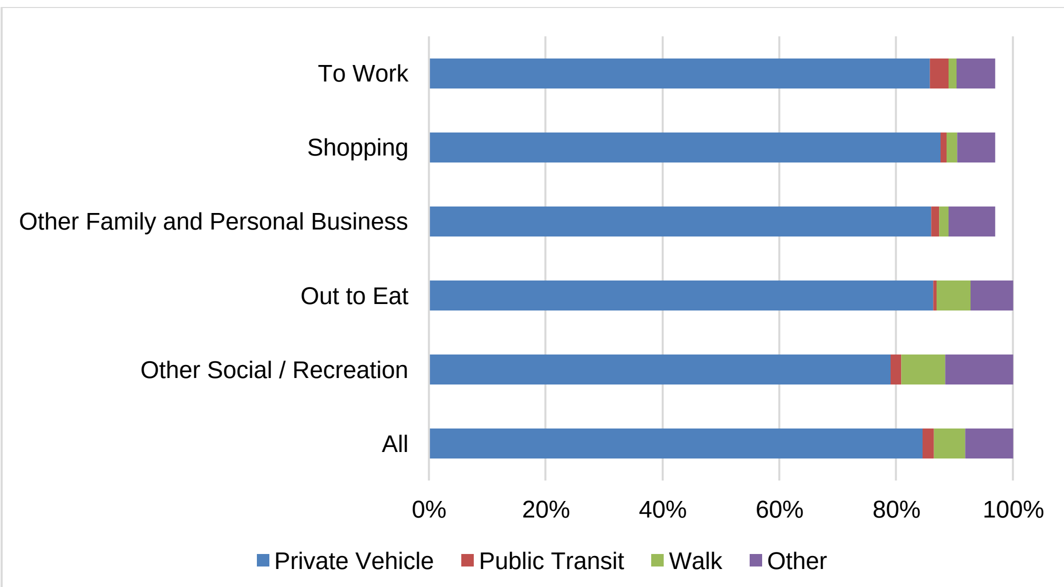

Figure 4. Percentage of daily person trips by mode and selected trip purpose, 1995

Source: data is from table 2-3, p 29 (Nelson et al., 2001).
Note: It does not include 3 percent of all trips for which a mode was not ascertained. "Other family and personal business" category includes the purchase of services such dry cleaning, auto repair, personal care, banking, and legal services; "Other social and recreational" category includes entertainment, recreation, and cultural events; "Public Transit" category includes taxicab; "Other" category includes school bus and bicycle. Data source is from USDOT, 1995.

Since the results presented in Figure 4 were based on the national survey data collected in 1990s from early literature, I analyze the percentage of daily person trips by mode and trip purposes using the most recent nationwide survey of travel - the 2017 National Household Travel Survey (NHTS) data (Federal Highway Administration, 2017). Each survey respondent reported trip purpose (e.g., school, work, out to eat, shopping, recreation, etc.), mode of transportation (car, walking, bicycle, bus, etc.), time of day travel, day of the week, and vehicle ownership and occupancy.

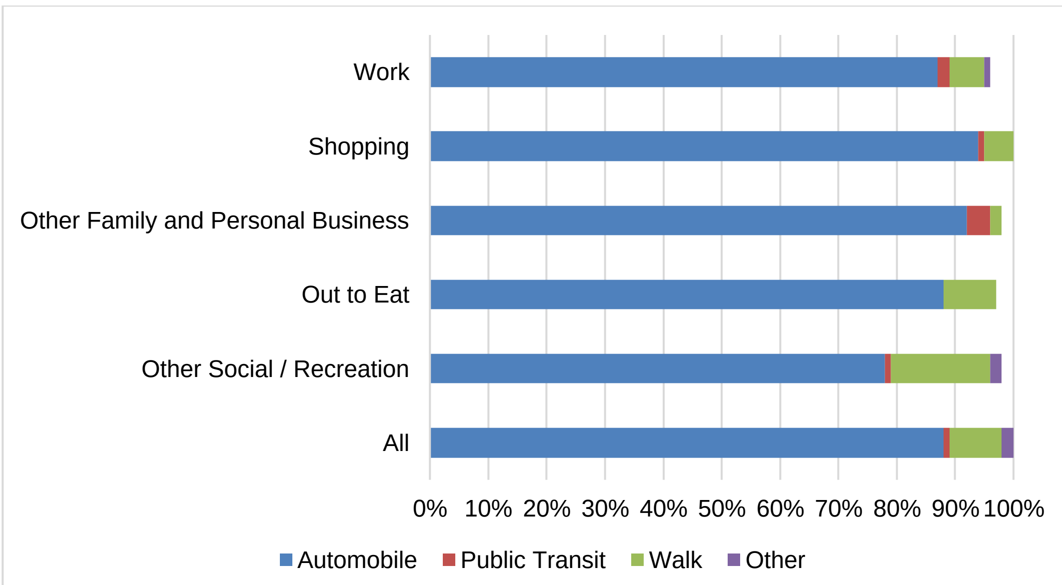

Figure 5. Percentage of daily person trips by mode and selected trip purpose, 2017

Source: data is from the 2017 National Household Travel Survey (FHA, 2017)
Note: “Other family and personal business” category includes the purchase of services such medical and dental services; “Other social and recreational” category includes entertainment, recreation, and cultural events; “Public Transit” category includes taxicab; “Other” category includes school bus and bicycle.

The results are shown in Figure 5. Similar with distribution in Figure 4, the dominant mode of transportation for all trip purposes is still driving. In specific, automobile mode accounts for 94% for “shopping” trips and 88% for “out to eat” trips. Walking is the second mode of choice, accounting for 17% for “other social / recreation” activities and 9% for “out to eat". Overall, an increasing trend is found for the use of the automobile for non-work activities by comparing Figure 4 and Figure 5.

Echoing these earlier studies, some of more recent empirical studies further examined divergent patterns of people going to non-work destinations and their accessibility differences. First of all, disparities were found across demographic and income groups. For example, Dunkley et al. (2004) found that Black predominant tracts of the Atlanta region had lower accessibility than their white counterparts to grocery stores, non-fast food restaurants, and movie theaters. Immergluck and Smith (2005) analyzed the economic investment trends and conducted a case study in Chicago. Their findings showed that an increase in the percentage of black or Hispanic residents resulted in decreases in commercial investment. Scott and Horner (2008) did a case study in Louisville and used a variety of accessibility measures to discover that people from at-risk socioeconomic groups overall (80%) experienced disadvantage in their ability to reach some of the important non-work travel destinations, such as grocery stores, hospitals, and post offices. Grengs (2015) examined transportation accessibility of vulnerable social groups, including African Americans, Hispanics, low-income households, and households in poverty. He found that people in the vulnerable groups experienced a distinct disadvantage in accessibility to shopping and supermarkets in the Detroit metropolitan region.

Generally speaking, cars provide greater access than transit to both work and non-work destinations (Grengs, 2010, 2015), and the use of car is the dominant travel mode for non-work activities for Americans. According to the 2017 NHTS data, most US households have at least one car (Federal Highway Administration, 2017). Cars also offer a more convenient mode for non-work trips, such as shopping and escorting children, which are more easily fulfilled via cars than other modes of transportation. People who have the option of driving, including those who have licensed drivers or who live in household with vehicles, choose car as their modes of non-work travel (Chatman, 2008; Polzin et al., 1999). Studies also show that differences exist in social groups in their ability to reach non-work destinations: behavioral patterns of non-work travel may widely vary across the demographic and socioeconomic groups.

(2) Measures of Non-work Travel Patterns

Measures of non-work travel include a variety of variables, such as travel speed, distance, number of trips by mode, vehicle miles traveled (VMT), and accessibility. Transportation planners use VMT to measure the length of non-work travel by automobile and use accessibility index to measure the ease to reach non-work destinations.

Trend in Household Vehicles and Non-work Travel VMT

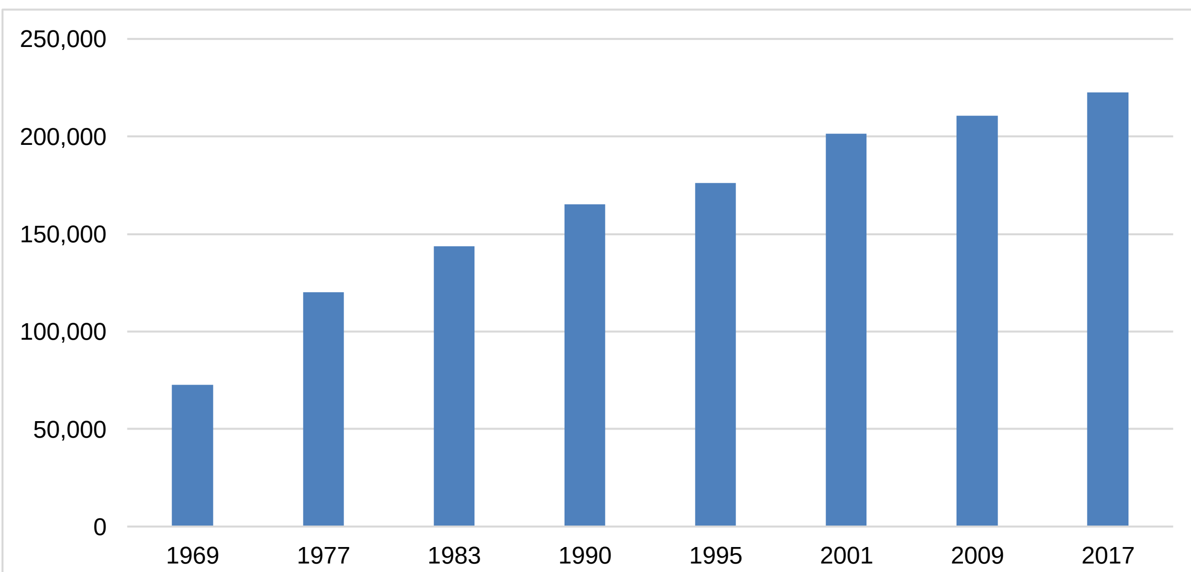

Figure 6. Number of household vehicles over time

Source: data is from table 1d, p8, McGuckin & Fucci, 2018. The count unit is in thousands.

Because the dominant means of non-work travel in the United States is by cars, the non-work travel VMT[6], which refers to the total miles traveled for non-work activities, can represent total travel length. First of all, the number of household vehicles shows increasing trends over the past fifty years (see Figure 6). The number of vehicles in households in 2017 was about 3 times more than the number of household vehicles in 1969 in the United States, and there are 1.88 vehicles per U.S. household on average.

People tend to travel more over the past fifty years (see Appendix A) and the total VMT for all trips estimated in 2017 (2,321,820 miles) is at a record high (Federal Highway Administration, 2017). In specific, compared to the rate in previous years, the total number of non-work VMT also shows an increasing trend over time (Federal Highway Administration, 2017; McGuckin & Fucci, 2018).

Non-work Travel Accessibility

Planners use "transportation accessibility" as a measure of the ease of travel when reaching a destination, suggesting a relative travel easiness for individuals to reach their destinations. Four components interact to affect accessibility: mode of transport, availability of infrastructure[7], land-use[8], and individuals[9] (Geurs & van Wee, 2004).

Two main accessibility methods are commonly used in previous studies, either "place-based measures", or "individual-based measures" (B. Y. Chen et al., 2017; Kwan & Weber,

---

[6] VMT is a unit to measure vehicle travel made by a private vehicle. Each mile traveled is counted as 1 vehicle mile regardless of the number of persons in the vehicle (Federal Highway Administration, 2017).
[7] The availability of infrastructure ensures movement and the associated travel disutility (Geurs & van Wee, 2004).
[8] Land-use provides the availability of opportunities at the destination (Ibid).
[9] Individuals represent the needs of traveling and the associated factors (e.g., temporal factors) constraining availability of opportunities. (Ibid).

2003). In this section, I describe them as two types of accessibility measure: "theorized accessibility" and "realized accessibility". The first is a distance-based method which counts the number of destinations (aka opportunities) that can be reached by a given constrain function, such as time, distance, or average cost (Geurs & van Wee, 2004). It is derived from a gravity-based model developed by Hansen (1959), still being the most widely used general method for measuring accessibility. As shown in equation below (El-Geneidy & Levinson, 2006):

$$A_{im} = \sum_{j} O_j \, f(C_{ijm})$$

where A_{im} is the accessibility at point i to potential activity at point j using travel mode m; O_j represents the opportunities at point j; and $f(C_{ijm})$ is the cost function to travel between point i and j using mode m. Based on this measure, accessibility is expected to decline the farther[10] the opportunities at j are from the origin i.

Compared to the theorized accessibility measure, the realized accessibility adheres to travel behavior theories (M. Ben-Akiva & Lerman, 1979; Neuburger, 1971). The measure requires an estimation of a cost function using empirical travel behavior data, which is a behavior-based method accounting for individual behaviors (El-Geneidy & Levinson, 2006):

$$A_n^i = ln\left[\sum_{\forall = C_n} \exp\left(V_{n(c)}\right)\right]$$

where A_n^i is the accessibility measured for individual n measure at location i; $V_{n(c)}$ is the observable temporal and spatial component of indirect utility of choice c for person n; C_n is the choice set of person n.

[10] "Farther" can be in terms of time, distance, or generalized cost (Ibid).

Theorized accessibility is theorized spatial accessibility, while realized accessibility is realized behavior. Theoretically, all people at origin i will experience the same spatial accessibility. However, in reality, people choose their destinations because of a variety of factors, such as the attractiveness of a destination, the personal preferences (perceived opportunities), the availability of travel mode, the cost of travel, and so forth (Cascetta et al., 2013; Jain & Lyons, 2008). Realized accessibility accounts for utilities, which is popular for its ability to capture the random nature of individual travel's preferences (Geurs & van Wee, 2004; Nassir et al., 2016).

2.2.2 Relationships among Travel Behavior, Travel Attitude, and the Built Environment

In travel behavior research, the built environment (BE) has long been considered as a potentially influential factor in shaping and changing travel behavior (Ewing & Cervero, 2001, 2010). The built environment-based travel demand management follows a microeconomic framework, modifying the cost of travel and the mode choice (Marsden, 2006; Weinberger et al., 2010). However, some more recent studies has found that attitudes and affective states also influence transportation decision making (Griffioen-Young et al., 2004; S. Huang & Hsu, 2009; T. Lin et al., 2017). Travel attitude is overall regarded as a more synthetic output that is affected by factors from economics, psychology, and sociology rather than an output only from a microeconomic utility model (McFadden, 2000).

Individuals' travel attitudes can be interpreted as their attitudes or perceptions towards their trips, including their emotions and opinions in reaction to their travels. Travel attitudes, including the overall satisfaction of travel, not only provide information about individuals' decisionmaking when choosing a particular travel mode but also indicate their perceptions of transportation infrastructure and services (Van Acker et al., 2010). The Theory of Planned

Behaviour (Ajzen, 1991) and the Theory of Reasoned Action (Fishbein & Ajzen, 1975) posit that a positive attitude leads to the formation of a greater behavioral intention (motivation), which is more likely in turn produce the behavior (Verplanken & Aarts, 1999). Parkany et al. (2004) reviewed literature in social psychology and transportation and found that attitudes are very important in travel mode choice. Thus, the respective transportation planning policies may not be able to have the expected effectiveness in promoting behavioral shifts. A better understanding of the dynamics between travel attitude and behavior is necessary for transportation planning and policy, which can add importance in light of recent evidence-based research on travel behavior and will help transportation and city planners to consider effective motives of sustainable urban transportation (Parkany et al., 2004).

Compared to the limited number of studies examining travel attitude, there are numerous studies focusing on the relationship between travel behavior and built environment (see reviews, e.g., Boarnet, 2011; Ewing & Cervero, 2001; Stead & Marshall, 2001; Stevens, 2017). Planners and designers consider built environment factors to be important mechanisms for encouraging mode shift. Elements such as the "5D's": density, diversity, distance to transit, destination access, and design (Ewing & Cervero, 2001, 2010) can reduce reliance on cars and parking, and to increase non-motorized modes' attractiveness. For example, Christiansen et al. (2017) found that higher density around destinations is associated with lower likelihood of using the car, and the odds also decrease when the end destination is closer to the city center. Stevens (2017) also found that compact development does make people drive less, even though the impact on reduction of vehicle miles traveled (VMT) appears to be small in magnitude.

Although a growing number of studies have explored the possible relationship over the past two decades, their relationship has yet to be confirmed. For example, because of the randomity in the attitude of travelers, some scholars argue that the observed correlations between built environment characteristics and travel behavior are at least partially attributable to travel-related attitude (Bagley & Mokhtarian, 2002). The mobility differences are largely explained by attitudes and the effect of the built environment mostly disappears when attitudes and socio-demographic factors have been accounted for (Handy et al., 2005). The complex relationships between behavior, attitude, and built environment still remains unresolved.

2.2 The Need to Use New data

In the field of transportation planning, traditional travel surveys have been widely used for obtaining attitudinal and behavioral information needed for planning management and decision-making. For example, travel attitude and behavior data - as measures for the perception of travel and behaviors, are commonly obtained from travel surveys of a population-based sample of participants in selected study areas.

However, it is important to recognize the main limitations of traditional data used for on non-work travel. First of all, most of them use data conducted from travel diaries, questionnaires, or interviews, ranging from just one day to several days and thus become less useful for understanding non-work travel on weekends, or in different seasons. In addition, constrained by the survey-based data collection method, they only can have a small sample size and may be difficult for conducting a robust statistical analysis. Therefore, although transportation planners can continuously to use traditional survey-based method to collect travel behavior data, they need to understand the limitations and challenges associated with such method.

With advancements in sensing and information technologies, recent years have witnessed an unprecedented increase in the volume, variety, and velocity of data from non-survey sources with often richer content and higher precision. Rich semantics of space and place associated with transportation perceptions and activities are embedded in these new "transportation big data", such as geotagged social media posts, mobile phone traces, check-in data, online reviews, crowdsourced traffic and street data, and so forth. In this section, I review the characteristics of new data and summarize their capabilities in travel behavior research.

2.2.1 Characteristics of Transportation Big Data

Fast development of sensing, computing, and networking techniques, social media and mobile devices have recently experienced a rapid growth, generating huge volumes of "transportation big data" almost in real-time, which brings us unprecedented opportunities for resolving transportation problems for which traditional approaches are not competent (Zhu et al., 2019). Transportation big data has the five V's characteristics of big data: volume, velocity, variety, veracity, and value (M. Chen et al., 2014; Tsai et al., 2015). Millions of users from different countries and regions in the world are producing massive amounts of information every second. The rich features embedded in these information, from their GPS coordinates, speed, to texts post on social media, recording their spatial, temporal, and emotional content that are associated with human movement (Torre-Bastida et al., 2018).

The basic properties of five main data types related to human mobility analysis from the literature are summarized in Table 3. They are "cellular service data", "WiFi data", "GPS location data", "geosocial media data", and "public transport smart card data". In general, all of these new transportation big data provide spatiotemporal information that can be used to estimate travel behaviors.

Table 3. Basic characteristics of common human mobility big data

Types of Human Mobility Big Data	**Collector**	**Properties**	**Reference**
Cellular service records	mobile phone companies	1) activity timestamps; 2) identifier of serving cell stations; 3) spatiotemporal information of mobile phone users.	Kang et al., 2010; Williams et al., 2015.
WiFi data	mobile devices via scanning surrounding WiFi access points	1) spatiotemporal data; 2) MAC address of the mobile device.	Sapiezynski et al., 2015; Traunmueller et al., 2018.
GPS location data	location data service companies (e.g., via smartphone apps)	1) spatiotemporal data with high accuracy; 2) identifier of mobile phone users.	Siła-Nowicka et al., 2016; Vazquez-Prokopec et al., 2013.
Geosocial media data	social media companies	1) user identifier; 2) text; 3) spatiotemporal information (timestamp, latitude, and longitude).	C. Yang et al., 2019; F. Yang et al., 2015
Public transport smart card data	automated fare collection systems of public transportation services	1) bus routes and metro lines; 2) spatiotemporal information of users; 3) operation information of public transportation systems.	Ma et al., 2013; Pelletier et al., 2011.

In addition, Table 4 further gives a review of specific application areas of these five main crowdsourced big data used in human mobility analysis. Overall, there are five common city-wide mobility subjects found from the literature: “distance and duration distributions”, “origin-destination matrices”, “individual activity-based mobility patterns”, “individual transportation mode inference”, and “travel attitudes and perceptions”. All these applicable mobility subjects are associated with human travel and help to understand urban human mobility. Some types of data are unable to be applied to some subjects. For example, WiFi access point data cannot be used to tracked human travel paths and distance estimations. Interestingly, for the “travel attitudes and perceptions” subject, only geosocial media data can be applied in this category. This is probably because geosocial media data has its unique textual information associated with each post (see Table 3), which contains related attitudinal information of travel.

Table 4. Crowdsourced mobility data types and human mobility research subjects

Types of Human Mobility Big Data	Common city-wide human mobility subjects				
	S1: Distance and duration distributions	**S2: Origin-destination matrices**	**S3: Individual activity-based mobility patterns**	**S4: Individual transportation mode inference**	**S5: Travel attitudes and perceptions**
Cellular service records	Kung et al. (2014) 16 million devices	Iqbal et al. (2014) 6.9 million devices	S. Jiang et al. (2017) 3.2 million devices	H. Wang et al. (2010) 1 million devices	-
WiFi data	-	-	Sapiezynski et al. (2015) 130 participants	Shin et al. (2015) 30 users	-
GPS location data	Alessandretti et al. (2017) 850 students	Ge & Fukuda (2016) 180,000 individuals	Vazquez-Prokopec et al. (2013) 582 participants	Ghorpade et al. (2015) 10 volunteers	-
Geosocial media data	Q. Huang & Wong (2015) Twitter	Yang et al. (2015) Foursquare	Hasan et al. (2013) Twitter	Mondschein (2015) Yelp reviews	Ayeh et al. (2013) TripAdvisor
Public transport smart card data	Yap et al. (2020)	Alsger et al. (2016)	Chakirov & Erath (2012)	-	-

Source: Table is based on table 1 in Y. Zhou et al., 2018.

According to the review of characteristics of these new transportation data, I identify “geosocial media data” and “GPS location data” as two potential data sources for me to investigate human travel to non-work destinations in this study. The first has user-generated content such as text posts or comments (Rybarczyk et al., 2018), serving as a potential data source to extract attitudes towards non-work travel. The second has fine-grained mobility attributes (Siła-Nowicka et al., 2016), which can be used to estimate the travel patterns to non-work destinations for a large group of demographics.

In the next two sections, I further review these two types of human mobility data's basic facts, features, and analytical methods.

2.2.2 Review of Geosocial Media Data

(1) Geosocial Media Data

Geosocial media data are the social media data having linked geographical information, which consists of characteristics of human mobility behavior in a spatial-temporal-social context. These links can be geo-tags in which location coordinates are *explicitly* attached to the social media texts, or place mentions in which texts are *implicitly* connected to the mentioned places. In recent studies, geosocial media data is being applied in multiple disciplines, including transportation, urban planning, public and political studies, management, information, decision systems, computer science, and business.

Before going further, the first question is: what is the definition and characteristics of such type of data? Obar & Wildman (2015) summarized social media definitions presented in the literature and identify the following commonalities among current social media services:

> *1) Social media services are (currently) Internet-based interactive web applications, which changed the way we interact with the online world and the other users we connect with through it because Web 2.0 applications have made the Internet more interactive;*
>
> *2) User-generated content is the lifeblood of social media. User-generated content such as text posts or comments, digital photos or videos, as well as data generated through all online interactions;*
>
> *3) Individuals and groups create user-specific profiles for a site or app designed and maintained by a social media service;*

4) Social media services facilitate the development of social networks online by connecting a profile with those of other individuals and/or groups.

With the support of social media technologies, individuals, companies, NGOs, governments, and other organizations can view, share, commend, and create information, ideas, and career interests (Barbier & Liu, 2011). An ever-increasing portion of the population makes use of social media in their day-to-day lives. Social media also plays a significant role in many aspects of daily travel behaviors, especially in information search, decision-making (before the trip), experiences / information share (during the trip), and post-travel evaluation (after the trip) (Chung & Koo, 2015; Munar & Jacobsen, 2013; Sedera et al., 2017; Xiang & Gretzel, 2010).

As shown in Table 5, well-known social media websites, including Facebook, YouTube, WhatsApp, Instagram, WeChat, TikTok, Telegram, Reddit, Twitter, Yelp, and Pinterest, have millions of monthly active users (DataReportal, 2021; DMR Business Statistics, 2021).

Table 5. Statistics of selected well-known social media platforms

Social Media Platforms	**Monthly Active Users (million)**
Facebook	2797
YouTube	2291
WhatsApp	2000
Instagram	1287
Weixin / WeChat	1225
TikTok	732
Telegram	550
Reddit	430
Twitter	396
Yelp	178
Pinterest	100

Source: data is from DataReportal, 2021 and DMR Business Statistics, 2021.

Social media platforms provide a variety of forms to the public to share information, including blogs, enterprise social networks, business networks, forums, microblogs, photo sharing, products/service review, social gaming, and video games (Aichner & Jacob, 2015). Each social media platform has its own focus. For example, Facebook seeks to build social networks; Reddit shares discussions and topics; Twitter focuses on sharing short messages and posts; and Yelp shares the reviews of dining, shopping, or recreational places. Table 6 summarizes the focuses and data formats by well-known social media platforms. Their focuses are subject to change with the changes in their business strategic plans.

Table 6. Focuses and data formats of well-known social media platforms

Social Media Platform	**Focus**	**Data Format**
Facebook	Social networking	Text, JSON, SQL
YouTube	Video	Video, Text
WhatsApp	Message	Audio, Text, JSON, SQL
Instagram	Pictures	Image, text, video
Weixin / WeChat	Message	Audio, Text, JSON, SQL
TikTok	Video	Video, Text
Telegram	Message	Audio, Text, JSON, SQL
Reddit	Discussion boards and forums	Text, JSON, SQL
Twitter	Microblog	Image, Text, JSON, SQL
Yelp	Reviews and ratings	Image, Text, JSON, SQL
Pinterest	Social commerce	Image, CSV, SQL, XML

(2) Travel Behavior Analysis Using Geosocial Media Data

In recent years, there has been rapid growth in geosocial media platforms, such as Yelp, Twitter, Foursquare and Facebook, which have attracted an increasing number of users and greatly enriched their daily urban experiences (Choe et al., 2017; Evans & Saker, 2017). Exploring the capability of new data sources such as social media to measure travel activity has become an emerging research area in the planning and design of urban transportation systems. For example, many transportation issues and behaviors can be linked with the volunteered geographic information in Twitter posts. Collins et al. (2013) use about 500 twitter texts to evaluate transit riders' satisfaction with a Sentiment Strength Detection Algorithm. Andrienko et al. (2013) extract the geotagged twitter information about everyday life of people – activities, habits, travel behaviors and experience – to understand movement patterns. Kurkcu et al. (2016) examine the spatial and temporal characteristics of human activity and mobility patterns and compare trip characteristics with satisfactory quantities using geo-located Twitter data. Kovacs-Györi et al. (2018) develop a methodology using tweets to extract visitors' spatiotemporal patterns along with the sentiments embedded in the text of tweets.

(3) Analytical Methods of Geosocial Media for Human Mobility Analysis

Past studies have applied a variety of analytical methods to process the geosocial media data for human mobility research. Researchers have developed multiple approaches to extracting information from geosocial media to track and analyze human movements. The development of data mining and machine learning allows travel experience information such as trip preferences and sentiments to be captured from geosocial media such as Twitter or Yelp (Ignatow & Mihalcea, 2016; Senaratne et al., 2017). Spatial analysis and visualization techniques also enable the user-generated geosocial media to be used to identify the most appreciated Points of Interest (POIs) and landmarks in a study area as well as to retrieve trip origins and destinations, durations, inferring activity types or classifying transportation modes (Chaniotakis et al., 2017; Nikšič et al., 2017).

2.2.3 Review of Mobile phone-based GPS data

(1) Mobile Phone-Based GPS Trajectory Data

Global positioning system (GPS) is widely adopted from industrial applications such as land surveying and aviation to personal applications such as navigation apps and devices, individual tracking, and location sharing. From the literature, I summarize the properties of three common types of GPS data: vehicle trajectories, social networking (with GPS information), and human trajectories, as shown in Table 7.

Table 7. Properties of common types of GPS data

Types of GPS Data	**Properties**	**Examples**	**Reference**
Vehicle Trajectories	1) time, GPS coordinates, velocity, accelerated velocity; 2) vehicle ID, object type, direction, change of direction; 3) name, origin, destination, station; 4) image, sound, surveillance video.	Taxi GPS trajectories, Automobile, Train/Metro, Ships etc.	Tang et al., 2015
Social Networking	1) time, GPS coordinates; 2) user ID, address; 3) text (e.g., posts such as tweets etc.), IP; 4) posts (MMS, Voice etc.)	Location-based social media datasets: Flickr, Twitter, Foursquare, etc.	D’Andrea et al., 2015
Human Trajectories	1) time, GPS coordinates, WiFi coordinates, velocity, accelerated velocity, gravity; 2) cell tower ID, service types; 3) user ID; 4) address.	Cell phone-based trajectories, smartphone-based GPS traces, etc.	Hardy et al., 2017

All these three types of GPS data in Table 7 have location coordinates and can be applied into transportation research and practices. For example, the vehicle trajectories, such as taxi GPS data (floating car data) can be used to estimate real-time travel time on roads with both high spatial resolution (10 m coordinate errors) and temporal resolution (less than 30 s).

In particular, the human GPS traces collected by mobile phones become a potential data source that can be used for analyzing human mobility patterns. With the development of Internet of Things (IOT) and communication technologies, location information is just in people's pockets. Mobile phone use has been widely across the world. There are 5.11 billion unique mobile users worldwide in 2019, and 2.71 billions of them use smartphones (Pew Research Center, 2019). The number of American smartphone users has grown by more than 200 million in the past decade. In 2018, there were 261.34 million smartphone users in the United States (Statista, 2019), and 35% of US smartphone users check their phones more than 50 times a day (Pew Research Center, 2019). As shown in Figure 7, in 2019, the vast majority of Americans - 96% - owned a cellphone, and the share of Americans that own smartphones was 81%.

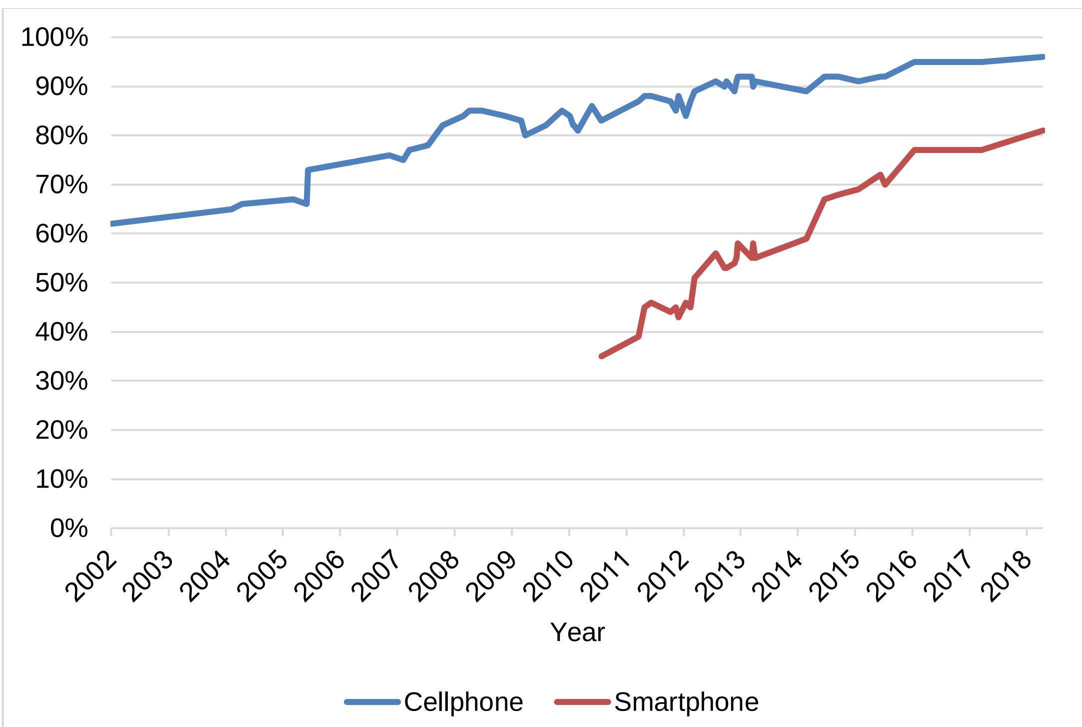

Figure 7. Percentage of U.S. adults who own mobile phones

Source: Pew Research Center, 2019.

Along with the popularity of mobile phones with built-in GPS, location information is available in many applications, such as location-based social networks, geocaching, and geotagging. Among all types of crowdsourced data, smartphone-based GPS trajectories have the high-quality localization performance (usually ranging from 1 m to up to 20 m) (C. Wu et al., 2015). Besides the GPS coordinates, smartphone-based GPS data also record time, velocity, accelerated velocity, accuracy, user identifier numbers, etc. As a result, this type of GPS data offers opportunities to accurately locate people's home, work, entertainment, and other locations.

(2) Travel Behavior Analysis using Mobile Phone-Based GPS Traces

Passively collected data, for example, trajectory data collected from mobile devices' global-positioning system (GPS) log, present potential opportunities to reflect aspects of people's travel behaviors. The increased market share of mobile phones in recent years and rapidly increasing adoption rate has accelerated the integration of this technology in travel behavior data collection and analysis.

To obtain an understanding of what information about human travel patterns can be detected from mobile phone GPS data, I review literatures that focus on travel behavior analysis using phone-based GPS traces. Overall, individual GPS data has been popularly used for analyzing trip characteristics, detecting human mobility patterns, and evaluating urban accessibility from recent studies (C. Chen et al., 2014, 2016; Marra et al., 2019; Reddy et al., 2008).

First, phone-based GPS data can be used to determine a variety of trip characteristics, such as speed, travel distance, direction, travel mode, etc. For example, Reddy et al. (2008) created a transportation mode classification system based on six participants' GPS data collected from their mobile phones, which can determine if a participant is stationary, walking, running, biking, or in motorized transport. Zhou et al. (2017) used several millions of urban residents' smartphone GPS records to examine their travel speed, trip distance, direction, as well as the mode of transportation used during a trip segment.

In addition to the trip characteristics, researchers use individual GPS records to analyze travel patterns in space and time. Khetarpaul et al. (2011) aggregated individuals' GPS traces collected from 62 users over a period of two years to analyze their frequently visited locations. They detected a total of 42 locations that were frequently visited by the participants in Beijing and Hangzhou, such information can be helpful for learning travel behavioral patterns, for planning billboard locations for advertising, and for various other city planning related tasks. Dang et al. (2018) investigated human mobility patterns based on a large-scale spatiotemporal data. They examined daily activity patterns of the participants in Singapore and Sydney and compared their levels of mobility across different regions and demographic groups.

(3) Analytical Methods of Phone-Based GPS Data in Human Mobility Analysis

Past studies have developed a variety of analytical methods to process the individual level phone-based GPS records for human mobility research, including trajectory data mining and visual analytics methods (B. Y. Chen et al., 2018; Mora et al., 2017). The advances in data mining, machine learning, and artificial intelligence allow human movement information to be extracted from GPS points. First, trajectory data mining is an interdisciplinary subfield of computer science involving with computational process of knowledge and pattern discovery using large-scale trajectory data (Dabiri & Heaslip, 2018). Besides detecting locations, the use of accelerometer may provide further detail of movement signatures to derive travel mode, if sampled at a high enough frequency. Given the large quantity of the raw GPS data, Python programming applications are useful to facilitate the estimation of travel mode, purpose, origins and destinations. In general, data cleaning and processing, clustering, classification, summarization, and abnormality detection and regression analysis are the representative methods (Zhao et al., 2016).

Besides trajectory data mining, spatial analysis and data visualization tools are widely used for analyzing the spatial relationships and showing spatial patterns of the traces. For example, the spatial analysis and visualization techniques can help to transform the collected GPS traces into appropriate visual representations (e.g., spatial clusters), which greatly improves efficiency of mobility pattern identification and analysis (Thomas & Cook, 2006). There have been many approaches and software systems that run such analysis for travel behaviors. For example, Charles-Edwards (2014) visualized GPS traces collected from 151 tourists in Noosa, Australia using ArcGIS heatmap. The heatmap assisted to summarize their participants' mobility and activity spaces and reveals marked homogeneity in the circuits of people visiting Noosa. Other GIS online applications also can be built for examining the detected geographic position and attributes of travel, such as route choice and travel time. Compared with trajectory data mining methods, spatial analysis methods allow decision makers to combine their background knowledge, creativity, and human flexibility to gain insight into complex problems (Keim et al., 2010).

2.3 Research Gaps and Opportunities

2.3.1 Research Gaps

Although the literature provides an important foundation for proceeding the travel behavior research, there still exists several research gaps.

First, travel behavior studies, despite the different data and analytical methods used in empirical studies, mainly focus on commuting trips. Non-work trips and travel behaviors have received attention in the urban travel literature. Although ongoing increased number of non-work trips and their consequences to traffic congestion has recently caused attention, a limited number of empirical studies examining travel to non-work destinations and associated accessibility inequity issues. Therefore, the first research gap lies in the need for more research on understanding non-work behavior and accessibility.

Second, the data quality and resolution used in existing studies need to be improved for providing a more comprehensive understanding of non-work travel and a discovery of detailed travel attitudinal and behavioral information. Travelers have much more flexibility to choose where they go for shopping, dining, supermarket, entertainment activities compared with the commuting trips. Existing empirical studies on such non-work travel largely rely on data collected from conventional travel surveys. While these surveys can gather information about the participants' demographic, socioeconomic, and trip-making characteristics, they have noticeable limitations. These challenges include declining sample sizes (scale), under-reporting of trips, imprecision or absence of locations and times, costly, and infrequently updated content. Potential data sources from cellphone call records, which can improve the data quality to some degree, especially in improving the quantity of data and sample size, but the geographical precision is still too low due to the mobile towers' service area (approximate 3 km^2). Meanwhile, such data sources may have new problems, such as issues in data representativeness because it is difficult to know travelers' demographics and trip details. Thus, it is necessary to explore finer grained data (more detailed, cost-effective, and high resolution in time and space) to explain travel behavior.

Third, while new transportation big data have been applied to study human mobility in recent years, some of the methods used in existing studies need to be carefully examined and adjusted to better explain attitudinal or behavioral patterns more specifically. Because many big data analytical methods such as data mining, big data integration, and machine learning, are developed based on case studies from the field of data science or computer science. When applying such methodological “pipelines” into travel data analysis to answer a specific travel behavior question, researchers need to validate the methods and examine hidden constrains. For instance, current studies utilize social media reviews, e.g., TripAdvisor travel reviews, to gather the overall attitudinal polarity of travelers toward a tourism destination (Lee et al., 2020). However, they deployed textual mining methods for the entire review texts, which may not be able to fully represent the travel related emotions since the entire review texts may include travelers’ attitudes and emotions irrelated to travel. Other studies use GPS-enable devices (e.g., movement tracker) to collected human movement data, occasionally, track information may be lost when the participant moves indoors where GPS signals cannot be received. More strict data processing methods are necessary to detect and address such constrains.

In addition, most of the current research on social media or other crowdsourced datasets are descriptive and do not often address whether limitations in these data can be mitigated conceptually or methodologically, or how they can address specific and realistic questions. For example, users of social media or mobile phones are not necessarily representative of the population at large. Specifically, there are differences in use of mobile devices and social media platforms across level of income, sex, age, and so forth. Further research is needed to address such “self-selection” bias associated with transportation big data (Spyratos et al., 2019). Introducing potential measurement to estimate the sample population of big data and is helpful to increase the representativeness and facilitate sensible use of big data for transportation planning decision-making.

2.3.2 Opportunities building on the Literature

To make up for the above research gaps and shortcomings as well as broaden the travel behavior studies, using crowdsourced data opens opportunities building on the literature. In this dissertation, I integrate geosocial media and phone-based travel data to analyze experience of travelers and their mobility patterns by taking Phoenix, Arizona as a case study area. Geosocial media data is collected from Yelp reviews, and the phone-based GPS trajectory data is collected from Phoenix residents. This dissertation research conducts three empirical studies to examine residents' everyday travel experience from the texts in geosocial media data and residents' travel behavior from the trajectory data, and further analyzes the associations between the experience, travel behavior, and the built environment. The exploration of these new data and methods can reveal new dynamics and open up new approaches to study travel experiences and human mobilities, as well as provide methodological framework for data-driven transportation planning and research.

Following this chapter, Chapter 3, 4, and 5 introduce each empirical analysis respectively. Each of them gives an evidence of how to leverage the data potential to analyze specific questions associated with non-work travel, explains the data capabilities systematically, and discusses their limitations in detail at the end.

CHAPTER 3

EMPIRICAL ANALYSIS 1: GEOSOCIAL MEDIA DATA IN TRAVEL EXPERIENCE INFORMATICS

In this chapter, I examine the information about travel to non-work destinations (non-work destinations, activity type, the experience of travel, etc.) using a spatially-precise location-based social network (LBSN). The exploration and empirical analyses in this chapter set the foundations for three publications during my doctoral time - a paper examining the effects of proximity to rail transit on non-work travel (Z. Jiang & Mondschein, 2019), a paper analyzing the associations with parking sentiment and parcel-level parking supplies (Mondschein et al., 2020), and a paper investigating parking sentiment and its relationship to parking supply and the built environment (Z. Jiang & Mondschein, 2021). A more detailed list of publications related to this dissertation can be found in Chapter 6.

3.1 Introduction

LBSNs, also called geosocial media, consist of shared human experiences, often with textual content, associated with geographic locations (Crampton et al., 2013; Rybarczyk et al., 2018). An ever-increasing segment of population uses geosocial media in their daily lives. For example, Yelp is a crowdsourced LBSN consisting of many non-work activity destinations. People write “tips” and “reviews” on Yelp platform when they have visited a destination in a city. Thus, Yelp reviews can be used for rating and describing non-work activity experiences. The activity reviews posted in Yelp not only provide information on experiences while at a business but can also indicate how reviewers travel to and from an activity.

In this chapter, I use a Yelp dataset with approximately 2 million reviews for Phoenix Metropolitan Area, Arizona, to examine how the public perceives their parking environment when driving to diverse non-work destinations. The density and spatial precision of observations allow us to categorize non-work activities in urban areas as well as allow us to examine the sentiment towards parking in a way that traditional travel surveys cannot. Methodologically, I first analyze the dataset characteristics, and then use textual analysis to extract transportation content and sentiment associated with reviewed activities in each major commercial and mixed-used district.

3.2 Related Work

3.2.1 Significance of Understanding Driver Attitudes and Perceptions towards Parking

Transportation planners acknowledge the complex relationship between parking and issues such as traffic congestion, mode choice, economic activity, and development patterns. They understand that simply providing more parking can be counterproductive, and searching for parking in commercial and mixed-use areas can waste fuel, contribute to traffic congestion, overload local parking supplies and spill into adjacent neighborhoods (Shoup 2006, 2011). While planners seek to manage urban parking, driver perceptions of parking availability are a critical component of the choice to park and demand for additional parking capacity. INRIX reports that sixty-one percent of US respondents reported feeling stressed looking for parking, and sixty-three percent stated that they avoid destinations because of expected difficulty finding parking (INRIX, 2017). Furthermore, customer perceptions of parking availability are a serious concern for business owners, who frequently see driving as the primary means of access to their establishments (Bureau of Transportation Statistics, 2018). Collectively, driver sentiments and their effects on business owners can place serious pressure on local planners and political leaders to provide more parking. However, contemporary planning best practice encourages planners to manage access to commercial and mixed-use destinations by providing shared or priced parking and by designing built environments amenable to alternative modes. Given this tension, a better understanding of driver attitudes and perceptions towards parking may inform planners seeking to foster multimodal, sustainable transportation and urbanization.

3.2.2 Relationships among Parking Choices, Facilities, and the Built Environment

The provision of parking has become an important component of suburban and even urban accessibility (Manville & Shoup, 2005), and parking availability can significantly affect the probability of choosing automobile travel mode option (Pandhe & March, 2012). North American cities have long included parking requirements for new urban development, but particularly in older areas, widespread automobility combined with relatively dense development has resulted in parking shortages, as perceived by drivers (Shoup, 2011). Today, as planners seek to facilitate multimodal, transit-oriented development in both cities and suburbs, parking is again being limited in many commercial and mixed-use areas (Dittmar & Ohland, 2012).

Rather than build more parking, transportation planners use both pricing and built environment strategies for reducing parking demand and encouraging mode shift from driving to more sustainable travel modes. For instance, charging for parking has become a widely-used approach to managing parking demand (Millard-Ball et al., 2014). Pricing is an important mechanism for controlling automobile use because (a) people are sensitive to parking cost, as well as parking search and walk times in choosing destinations and mode, and (b) parking supply and price are at least partially controllable through policy levers, such as zoning, regulation, and taxation (Inci, 2015). At the same time, planners and designers consider built environment factors to be important mechanisms for encouraging mode shift. Elements such as the "5D's": density, diversity, distance to transit, destination access, and design (Ewing & Cervero, 2001, 2010) can reduce reliance on cars and parking, and to increase non-motorized modes' attractiveness. For example, Christiansen et al. (2017) found that higher density around destinations is associated with lower likelihood of using the car, and the odds also decrease when the end destination is closer to the city center. Stevens (2017) also found that compact development does make people drive less, even though the impact on reduction of vehicle miles traveled (VMT) appears to be small in magnitude.

Pricing and built environment-based parking demand management follow a microeconomic framework, modifying relative costs of driving and other modes (Marsden, 2006; Weinberger et al., 2010). However, attitudes and affective states also influence transportation decision making (Griffioen-Young et al., 2004). The Theory of Planned Behaviour (Ajzen, 1991) and the Theory of Reasoned Action (Fishbein & Ajzen, 1975) posit that a positive attitude leads to the formation of a greater behavioral intention (motivation), which is more likely in turn produce the behavior (Verplanken & Aarts, 1999). Parkany et al. (2004) reviewed literature in social psychology and transportation and found that attitudes are very important in travel mode choice. Parking behavior may be determined by attitudes and intentions. For example, Bamberg et al. (1999) found that attitudinal factors toward parking fees, parking space availability, and gas tax rises affect travel mode. The decision of whether and where to park is based on perceived impedances as well as affective qualities of travel, such as the stress of finding parking (INRIX, 2017).

For transportation planners, an equally important relationship is that between transportation experiences and attitudes towards specific planning interventions. Support for road building, for example, is associated with more driving (Börjesson et al., 2015) and increasing regional congestion (Rose, 1990). Therefore, a better understanding of the sentiment towards parking - which is a measure of positive or negative feelings and attitudes toward a thing or phenomenon - may inform planners 1) better understanding how the customers evaluate the parking management strategies and interventions, and 2) seeking to foster multimodal, sustainable transportation and urbanization as well as help shed light not just on the behavioral effects of those strategies but their political feasibility.

3.2.3 Transportation Research using Geosocial Media Data

In recent years, large geosocial media datasets, such as Yelp, Twitter, TripAdvisor, and Facebook, have expanded rapidly, attracting an increasing number of users, who often use these services to help make destinations and route choices for travel (Evans & Saker, 2017). With text mining methods, transportation researchers are able to extract travel information from online text reviews and connect it with specific locations (Sekar et al., 2017), investigate travel mode choice to non-work destinations (Z. Jiang & Mondschein, 2019), and use travel-related reviews to implement a planning decision support system (X. Zhou et al., 2017). These data have the potential to address documented limitations with traditional travel surveys: declining sample sizes (P. R. Stopher & Greaves, 2007), under-reporting of trips (Forrest & Pearson, 2005), imprecision or absence of locations and times (Arribas-Bel & Bakens, 2019; P. Stopher et al., 2005), and infrequently updated content (Chen et al., 2010). Compared to traditional survey data, textual analysis methodologies can provide distinctive insights from LBSNs and supplement existing travel analysis, as well as allow investigation of variability in travel attitudes linked with destinations across neighborhoods, cities and countries, at high volume and spatial precision (Sekar et al., 2017).

Sentiment analysis (SA) estimates people's opinions, attitudes, and emotions from written language. SA is a component of natural language processing (NLP) and is also widely utilized in text mining and machine learning (Liu, 2012). The development of SA methods has allowed LBSN sentiment mining in order to estimate attitudes in geographic contexts. Specifically, with the help of improved NLP techniques (Aggarwal & Zhai, 2012), the text extracted from LBSNs can be analyzed to identify the emotional content of behaviors in urban environments (Roberts et al., 2019). However, the analysis of sentiment for "big" textual data is challenging due to the fact that human interpretation of each observation would be too time consuming to be effective with limited time and money. This challenge can be addressed by means of automated SA techniques focusing on determining the polarity – positive or negative – of natural language text. Among these techniques, lexicon-based SA methods for classifying the polarity of texts have gained attention in recent work and their performance has been shown to be robust across domains and texts (Ding et al., 2008; Taboada et al., 2011).

In addition, a majority of SA literature primarily focuses on broad geographic scales, such as cities or regions. For example, Caragea et al. (2014) performed sentiment classification of user posts in Twitter during the Hurricane Sandy and visualized these sentiments at global and regional scale. Mitchell et al. (2013) investigate correlations between individuals' posts and a wide range of emotional, geographic, demographic, and health characteristics using geo-tagged Twitter data. However, these prior analyses lack more geographically specific analysis of factors that may affect travel attitudes and behavior.

3.3 Yelp Data

3.3.1 Overview of Yelp Dataset

Yelp is a popular LBSN dataset with nearly 180 million unique visitors per month (see Table 5). It provides a platform where reviewers rate “businesses,” including a variety of destination types, and contribute long-form text reviews so that users can make more informed non-work activity choices. The online text of Yelp reviews contains relatively rich information about the travel experiences of a variety of travel modes (Z. Jiang & Mondschein, 2019; Majid et al., 2013; Mondschein, 2015).

I use the 2019 release of the *Yelp Academic Dataset: Round 13* (Yelp, 2019a), which provides over 6 million full text reviews in ten metropolitan areas in North America: “Champaign, Illinois”, “Charlotte, North Carolina”, “Cleveland, Ohio”, “Las Vegas, Nevada”, “Madison, Wisconsin”, “Phoenix, Arizona”, and “Pittsburgh, Pennsylvania” in the US, “Calgary, Alberta”, “Montreal, Quebec”, and “Toronto, Ontario” in Canada. These reviews are collected from October 12, 2004, the time Yelp started to have review records, to November 14 2018, the end time of the Round 13 dataset. A total of 6,685,900 reviews associated with 192,609 businesses from 1,637,138 distinct users are included in this dataset. Figure 8 shows the Yelp review count by year, suggesting an overall sharp increasing trend [11] and an exponential increase trend in the early years.

[11] The data collected in year 2018 is from January 1 to November 14, so the total number of reviews in year 2018 was slightly lower than 2017.

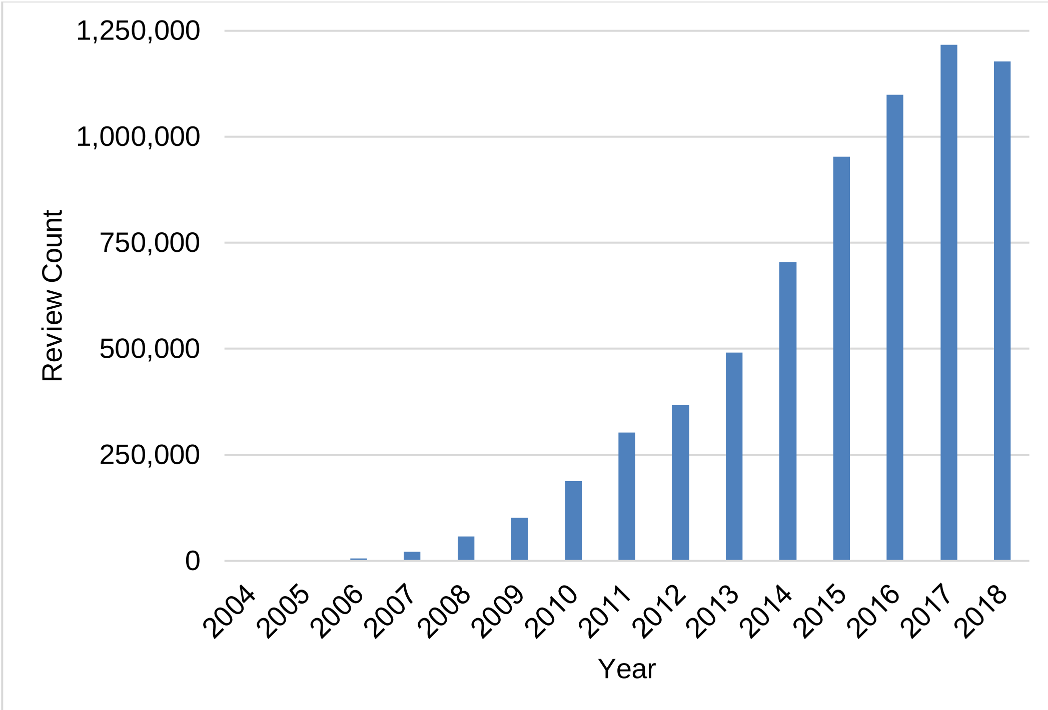

Figure 8. Count of Yelp reviews by year

3.3.2 Time and Space Information embedded in Yelp Dataset

Each Yelp review has the precise latitude and longitude of each reviewed business and is timestamped in terms of when the review as submitted (not when the activity took place). Besides the spatial location and timestamp information, each business in the Yelp dataset is originally categorized using a multi-label classification approach (Tung, 2015) with nearly 1,000 categories (Yelp, 2018). I reclassify these categories into Yelp's reported "10 big categories," which are *Active Life, Arts, Automotive, Health, Hotels & Travel, Nightlife, Other, Restaurants, Service,* and *Shopping*, transforming each business from multi-label to single-label by using an identification algorithm to match the business within the 10 big categories. Because each business in the raw dataset includes multiple category labels with the first being the main category, the algorithm selects the first label from the raw dataset and assigns a single category to the business (Z. Jiang & Mondschein, 2019). As shown in Figure 9, *Restaurants*, *Service*, and *Shopping* have larger number of counts than other non-work activity types in the dataset. Among them, *Restaurants* category takes the largest share - over 60,000 restaurants are included in the Yelp dataset.

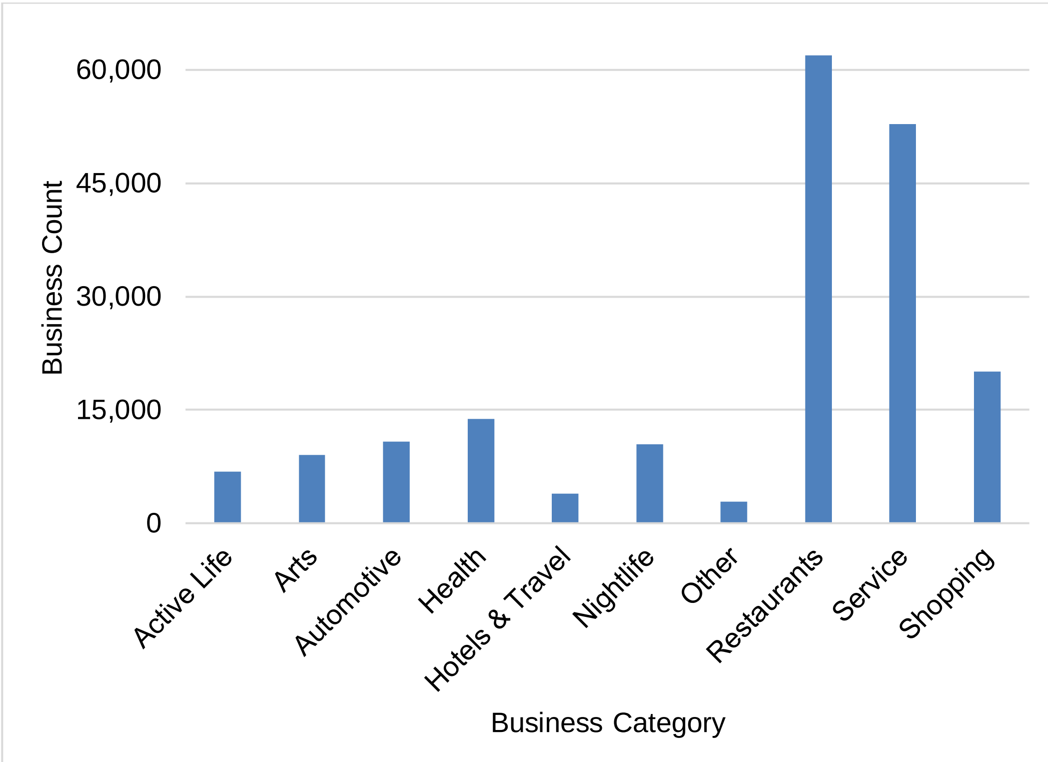

Figure 9. Count of business by category in Yelp dataset

What about these businesses' review count? Are their review count consistent with their quantity count? As shown in Table 8, restaurants attract the largest number of people to write Yelp reviews which takes more than half (52%) of the total reviews. Nightlife and service businesses comprise 14% and 13% of the total reviews, respectively.

Table 8. Count of reviews by business category in Yelp dataset

Business Category	Review Count	Percentage
Active Life	137,432	2%
Arts	370,915	6%
Automotive	214,225	3%
Health	196,630	3%
Hotels & Travel	173,110	3%
Nightlife	906,292	14%
Other	29,055	< 1%[1]
Restaurants	3,502,689	52%
Service	852,807	13%
Shopping	302,745	5%

Note [1] The number of "*Other*" business type is 29,055, and the total number of reviews is 6,685,900, the percentage is 0.4%.

3.3.3 Parking Attributes embedded in Yelp Dataset

In addition to textual review information, most businesses have associated "parking attributes", a set of binary categories (*True/False*) indicating the availability of five parking attributes at each business, such as "*parking garage*", *"parking lot"*, *"street parking"*, *"parking valet"*, or *"validated parking"*. The parking attributes provide a means of ground-truthing the type of parking supplied in different neighborhoods across the study area, though they do not indicate the absolute quantity of parking supplied. In addition, they represent specific strategies used by businesses, business collectives such as business improvement districts, and planners to more effectively manage parking supplies in commercial and mixed-use districts. A closer analysis of parking attributes is in Chapter 5.

3.3.4 Representativeness of Yelp Dataset

One limitation of using social media data such as Yelp is it lacks embedded demographic and socio-economic information about each reviewer. For this analysis, my population of interest is patrons of establishments in urban commercial and mixed-use districts. While these patrons may not represent all urban residents, the sentiments of this self-selected group are likely to have significant impacts on local parking demand as well as local planning and decision-making. Still, Yelp users themselves may not be representative of all patrons of establishments in urban commercial and mixed-use districts. Therefore, I use empirical methods, where possible, to determine demographic characteristics from available data in order to assess whether these factors are likely to have a significant impact on the outcome variable. These approaches, described here, include comparison of Yelp users to aggregate population data, imputation of demographic characteristics, and the use of proxies for demographic information.

(1) Aggregate Demographic and Socio-Economic Information

First, I consider Yelp users in the aggregate relative to the population as a whole. As of 2019, Yelp has an average of over 180 million monthly unique users (Yelp, 2019b). Comparing the demographics of Yelp users from a Quantcast survey (Quantcast, 2017) to US Census (U.S. Census Bureau, 2018) and Canada Statistics (Statistics Canada 2018) data on the general population, Yelp users are more female (61% of users) than US and Canada census respondents. Yelp users' households are also slightly more educated and wealthier on average than households in the US and Canada (Yelp, 2019b).

(2) Gender

Given the aggregate differences between Yelp user demographics and the population as a whole, I assessed whether a significant relationship may be observable between the outcome variable of interest, parking sentiment, and specific demographic factors. Note that I describe the sentiment analysis methodology further in the section "Sentiment Analysis of Parking Reviews in Yelp" below.

I considered gender as a demographic characteristic of the Yelp users that may influence parking sentiment, using a name-based prediction method to predict the users' gender using the "*user_name*" variable for each review. This variable is the username chosen by users when they register. Most of these usernames (~90%) were a standard name word (the first name), I used the R package "*gender*" (Mullen, 2020) to predict each user's gender based on their username. The prediction of the gender package is based on first names using historical datasets. After prediction, I found that the percentage of the predicted female group was 55% (# of count: 809,482) and the predicted male was 45% (# of count: 652,229). 89.2% usernames were used to conduct prediction. About 10% usernames were unpredictable inputs, since they were a single letter or character combinations that cannot be found as a name in the historical name datasets.

Using the names with assignable male or female genders, I linked the predicted gender information with the parking reviews, and the Spearman correlation analysis (Schober et al., 2018), a method that can measure the association between the continuous data and ordinal data, shows near-zero correlations between parking sentiment (values are from the SA results) and predicted gender being male (r=0.006). Thus, I do not find a significant correlation between the parking sentiment and users' gender groups.

(3) Income

In order test whether a significant relationship may exist between income and sentiment, I used the cost of restaurants (included in the Yelp dataset at 4 levels[12]) as a proxy for patron income, with the reasonable assumption that more expensive restaurants will be patronized by higher income Yelp reviewers, ceteris paribus. While restaurant price is a reasonable proxy for the income of its patrons, on average, this approach would still not reveal specific interactions between individual income and restaurant price. The Spearman correlation analysis shows near-zero correlations between parking sentiment (values are from the SA results) and restaurant cost (r=0.028). Thus, I find no significant correlation found between the parking sentiment and this proxy for income levels.

(4) Race

While I was able utilize Yelp data to examine gender and income associations with the outcome variable, other demographic factors are more difficult to assign to reviewers. For example, I sought approaches to investigate the race and ethnicity of Yelp users. However, the Yelp dataset refers only to users' first (given) names. While a few R packages, e.g., *predictrace* (Kaplan, 2020), *wru* (Khanna, 2020), can use the last name (surname) to predict race or ethnicity, this approach is not validated for first names. Similarly, I am unable to determine reviewer age from available data. I discuss how these limitations could affect interpretation of the results in the Discussions and Conclusions section.

(5) Overall Business Activity Experience

[12] The four cost levels are used dollar signs on the Yelp website: *$, $$, $$$,* and *$$$$,* ordering from least costly to most costly.

Finally, I also examined whether the overall review stars are correlated with the parking sentiment, given the concern that overall activity experience and parking sentiment may be correlated. Because I extract only the text segment that specifically describes the parking experience (Mondschein et al., 2020), I expect that the parking sentiment should be isolated to the parking experience itself. When I test this association, the Spearman correlation analysis shows near-zero correlations between parking sentiment (values are from the SA results) and the review stars in Phoenix (r=0.015).

3.4 Study Area and Major Commercial and Mix-use Districts

3.4.1 Study Area

Geographically, I use a subset of the full dataset, focusing on Phoenix Metropolitan Area, Arizona. I extracted all the Phoenix data from the Yelp dataset. Since major commercial and mix-use districts in Phoenix attract visitors and traffic to them, Phoenix city government requires some amount of parking for commercial developments in their zoning codes (Phoenix City Council, 2015). As a variety of urban activities and the associated travels occur in big cities, there is an increasing need to establish a highly functional and efficient parking management solution that ensures resident satisfaction and utilizes the existing parking facilities throughout the city.

Table 9. Statistics of business and reviews by categories of Phoenix Yelp review

	Business (total #: 55,775)		Reviews (total #: 2,058,864)	
Category	**Count**	**Percentage**	**Count**	**Percentage**
Active Life	2,086	4%	47,317	2%
Arts	2,303	4%	91,125	4%
Automotive	3,875	7%	102,094	5%
Health	6,109	11%	89,799	4%
Hotels & Travel	939	2%	37,843	2%
Nightlife	1,903	3%	240,287	12%
Other	1,007	2%	11,223	1%
Restaurants	11,863	21%	981,533	48%
Service	20,081	36%	354,627	17%
Shopping	5,609	10%	103,016	5%

As shown in Table 9, the top three business categories in Phoenix, by percent of all 55,775 businesses in the Yelp dataset, are *Service, Restaurants*, and *Health*, with a percentage of 36%, 21%, and 11%, respectively. The combined number of reviews in Phoenix is more than 2 million. Among all the reviews, about 48% of them are about "*Restaurants*" category.

3.4.2 Frequently Visited Commercial and Mixed-use Districts in Phoenix

The analysis is at the level of major commercial districts and centers in the Phoenix region. Analysis at the district level allows for aggregating the data that represents an overall parking sentiment of a non-work activity site. Six unique districts are identified in the Phoenix Metropolitan Area. These districts are selected based on 1) their concentrated business activities (density), 2) their representativeness of areas with a mix of non-work activities (diversity), and 3) the number of visitors.

Specifically, in order to identify major commercial districts, first, I retrieve a comprehensive list of businesses in the study area from the *Infogroup Business USA* dataset (Infogroup, 2016), a dataset including all the US businesses' detailed information, such as their business names, locations, categories, etc. Given the dissertation focuses on non-work activities such as shopping, dining, leisure, etc., I filter businesses of interests and then perform a spatial cluster analysis of these businesses' locations on the map. According to the spatial density analysis, I examine these businesses' spatial clustering patterns. Based on the spatial cluster results, I identify seventeen commercial districts which can represent places with concentrated commercial activities in the study area. Then, I calculate the number of visits by cars in the 17 commercial districts using residents' GPS traces in Chapter 4. Finally, I select the top 6 districts and identify them as the residents' frequently visited districts which can represent places with intensive non-work activities in the Phoenix Metropolitan Area.

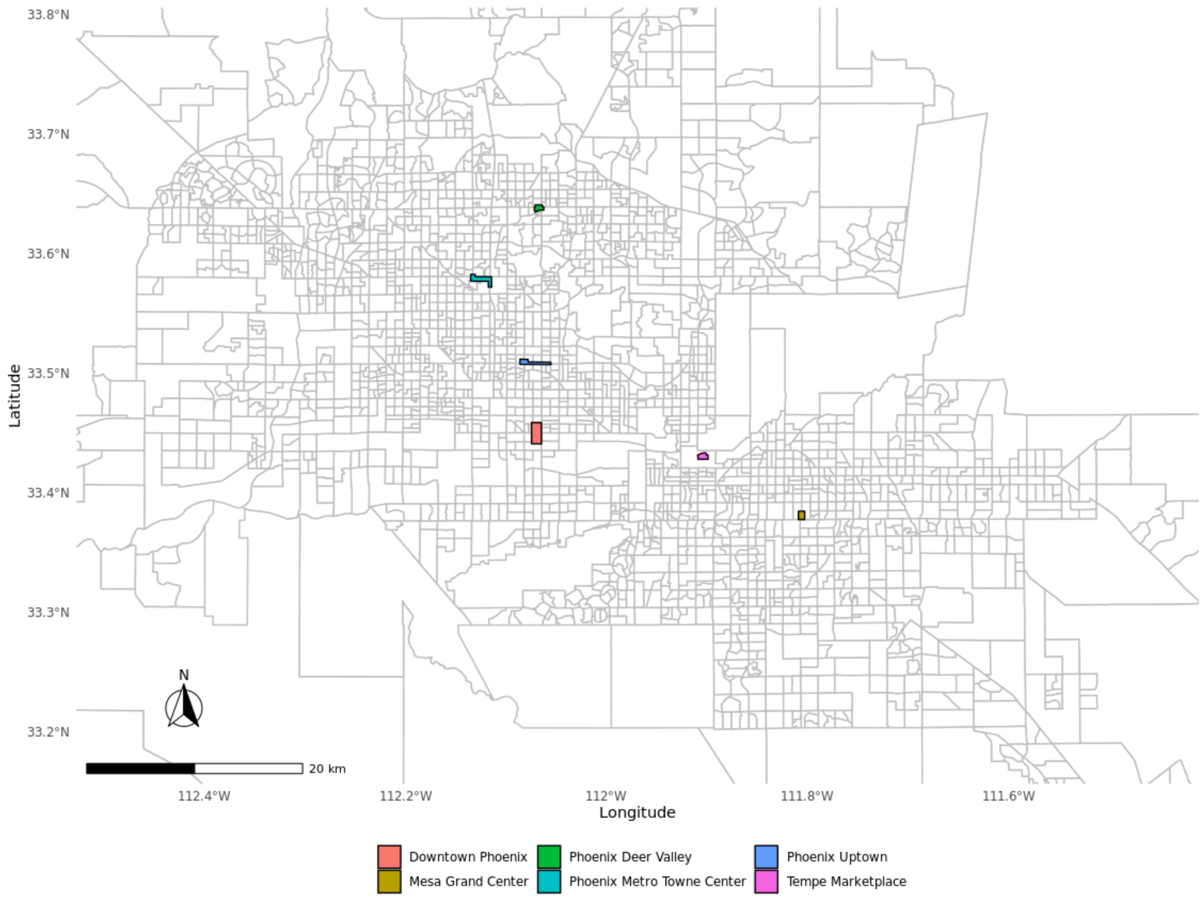

Figure 10. Major commercial and mixed-use districts in Phoenix, AZ

Figure 10 shows the locations of these six major districts. These six districts may not represent all potential districts for non-work travel study in the metropolitan area, but the intent in this research is to develop methods that can be applied widely in subsequent research and capture reviewer sentiment towards parking in some of the major districts through a data-driven approach using online reviews. Phoenix, Tempe, and Mesa are the main three cities in the Phoenix Metropolitan Area. As shown in Figure 10, four major commercial districts (Phoenix Deer Valley, Phoenix Metro Towne Center, Phoenix Uptown, Downtown Phoenix) are in Phoenix, one (Tempe Marketplace) is in Tempe, and one (Mesa Grand Center) is in Mesa.

3.5 Analytical Methods and Results

3.5.1 Transportation Content in Yelp Reviews

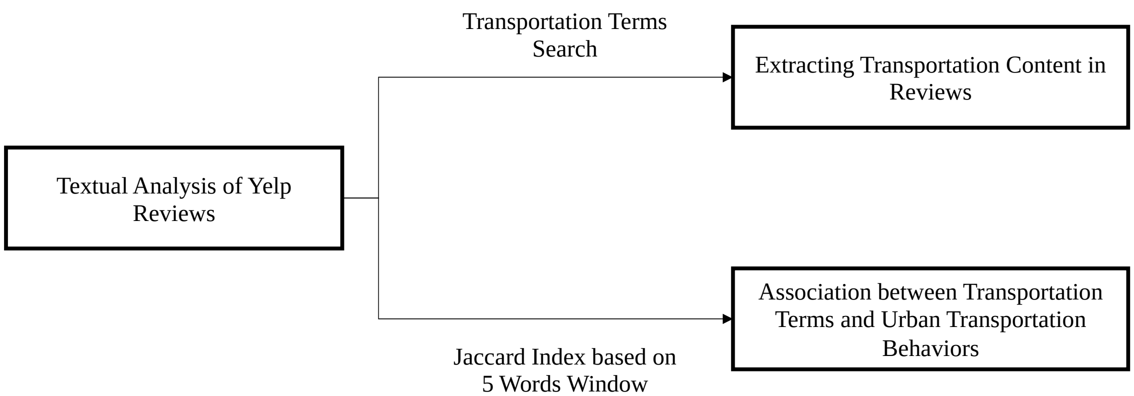

Figure 11. Methodology framework of mining transportation content in Yelp reviews

Figure 11 illustrates the sequence of analyses. I use textual analysis to extract transportation content in Yelp reviews, including a query of travel related terms to identify the texts mentioning travel experiences, and a *Jaccard similarity index* to examine the association between transportation content in Yelp and urban transportation behaviors.

(1) Extracting Transportation Content

Yelp reviews frequently include transportation content, describing the travel experience to or from a business or other destination. Examples from the dataset:

"They have their own free parking lot...very cool."

"It's a great place for running, biking, walking, etc. It's a great way to travel on bike between Old Town and Arcadia."

"There's parking validation for the structure adjacent to the theater, so that's cool. It's a bit of a walk, for handicapped, elderly, or lazy people."

"I'm a fan of this place because of the light rail convenience and the low prices."

"Hopefully the cities ramp up interests in their mass transit systems."

Table 10. transportation terms frequency for Phoenix Yelp review

Category	Terms	Phoenix Metropolitan Area	
		N	%
Auto	**car**	81,796	22%
	drive, drove	32,407	9%
	parking, parked	55,775	15%
	traffic	6,259	2%
Public Transport	**rail, train**	32,957	9%
	bus, streetcar	4,740	1%
	transit	2,381	1%
Active Travel	**bike, biked, biking, bicycle**	10,949	3%
	walk, walked, walking	144,552	39%
Total Auto		176,237	47%
Total Public Transport		40,078	11%
Total Non-auto		155,501	42%
Total Transportation Terms		371,816	100%
Total Reviews		2,058,864	

To analyze the large number of reviews with transportation content, I use a text mining approach by identifying and extracting the mentions of a particular travel experience (Hu & Liu, 2012; Krippendorff, 2012). I seek specifically modal experiences within a given review, generating measures of transportation mode experience frequency. Table 10 summarizes the frequencies of travel modal transportation terms in the Phoenix Yelp reviews. The textual analysis method is from Mondschein (2015). A mode is defined by multiple terms, such as "drive" and "drove," or "parked" and "parking." 18.1% of reviews in the dataset have identified transportation content. Note that this may be an underestimate, since not all possible terms related to transportation may be included in the selected set of terms.

"Auto" and "Active Travel" are the most frequent modes in this dataset with 47% and 42% of all reviews, respectively. Note that for the analysis, I divide auto-based terms into "driving" categories including "car", "drive", "drove", and "traffic", and a "parking" category including, "parking" and "parked". "Public Transport" category, including rail, transit, and bus are mentioned less frequently than other modes.

(2) Association between Transportation Terms and Urban Transportation Behaviors

In order to understand the usage of transportation terms in the reviews, I use the Keyword-in-Context (KWIC) technique (Ignatow & Mihalcea, 2016; Jockers, 2014; Vinithra et al., 2015), an approach examining the associations among each transportation keyword and the words that surround it – specifically, five words to the left and right of the transportation term. The analysis is completed with KH Coder software (Higuchi, 2012, 2014), using a *Jaccard index* of "shared phrases/all phrases" (Markov & Larose, 2007) to reflect the strength of concordance. Given a review's content, the similarity between every pair of noun or adjective within 5-word window and the transportation term is measured by Jaccard similarity coefficient, a statistic commonly used for comparing the similarity and diversity of sample sets.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

The index $J(A,B)$ is the ratio of the number of reviews including both A word and B word over the number of reviews including either A word or B word. Jaccard similarity coefficient ranges from 0 to 1. The strength of concordance is the association level between the target word and the substantive word. If $J(A,B) = 0$, this means A word and B word are totally unassociated. If $J(A,B) = 1$, this means A word and B word are exactly coexistence.

Table 11. Word associations with key transportation terms, nouns and adjectives by Jaccard similarity index

Parking				**Transit/Rail**				**Walk**				**Drive**			
Assoc. nouns[1]		Assoc. adjectives[2]		Assoc. nouns		Assoc. adjectives		Assoc. nouns		Assoc. adjectives		Assoc. nouns		Assoc. adjectives	
Word	JI[3]	Word	JI	Word	JI	Word	JI	Word	JI	Word	JI	Word	JI	Word	JI
lot	0.27	available	0.06	train	0.19	convenient	0.11	people	0.05	great	0.04	window	0.09	fast	0.07
spot	0.14	far	0.06	ride	0.18	central	0.11	place	0.04	enjoyable	0.04	food	0.08	favorite	0.06
space	0.13	close	0.06	metro	0.11	public	0.08	area	0.04	far	0.03	order	0.08	long	0.05
garage	0.12	small	0.06	downtown	0.11	light	0.07	parking	0.05	easy	0.03	location	0.07	good	0.04
street	0.11	open	0.06	dollar	0.11	right	0.07	building	0.04	fun	0.03	minute	0.07	nice	0.03
management	0.08	near	0.06	school	0.11	empty	0.07	street	0.04	central	0.03	coffee	0.07	quick	0.03
structure	0.08	difficult	0.06	railroad	0.10	easy	0.06	office	0.03	cool	0.03	way	0.06	worth	0.03
car	0.08	complex	0.06	stop	0.09	clean	0.06	outside	0.03	open	0.03	miles	0.04	short	0.02
area	0.07	enough	0.06	museum	0.09	urban	0.06	home	0.03	good	0.02	time	0.04	easy	0.02
traffic	0.07	designated	0.06	ticket	0.09	located	0.05	center	0.03	long	0.02	place	0.04	wrong	0.02
maintenance	0.07	great	0.06	railway	0.08	walkable	0.05	dog	0.02	various	0.02	home	0.03	down	0.02
valet	0.06	free	0.03	station	0.07	non-flexible	0.05	space	0.02	nearby	0.01	hour	0.03	friendly	0.01

[1]"Assoc. nouns" is short for "associated nouns."
[2]"Assoc. adjectives" is short for "associated adjectives."
[3]"JI" is short for "Jaccard Index."

Demonstrated by the Jaccard analysis (see Table 11), the usage and intent of transportation terms are revealed in the words proximate to the transportation terms. For "*Parking*," reviewers associate nouns like "lot,", "spot," and "space," and adjectives like "available," "far," "close," and "small." Associated adjectives for "*Transit/Rail*" are like "convenient," "central," "public," "right," "right," "empty." "*Walk*" association words are somewhat more diverse, but the majority of nouns and adjectives associated with "walk" are related to outdoor walking experiences. I exclude bicycle terms specifically because they often refer to bicycle shops.

Note that this conceptualization is supported by an examination of transportation content over time. I observe that the review-derived mode split is generally very stable except in the case of "*Rail*" in Phoenix, where the opening of a light rail line during the period revealed a significant increase in rail reviews. This responsiveness to a major change in the network supports the linkage between travel mentions/information in the reviews and the experience of transportation users.

Importantly, the modal categories presented here are not necessarily mutually exclusive – those who mention "parking" almost certainly drove, and those who mention "rail" almost certainly walked or biked. However, distinctions between these modal terms allow us to identify what is most important when accessing non-work destinations with intensive commercial activities.

3.5.2 Sentiment Analysis for Parking Reviews

The analysis in section 3.5.1 indicates that Yelp reviews frequently include transportation content. To further examine the parking content and its associated sentiment in major commercial and mixed-use districts in Phoenix, I extract the parking reviews only and deploy a sentiment analysis.

(1) SA Framework for Yelp Parking Reviews

First of all, I find that the average word count of a Yelp review containing parking keywords in Phoenix is 205 words, potentially including information of individual users' opinions and their parking experiences, or the reasons why they choose or do not choose parking when they travel to certain businesses. Examples from the dataset[13]:

"The parking is free and easy. That is awesome."

"Limited menu a hard place to find using GPS, parking can be a little hectic too"

"... just frustrated in trying to find a parking spot."

"... as always, parking is a little tough uptown."

"One downfall is that parking is horrible, with narrow spaces and not a lot available."

"Located in a very busy intersection, plenty of commuters and parking is pretty bad."

"It's a good place for quick meets with easy parking and easy access along Dundas."

"My only complaints, it was pricey and parking was challenging."

Various methods can be used for conducting a sentiment analysis. In this section, I use a lexicon-based approach to measure the emotional content of the large number of reviews with parking experience. A summary of SA steps is shown in Figure 12.

---

[13] These examples only show sentences containing parking keywords.

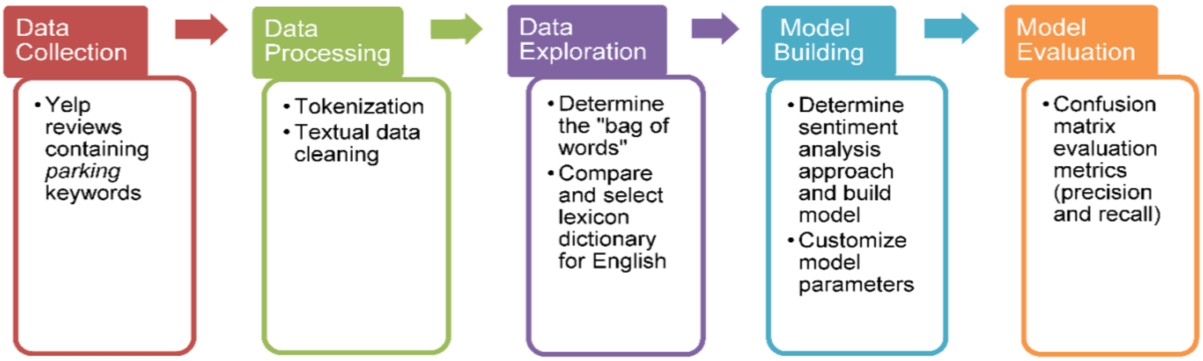

Figure 12. Sentiment Analysis framework for Yelp parking reviews

(2) Data Preprocessing

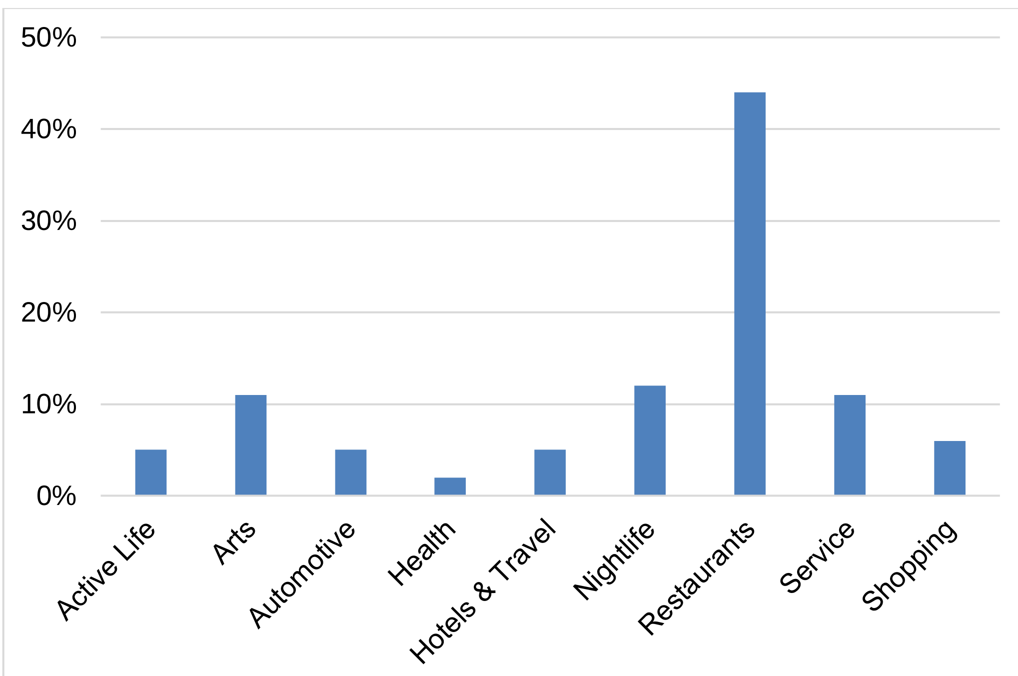

Figure 13. Percentage of parking reviews across business types

First, I use a text mining approach to identify parking reviews. In order to focus the search for parking reviews, I only use keywords "parking" or "parked" as search criteria within a given review, generating the frequency of parking reviews. As shown in Figure 13, people tend to write "tips" or "reviews" about parking when they visit "Arts", "Nightlife", "Restaurants", "Service", and "Shopping" businesses. Among them, "*Restaurants*" category has the largest number of parking reviews which takes more than 40% of all parking reviews.

Reviews with parking content from Phoenix are 15% of all transportation reviews and 2.7% of all reviews (see Table 12). 75% of parking reviews are associated with businesses providing parking attribute information. Note that this might be an underestimate, since not all possible terms related to parking may be included in the selected set of terms. In the six major commercial mixed-use districts, 5% of all reviews have parking content. 73% of these district parking reviews have parking attribute information. I use the 55,775 parking reviews that can be linked to businesses with parking attribute information for the sentiment analysis.

Table 12. Statistics of parking reviews in Phoenix Yelp reviews

	Geographical Area	
	Phoenix Metropolitan Area	**Six Major Commercial Districts**
Total number of reviews	2,058,864	108,399
Total number of parking reviews[1]	55,775	5,809
Percentage of parking reviews	2.7%	5%
Total number of parking reviews with parking attributes[2]	41,573	4,266
Percentage of parking reviews with parking attributes	75%	73%

Note [1] Parking reviews refer to reviews that mention keywords such as "parked" or "parking" in the Yelp dataset.
[2] Parking attributes refer to the parking information provided for businesses in the Yelp dataset. It includes the availability of street parking, parking lot, parking garage, parking valet or validated parking service, stored in a binary format using "true" or "false" index.

Data preprocessing includes textual data tokenization and data cleaning. Tokenization splits long strings of text into smaller pieces, or tokens. To find the best token to represent the parking reviews, I tokenize each review as paragraphs, sentences, and smaller word chunks first. Paragraph is defined by a new line in the review, sentence is defined by the ending sentence punctuation, and word chunk is defined by punctuation in the middle of a sentence. Each of these tokens must contain at least one parking keyword. Then, I compare and determine the best tokenization for the SA. I cleaned the data for each tokenized string of text, by using the ‘tm’ package in R statistical programming language (Feinerer, 2018). The cleaning process involves a sequential process for each tokenized string: making a corpus of words, converting into lowercase, removing punctuation, numbers, and URLs, stripping whitespace, and removing words irrelevant to SA, such as “the” or “an.”

Data Exploration

The word chunk is chosen as the token unit since it shows most appropriate representation of parking experience information. Specifically, neither paragraph or sentence is good enough for this case. Sentences with parking terms may be very long since some people use multiple commas instead of periods. I cut sentences into word chunks that can actually describe parking experience. The analysis uses the Harvard IV dictionary, a general-purpose psychology-based dictionary. It includes greater than 11,000 words with 1,915 positive and 2,291 negative sentiment words (Stone et al., 2007). This dictionary is able to capture sentiment through different sets of words associated with quantified sentiments (Saxena et al., 2018).

Model Building

To estimate the sentiment scores of parking reviews represented as word chunks, I use *analyzeSentiment()* function in the *SentimentAnalysis* package (Feuerriegel & Proellochs, 2018) in R to generate the initial sentiment scores. This approach is a lexicon-based approach that can classify the sentiment, returning the sentiment scores for each selected dictionary. The scores range from -1 to +1 with -1 showing an extremely negative sentiment and +1 representing most positive, with 0 being a "neutral" parking experience. The best model fits each parking review with an estimated sentiment score based on the degree of positivity and negativity in the bag of words, including assigning weights that are most predictive in the context of the observed corpus (dictionary corpus).

Model Evaluation

In terms of the nature of the large dataset, the total number of parking experience word chunks is more than 60,000, making reading each review and assigning a manual score impossible. In order to efficiently evaluate the SA model performance, I adopt a two-step performance evaluation for the model results. In Step 1, I read and check all of the predicted min sentiment scores and max sentiment scores for each district and for each business type. It produces a 2 (sentiments) * 6 (district) matrix, results are listed below:

District 1 Phoenix Deer Valley:

review (high): "Plenty of parking."

review (low): "Ridiculous parking lot!"

District 2 Phoenix Metro Towne Center:

review (high): "Easy parking."

review (low): "Hard to find a parking spot."

District 3 Phoenix Uptown:

review (high): "PLENTY of street parking nearby!"

review (low): "Parking is impossible!"

District 4 Downtown Phoenix:

review (high): "Easy parking."

review (low): "Parking is difficult!"

District 5 Mesa Grand Center:

review (high): "Plenty parking outsides."

review (low): "Overcrowded, and difficult to find decent parking!"

District 6 Tempe Marketplace:

review (high): "Plenty of parking and easy access to the mall."

review (low): "Parking is too annoying!"

I have a clear general impression of the classification from step 1's results. Then, in Step 2, I review a random sample 500 (~10% of entire dataset) of estimated sentiment scores and the corresponding original parking reviews. I manually read them one by one and create a confusion matrix (using three categories: "positive," "neutral," and "negative") to compare the precision and recall between the model results and the human-judged results. The percentage of accurate categorization is 80%. According to the precision and recall metrics, TF (positive sentiment categories erroneously classed as non-positive) = 5%, FT (non-positive sentiment categories erroneously classed as positive) = 15% (Sokolova et al., 2006). 80% accuracy, while introducing error into the subsequent analysis, is favorable for sentence-based SA using current methods. Importantly, the error is distributed across both positive and negative predictions, with some bias toward overprediction of positive sentiment. The results of the subsequent analysis should be understood keeping this potential bias in mind (Taboada et al., 2011).

(3) Sentiment Analysis Results

Parking Sentiment in Phoenix Metropolitan Area

First, I investigate the parking sentiment phenomena across the region. The SA results are shown in below. There are 26,774 positive parking reviews, 22,025 neutral reviews, and 20,226 negative reviews. The percentage is shown in Figure 14. Overall, the share of positive sentiment is a little bit higher than the negative sentiment.

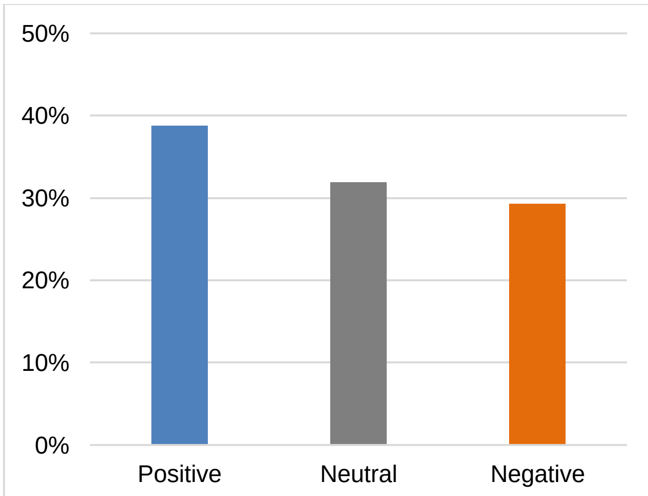

Figure 14. Percentage of sentiment categories for Phoenix parking reviews

The below Figure 15 shows the locations of businesses that have reviews mentioning parking. In general, the spatial distribution of businesses having parking reviews is across the Phoenix Metropolitan Area.

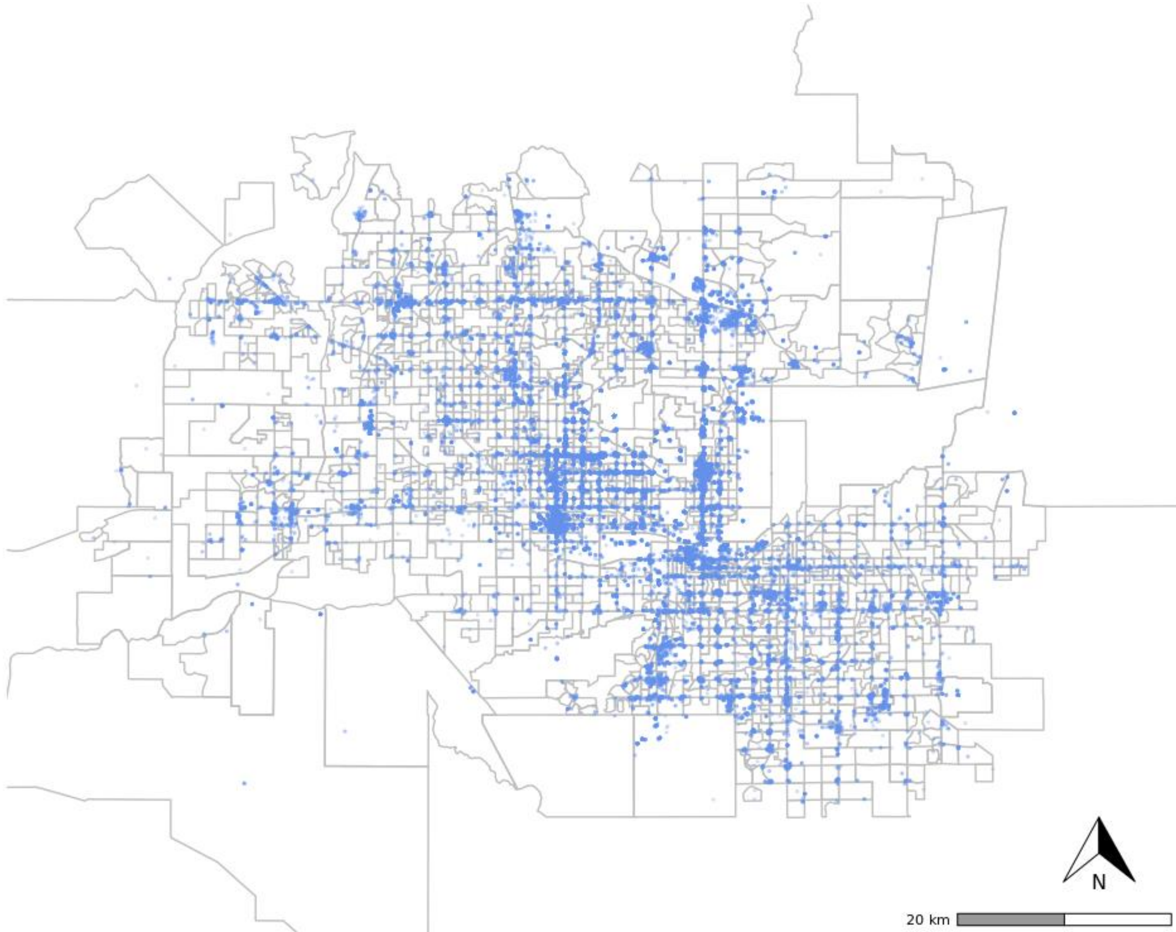

Figure 15. Locations of business with reviews mentioning parking in Phoenix

Parking Sentiment in Commercial Districts

Then, I examine the attitudes towards parking in the six selected commercial districts only. Figure 16 shows the share of parking sentiment (positive VS. negative) in each commercial area. The dominant parking sentiment in “Phoenix Deer Valley” and “Phoenix Metro Towne Center” is positive, suggesting an overall satisfaction of parking environment in these two commercial districts. By contrast, negative reviews dominate in all other four districts except the Mesa Grand Center which has almost the same share of negative and positive parking sentiment. In particular, parking reviews for “Downtown Phoenix” and “Tempe Marketplace” are both negative-dominant, which suggests an overall difficulty in parking for non-work activities. In general, commercial centers such as Phoenix Deer Valley, Phoenix Metro Towne Center, and Mesa Grand Center, have more positive parking sentiment than Phoenix Downtown and Uptown. Among them, only Metro Towne Center is a strip commercial district and other two are mall-like commercial districts. Mall-like districts, especially those located in suburban areas, often provide sufficient surface parking capacity which may allow drivers to find a place to park their cars easier.

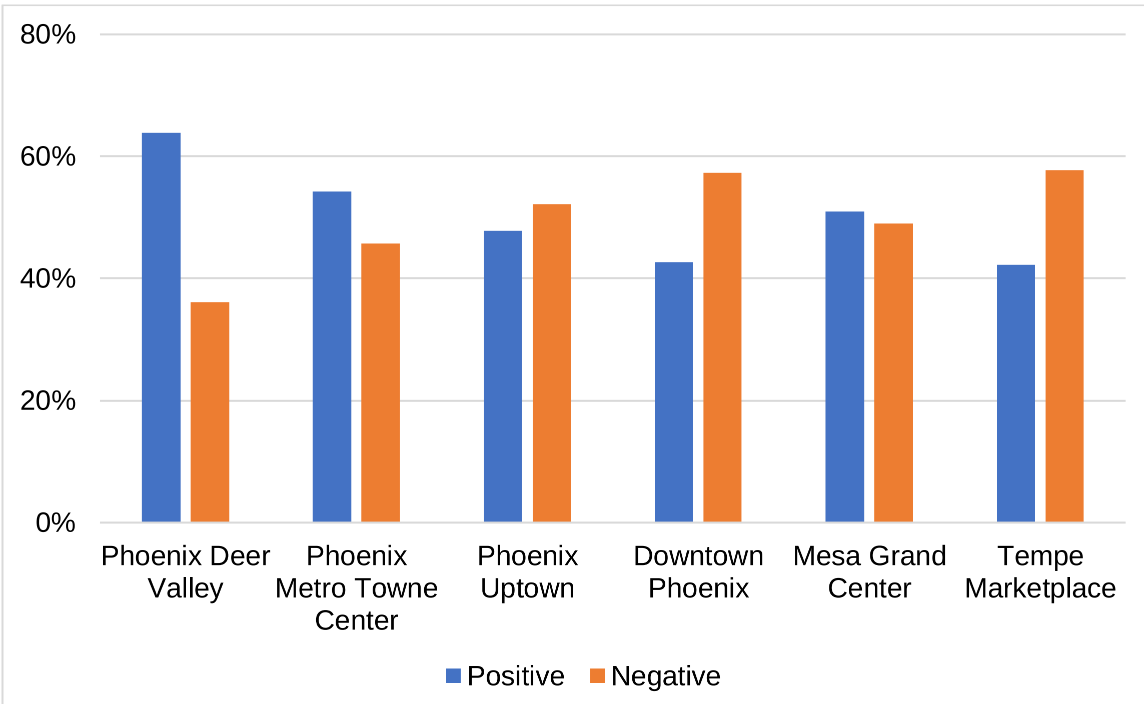

Figure 16. Positive and negative sentiment of parking across commercial districts

Based on the share of parking sentiment in each commercial district, a parking index is calculated to represent the ease to find a place to park in each district (see Table 13). This index[14] serves as an indicator showing the *parking positiveness* (including attitudes like "easy to find a place to park", "convenient parking", "free parking", "available parking") in the site.

---

[14] This index of each commercial district is calculated as the share of positive parking divided in the share of negative parking in each site.

In next chapter, I use the smartphone-based GPS data to further analyze travel behaviors in these districts.

Table 13. Parking sentiment and index of each commercial district

District Name	**Parking sentiment**		**Parking index**[1]
	Positive (%)	Negative (%)	
Phoenix Deer Valley	63.83%	36.17%	1.76
Phoenix Metro Towne Center	54.29%	45.71%	1.19
Phoenix Uptown	47.80%	52.20%	0.92
Downtown Phoenix	42.70%	57.30%	0.75
Mesa Grand Center	51.02%	48.98%	1.04
Tempe Marketplace	42.24%	57.76%	0.73

[1]: This index of each commercial district is calculated as the share of positive parking divided in the share of negative parking, as an indicator showing the *parking positiveness* (including attitudes like "easy to find a place to park", "convenient parking", "free parking", "available parking") in the site.

3.6 Discussions and Conclusions

3.6.1 Discussions

This investigation examines the transportation content embedded Yelp reviews and analyzes the sentiment towards parking for non-work activities. Broadly, the results demonstrate that a textual mining of geosocial media data can provide useful information about travel attitude - the experience of travel, particularly. The findings demonstrate that travelers are sharing experiences on online review platforms such as Yelp when going to varied types of non-work destinations. Specifically, I find that parking is of interest to Yelp reviews, which can be used to gauge sentiment about parking in commercial and mixed-use districts and centers.

For “everyday” activities such as shopping, services, eating out, and going to bars, driving plays a dominant role in mode of transportation in Phoenix. The emotions towards parking can represent the drivers’ travel attitudes. Making all drivers “happy” or “easy to find a place to park” is not the goal of transportation planning. I present this empirical analysis as an example of an exploratory analysis to assess whether online business reviews, in this case Yelp data, are of use to planners and policymakers to better understand how parking is associated with customer sentiment. Transportation access, which for many businesses includes parking, are an important policy area that can be better informed through the use of user-generated business reviews.

3.6.2 Limitations

This analysis also has some limitations. First, there may some inherent limitations of the LBSN data, for example, the issues of self-selection and representativeness in Yelp dataset. The demographics of Yelp users may not necessarily represent the entire population. Compared to the average U.S. adult population, Yelp users are younger, more highly educated, wealthier, and tech-savvy. In this chapter, I use multiple approaches to deal with Yelp data’s representativeness - an issue more broadly associated with all other self-selecting LBSNs. Addressing these issues requires a combination of empirical methods to diagnose or control for potential biases, as well as clear caveats and recognition of potential effects of biases that cannot be controlled.

As shown in section 3.3.4, for gender and income, two demographic factors that could reasonably be imputed or analyzed by proxy, I found that neither has a significant relationship to parking sentiment. However, I was not able to impute or assess the effects of age, race, or ethnicity of reviewers. For age, increased need for comfort during travel and a reduced desire to walk longer distances for utilitarian purposes has been documented (Hess, 2012; Keadle et al., 2016). Therefore, I might expect that parking sentiment among older adults would be lower for parking strategies that require more utilitarian walking. For race and ethnicity, parking experiences that require interaction with individuals, such as valet parking or even parking validation, could potentially result in distinctively racist or alienating experiences. These experiences are well documented in the context of transit, taxi, and rideshare travel, but so far undocumented for parking experiences (Purifoye, 2015; Sarriera et al., 2017). While this study's research design would not be able to address this question, further research could potentially mine Yelp's extensive reviews to identify whether parking experiences are perceived specifically as racist or alienating.

In addition, I find that there are approximately 40% of the parking reviews are from the "Restaurant" reviews. Although restaurants are more sensitive to parking woes as demand for restaurants can be more elastic as people have many choices about where they can go, future research may need to consider more types of businesses, especially parking attitudes and sentiments towards non-restaurant businesses as a complementary to the Yelp data analysis.

Broadly, working with LBSNs requires clear understanding of each dataset's strengths and limitations. Ultimately, some research questions may continue to require purpose-built survey efforts that reach populations that do not participate as readily in LBSNs, or where LBSNs do not supply critical information to answer those questions. Additional factors, such as time of day, day of week, may affect the experience of parking as well (Litman, 2006; Millard-Ball et al., 2014). Still, while a survey or online poll can provide detailed information about parking sentiment and parking behaviors, the reality for urban planners is that these surveys are rarely undertaken, and usually only in specific neighborhoods where there has been demand and funding for a parking study. The LBSN-based approach allows for a much broader look across a city, to allow for better comparisons across neighborhoods and even between cities, allowing for more empirically robust understandings of parking management and build environment strategies that result in positive parking experiences that are compatible with broader goals towards reduced total parking, densification, and multimodal travel in commercial and mixed use areas. Additionally, in future research spatially precise parking utilization data could be integrated into analysis of parking experiences in order to understand how supply, utilization, and travel experiences covary in different locations.

3.6.3 Conclusions

Yelp and similar LBSN data can be used to understand how travel may vary at fine geographic scales, controlling for factors of interest such as activity purpose and regional differences. Spatially-precise information from these data can provide insight both toward highly local, as well as regional transportation and land-use relationships. Future research could integrate additional factors, including population density, land use, employment, parking facilities, or local socioeconomic factors, to further examine what causes individuals to assign value to the modes they prioritize when going out. Also, analysis of other datasets or Yelp data for additional cities with extensive road networks such as New York or San Francisco, can extend the findings in this examination. In addition, for urban planners, these data, or similar social media data, allow a deeper look at how travelers are using the information shared on the platforms to adjust their travel choice decisions.

CHAPTER 4

EMPIRICAL ANALYSIS 2: SMARTPHONE-BASED GPS TRACES IN TRAVEL BEHAVIOR ANALYSIS

4.1 Introduction

Individuals and households have different needs for non-work activities in urban commercial districts, exhibiting divergent travel behavioral patterns when going to these places. They choose their destinations because of a variety of factors, such as the attractiveness of a destination, the personal preferences (perceived opportunities), the availability of travel mode, the cost of travel, and so forth (Cascetta et al., 2013; Jain & Lyons, 2008). Previous studies examined these various travel patterns and destination preferences and found residents from disadvantaged neighborhoods may experience an inequity of accessibility, e.g., they mainly go to lower price places as a result of low income (Neutens et al., 2010; Ren et al., 2014).

These disparities in accessing different non-work destinations also reveal that the urban isolation and segregation extend beyond individuals' residences. Therefore, understanding residents' mobility patterns to commercial districts is crucial for urban planning, facility management, and business strategies (Crane, 2000).

In this chapter, I explore and quantify individuals' travel to six major commercial districts with a specific focus on the mode use by cars, which is the main mode that used by Phoenix individuals and households for non-work travel. People there drive to a variety of non-work activities. Smartphone-based GPS data from over 90,000 individuals in the metropolitan area of a month period are used for this empirical analysis.

4.2 Related Work

4.2.1 Travel Behavior and Accessibility to Non-work Destinations

Travel to non-work activities, such as meals, shopping, recreation, and socializing, travel patterns are less routinized than commuting (Chatman, 2008; Walle & Steenberghen, 2006). Individuals demonstrate a variety of spatial and temporal needs for such non-work activities, exhibiting divergent patterns of mobility. To fully address variability in traveler preferences and behaviors, transportation planners need to have a better understanding of non-work travel. Accessibility as a measure to represent the ease of travel when reaching a destination, is widely used in travel behavior research. Research has shown that residents of disadvantaged / vulnerable neighborhoods (e.g., African Americans, Hispanics, low-income households, and households in poverty) don't have the same accessibility to non-work destinations as the residents of advantaged neighborhoods (Grengs, 2015; Stanley et al., 2011; Wee et al., 2011). In other words, even though theoretically residents from disadvantaged neighborhoods are able to freely choose their non-work travel destinations, their average travel time, distance of travel, ease of travel, etc., still have a significant difference from the residents of advantaged neighborhoods.

Researchers described the above disparities as inherent inequality, and a variety of factors might be associated with it (Neutens et al., 2010; Ren et al., 2014). In Chapter 2, I summarize main findings from previous studies: a variety of factors might contribute to the such divergent behavior patterns. For example, travel attitude, built environment (transportation and infrastructure plans), or personal factors have an influence on behavioral choice. With noting this, planners need to understand travel behaviors before establishing any transportation plans. Thus, as an important first step, an investigation of the non-work travel behavior and its associated accessibility measures become critical for providing future planning guidelines.

4.2.2 Travel Behavior Research using Phone-based GPS Trajectories

(1) Examining Travel Patterns

For a long time, the lack of fine-grained data impedes the travel behavior measurement at high resolution. Traditionally, data for travel behavior analysis has largely relied on traditional data acquisition methods with limited sample sizes that impede temporally and spatially fine-grained analysis (C. Chen et al., 2016). For example, these traditional data acquisition methods, such as surveys and diaries, always cannot provide enough information about precise spatial locations or timestamps when assessing local travel patterns. In other words, the resolution of the traditional survey data becomes insufficient to discover detailed travel information. If the resolution of the data can increase both in time and space, it will be of great use in explaining the travel behavior.

The advances in communication technologies have enabled researchers to collect travel data based on ubiquitous and location-aware smartphones, with massive GPS space-time data at a fine scale (Dabiri & Heaslip, 2018). Such individual-level GPS data has been found useful for analyzing trip characteristics, detecting human mobility patterns, as well as evaluating urban accessibility in recent studies (C. Chen et al., 2014, 2016; Marra et al., 2019; Reddy et al., 2008). In Chapter 2, I summarize these emerging big data's applications in travel behavior research. For example, phone-based GPS data can be used to determine a variety of trip characteristics, such as speed, travel distance, direction, travel mode, etc.

(2) Examining Accessibility

Furthermore, urban accessibility can also be assessed based on the trajectory data mining and analysis of individuals' GPS records (B. Y. Chen et al., 2018; Mora et al., 2017). Since the paramount goal of urban planning is to provide sufficient opportunities for residents to access various urban services, it becomes critical to allow residents to have good accessibility to these places for urban life viability. Evaluating accessibility of different regions and social groups has been the first important step for planning and management to provide equitable access (Neutens et al., 2010).

Accessibility has been evaluated by either "place-based measures", or "individual-based measures" (B. Y. Chen et al., 2017; Kwan & Weber, 2003). An example of the place-based measure and individual-based measure can be found in section 2.2.1 in Chapter 2. Briefly speaking, the place-based measures are conceptualized mainly in terms of locational proximity to individuals' residences. On the contrary, individual-based accessibility measures consider the individuals' activities and travel behaviors which has been widely used to evaluate accessibility of individuals in different geographical regions and socioeconomic groups. To easily distinguish these two accessibility measures, Chapter 2 describes that the theorized accessibility is theorized spatial accessibility, while realized accessibility is realized behavior.

Recent technical advancements make the phone-based GPS data as a new source with spatiotemporal information to help evaluate realized accessibility of individuals in different socioeconomic groups and geographical regions (Kwan & Weber, 2003; Miller, 2007). Precise information captured by the GPS traces allows to examine realized accessibility at fine-grained spatial and time scale (J. Chen et al., 2011; Neutens et al., 2010; Ren et al., 2014). For example, Chen et al. (2018) used massive mobile phone tracking data collected in Shenzhen, China to evaluate accessibility to urban services over different neighborhoods and social groups. They found accessibility of disadvantaged people was at a lower level than advantage people in general. Mora et al. (2017) evaluated accessibility based on GPS traces collected in the mobile devices of individuals with and without disability. They compared a series of parameters between these two groups, including the frequency of routes, and the distribution of the main origin and destination locations. Their findings showed the difficulties faced by disabilities to access urban amenities which calls for a better urban management to help disabled groups.

4.3 Smartphone-based GPS Data

The smartphone-based GPS trace data used in this investigation was collected by X-Mode Social, a location data industry partnering with over 300 app developers in the United States (X-Mode, 2021). Because I already examined transportation information using Yelp reviews in Phoenix in Chapter 3, in order to integrate the results with travel behavior information in next chapter, I select the same geographical area - Phoenix Metropolitan Area - for the GPS data collection.

First, I collect smartphone-based GPS trace data in Phoenix from October 1st to October 31st in 2018. The raw dataset consists of a sample of 297,406 individuals with 578,154,919 location records. I filter out nonresidents' data and obtain a sample of 92,032 residents' data. Residents were defined as individuals who had at least 20 days of GPS valid data[15] in the study area. After cleaning the dataset, I obtain 530,255,510 GPS location records of all the 92,032 residents.

The GPS data collection was through partner applications and relied on using modern

[15] To improve the accuracy of residents' identification, I particularly examined valid GPS location data recorded during nighttime.

smartphone devices' internal GPS hardware with high spatiotemporal resolution. Table 14 shows the variables in the GPS dataset, including a variety of individuals' location information, such as the anonymized individual ID, latitude, longitude, timestamp (in seconds), dwell time, dwell type, and speed (meters in seconds).

Table 14. Description of variables in the GPS dataset

Variable	**Description**
id	a unique id number for each individual
latitude	recorded latitude for the location
longitude	recorded longitude for the location
speed	recorded speed for the location (meters in seconds)
dwell time[1]	recorded the amount of time at the location (milliseconds)
dwell type[1]	estimated dwell type (high, medium, low, moving) for the location
timestamp	recorded timestamp for the location

Note: [1]"dwell time" and "dwell type" are calculated based on an algorithm developed by the location dataset provider. The dwell time calculation is contextual around the time period it is run and based on records which are contained in a point of interest (POI); the dwell type provides a description of the dwell time (high, medium, low, moving) and helps to determine a confidence in the POI and dwell time.

As a preliminary data examination, I count the number of daily GPS records for two randomly selected individuals. As shown in Figure 17, I find that individual traces may not be available for some days due to a lot of reasons, such as the battery running out, signal lost, or location service off, etc. For example, resident #1 (green line) did not have data on Oct 28, 2018. Also, both of their counts of GPS daily data were not stable. In some days they might be recorded for more than 200 GPS locations, while in some days they were recorded only for a small number of locations.

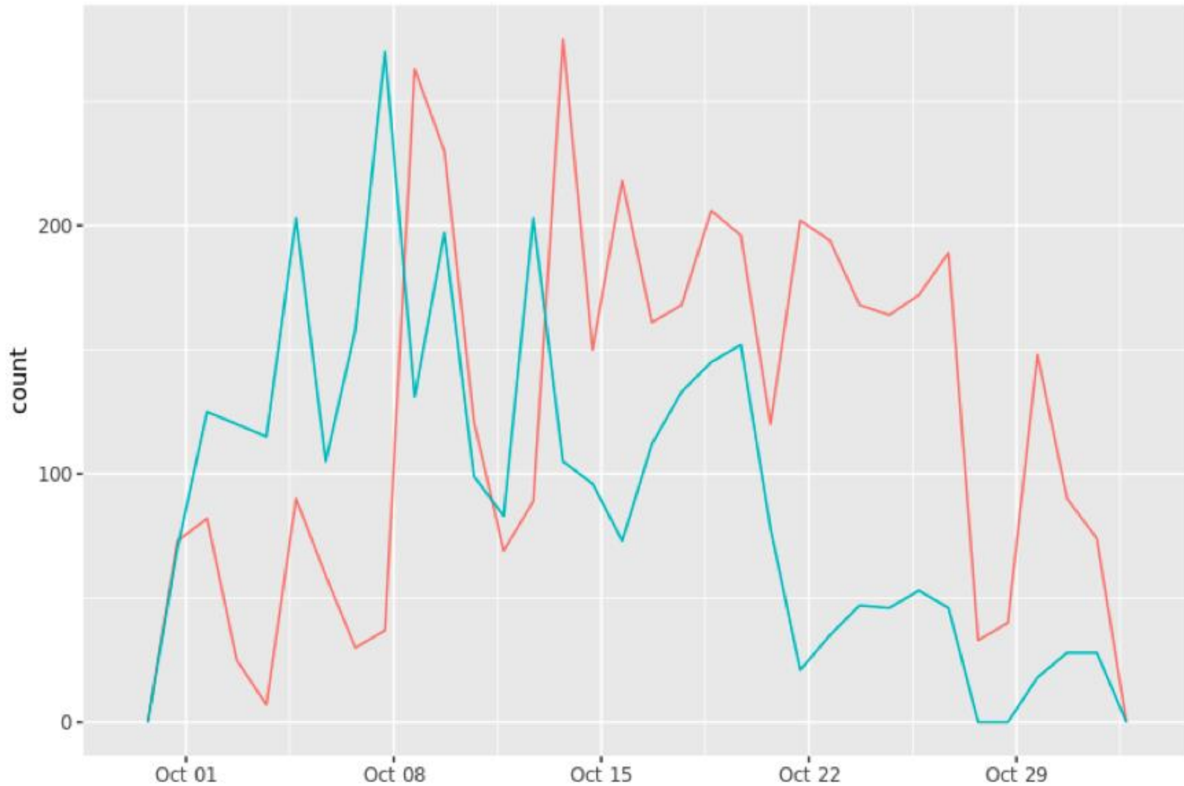

Figure 17. Histogram of two individuals' daily GPS record count

Note: Individual #1: the green line; Individual #2: the red line. Both of them are randomly selected individuals from the GPS dataset.

4.4 Methods

4.4.1 Estimating Home Location

First, I use each resident's GPS records during nighttime (12am to 6am) to estimate their home locations. As shown in **Error! Reference source not found.**, ten randomly selected r esidents' one-day (Oct 15, 2018) nighttime GPS records were visualized on the map. Some of them might still move during the nighttime, while a majority of them just stayed at one location and didn't travel as they had concentrated point clusters at night. Taking account all the 31 days' data, I estimate dwelling locations according to a density-based spatial clustering method (Mennis & Guo, 2009). Specifically, the centroid of the largest cluster within a 200-meter threshold was set as a resident's home location.

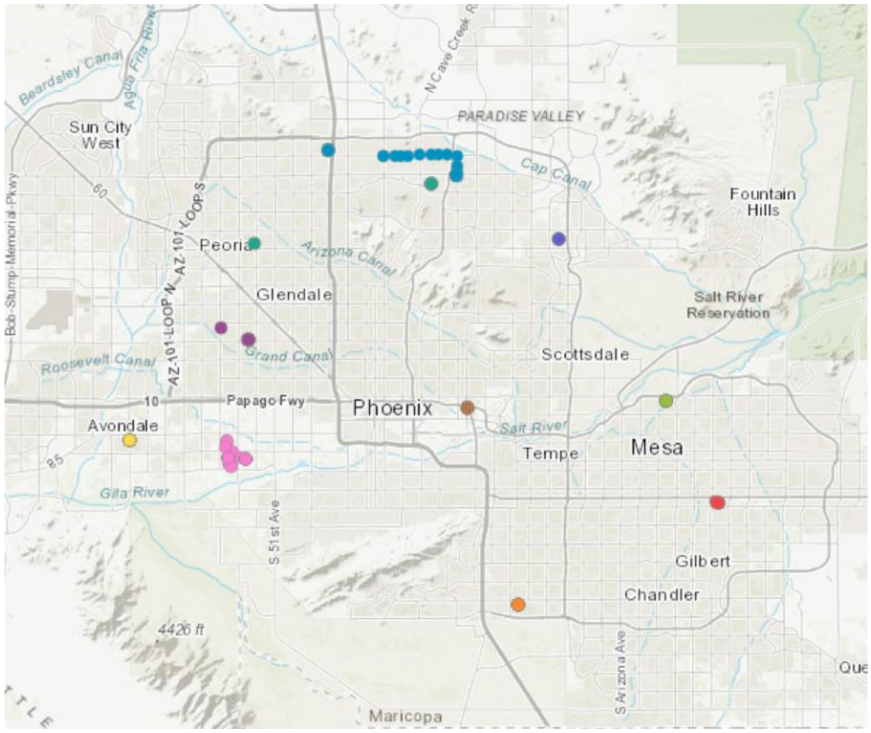

Figure 18. Map of ten residents' nighttime GPS records on Oct 15, 2018

Note: Ten different colors represent ten residents' nighttime location records. All of them are randomly selected from the GPS dataset.

After estimating the residents' home locations, I spatially join their home locations with the corresponding census block group (CBG) id (a 12-digit unique id). The CBG data was acquired from the 2014-2018 American Community Survey (ACS) data (U.S. Census Bureau, 2018), which provides detailed demographic, social, and economic characteristics for all of the block groups in the United States. This information allows me to evaluate accessibility level of individuals living in neighborhoods with varied demographical and socioeconomic groups in the following sections.

4.4.2 Examining the Representativeness of GPS Dataset

Although the vast majority of Americans - 96% - owned a cellphone, and the share of Americans that own smartphones was 81% (~ 300 million population) in 2019 (see Figure 7), it is still necessary to examine the potential bias issue (the representativeness) associated with the collected GPS data.

The location dataset contained over 90,000 Phoenix residents' GPS traces for an entire month, however, it is still a subset of the whole population. Usually, researchers use statistical sampling strategies to assign an appropriate weight to each GPS resident, which is illustrated as a method to adjust for selection bias and can help to make the estimations more representative (Garber et al., 2019). Thus, I account for weighting functions in this analysis.

First, I calculate the number of census population in each census block group (CBG) based on American Community Survey data (U.S. Census Bureau, 2018). Then, I perform a correlation analysis between the estimated number of GPS residents and the census population in each CBG to preliminarily examine whether the collected GPS dataset had a significant spatial bias in Phoenix. The paired Pearson Correlation result was 0.70, which suggested that the two count numbers were strongly correlated with each other at CBG level. Thus, this spatial correlation analysis indicated that the GPS data had a similar spatial distribution with the census data overall[16].

In order to further match the GPS sample's population with the census population, I calculate a weight value for each GPS resident based on the following equation, which is similar to an inverse-probability-of-selection weighting method that has been widely used to adjust for selection bias (Griffin & Jiao, 2015; Heesch & Langdon, 2016).

---

[16] Echoing this analysis, I also perform a Moran's I spatial autocorrelation (Moran, 1950) for these two groups (the GPS sample and the census population) to compare their spatial differences.

$$W_{ij} = \frac{P_j}{P'_j} * \frac{\sum P'}{\sum P}$$

where W_{ij} refers to the weight of resident i who is from CBG j. P_j and P'_j represent the census population and the GPS population in CBG j, respectively. $\sum P'$ and $\sum P$represent the total GPS population and the total census population of all the CBGs. I consider the weight in the analyses of the following sections in order to better interpret the results based on a representative sample.

4.4.3 Frequency of Visiting Commercial Districts by Cars

In this section, I calculate the frequency of visiting commercial districts by cars using the GPS location dataset. The methodology framework is shown in Figure 19. First, I identify travel mode to seventeen commercial districts which are selected to represent places with concentrated commercial activities in the study area. More details about the identification of these seventeen districts can be found in section 3.4.2 in Chapter 3.

Then, I calculate the number of visits by cars in each district using residents' GPS traces. I select the top six commercial districts as major districts (see Figure 10). Following that, I calculate the average number of car trips per person to the six districts during the one-month period (see Table 15). Phoenix residents make an average of 3.9 trips in weekend evenings, 2.5 trips in weekend days, and 2 trips in weekend evenings in October, 2018. Then, I summarize the total number of visits by cars to each of the six major district in three different time periods: weekday evenings, weekend days, and weekend evenings[17].

---

[17] Time of day traveling to commercial districts: weekday evening (5pm - 10pm); weekend day (9am - 5 pm); or weekend evening (5pm - 10pm).

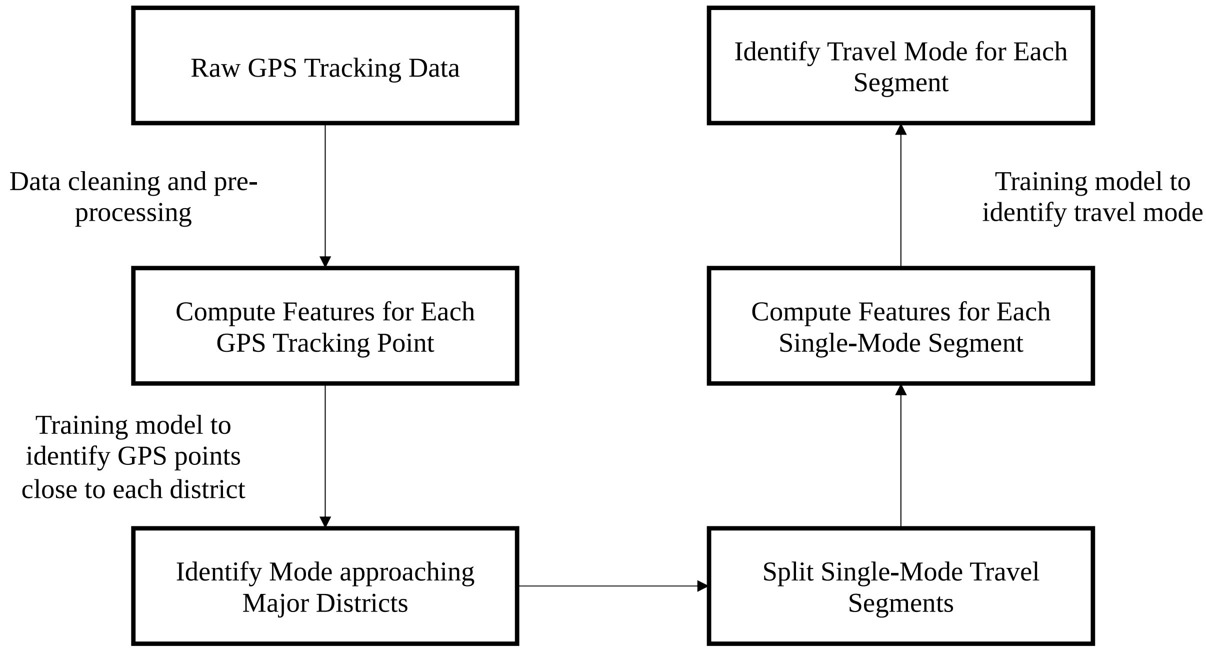

Figure 19. methodology framework of travel mode analysis using GPS traces

Table 15. Basic statistics of trips to selected commercial districts

Time of Travel	**Average Trips[1] Per Person (One Month)**
Weekday Evenings	3.9
Weekend Days	2.5
Weekend Evenings	2.0

Note: [1]These are trips by cars to six major commercial districts.

4.4.4 Measuring Accessibility to Major Commercial Districts

(1) Examining Theorized Accessibility

As described in section 4.2.2, the theorized accessibility is theorized spatial accessibility, according to place-based measures. In this section, I specify that each resident's theorized accessibility to a commercial district equals to 1 divided by the distance d_{odi} between residence and destination district's centroid. The closer of the resident's home location is to a district D_i, the larger of the value is, which suggests a higher level of access.

Furthermore, I aggregate individual-level accessibility values to district-level values. For a district D_i, I hypothesize that all the residents have travel needs to go to D_i. Then, I use the socio-demographical variables at CBG level (median household income and percentage of white, Latinx, black, Asian) to multiply the theorized value (1/distance) to calculate a summation socio-demographics at D_i. Later on, the summation divided by the total number of residents gives an average value, which can represent the mean socio-demographics at D_i based on theorized accessibility measures. After this calculation, for example, for district one D_1, it will have a set of variables that describe "theorized travelers" to D_1. Specifically, these variables include theorized travelers' average median household income and percentage of white, Latinx, black, and Asian.

(2) Examining Realized Accessibility

It is noticeable that the set of each district's average socio-demographic characteristics calculated based on theorized accessibility shown in the above section only can represent facts under a theoretical assumption: all residents (theorized travelers) have inverse-distance probabilities to go for commercial and mixed-use activities in Phoenix. However, in reality, people choose their travel destinations because of a variety of factors, such as the attractiveness of a destination, personal preferences (perceived opportunities), availability of travel mode, cost of travel, and so forth.

Thus, based on travel behaviors examined from the GPS data, I calculate a set of socio-demographic characteristics of each district based on realized accessibility. In this calculation, I only consider residents who actually visited the selected districts. In addition, since the GPS data provides travelers' time of non-work travel, I include temporal dimensions in the calculation as well. As a result, each district has three sets of socio-demographic characteristics[18] in weekday evenings, weekend days, and weekend evenings, respectively.

(3) Comparing the Difference between the Theorized and Realized Accessibility

After step (1) and step (2)'s calculations, I have four sets of district-level socio-demographic characteristics based on theorized accessibility and realized accessibility, respectively. Because all the calculated socio-demographic characteristics represent the group means, a Pearson Chi-square test (Pearson, 1900) is used to show whether the mean values distribute statistically significant different across groups.

(4) Characterizing Neighborhoods and Examining Non-work Travel Disparities over Different Neighborhoods

To categorize more characteristics of the Phoenix neighborhoods, I retrieve more variables from the census data at CBG level after the spatial join between home locations and their corresponding CBGs. Specifically, I label all the CBGs with these categories: Poor/Nonpoor and Race/Ethnicity (white majority, Latinx majority, black majority, or Asian majority). Poor CBG is defined as over 30% of the CBG population under the poverty line. The race/ethnicity category is labelled based on the dominant racial group, which has the largest racial percentage in the CBG.

To examine non-work travel disparities over different neighborhoods, I select travel distance as a variable and compare how different in travel distance of each neighborhood category when going for non-work activities.

---

[18] These variables are the same as the step (1)'s calculation, including "realized travelers" average median household income and percentage of white, Latinx, black, and Asian.

4.5 Results

4.5.1 Home Locations

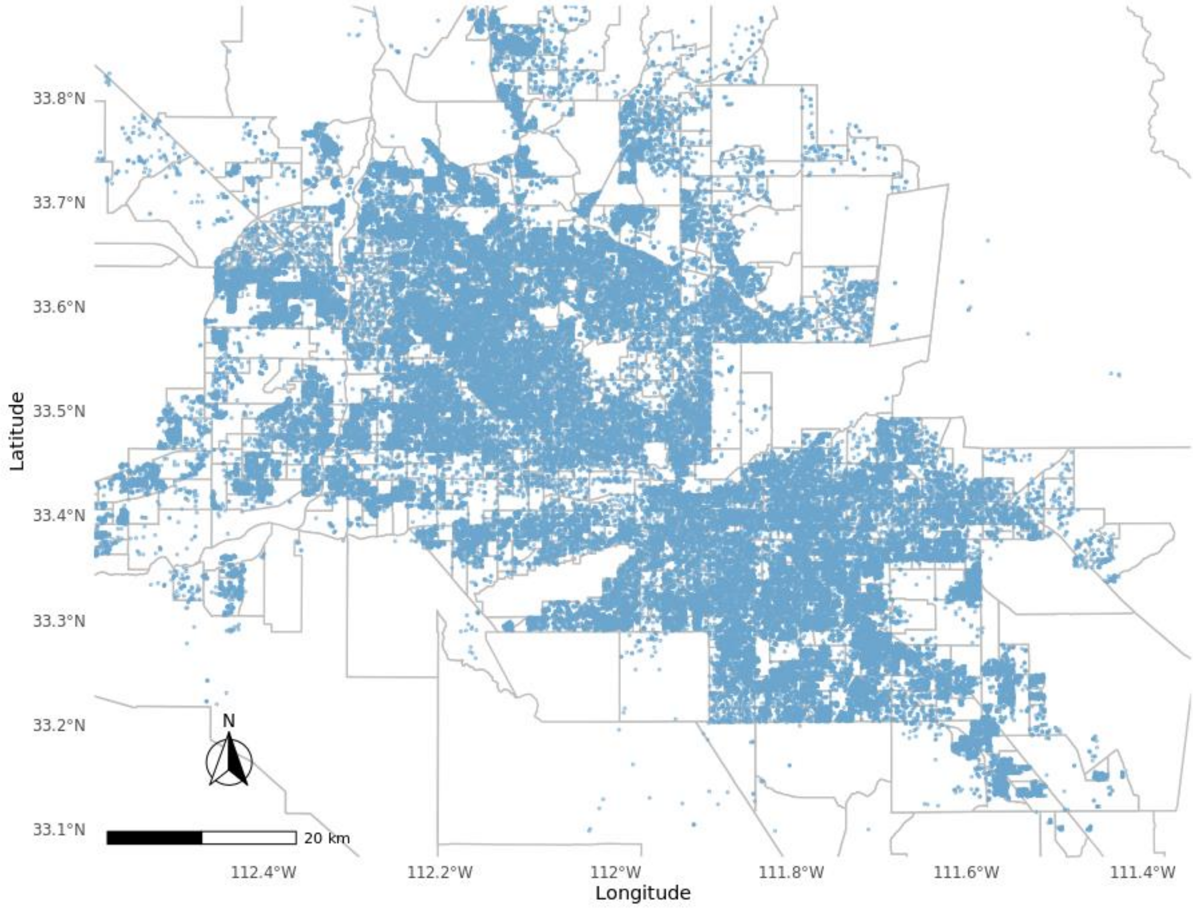

Figure 20. Map of the estimated home locations of GPS residents

Figure 20 shows the identified 92,032 residents' home locations (blue dots) of 2,564 GBGs in the study area. After spatially joining them to the geographical block groups and associated census dataset (U.S. Census Bureau, 2018), each resident has a series of their dwelling CBG's demographic and socioeconomic information.

4.5.2 Divergent Non-work Travel Behaviors

Figure 21 shows a visualization example of mapping one-day's (Oct 15, 2018) GPS traces of three residents randomly selected from the GPS dataset. It also provides a proof of evidence showing how individual locations are recorded in the smartphone-based GPS dataset. As shown in Figure 21, the travel path of resident #1 (blue dot) was between northern rural areas and Downtown Phoenix, along with the Interstate 17 Highway. Resident #2 (green dot) traveled between the suburban Paradise Valley to the Tempe-Mesa region, with concentrations of points recorded in Paradise Valley and a neighborhood Eastern Mesa. In contrast, resident #3 (red dot) almost demonstrated a much smaller travel trajectory compared with the other two residents - with only moving around several neighborhoods in Chandler.

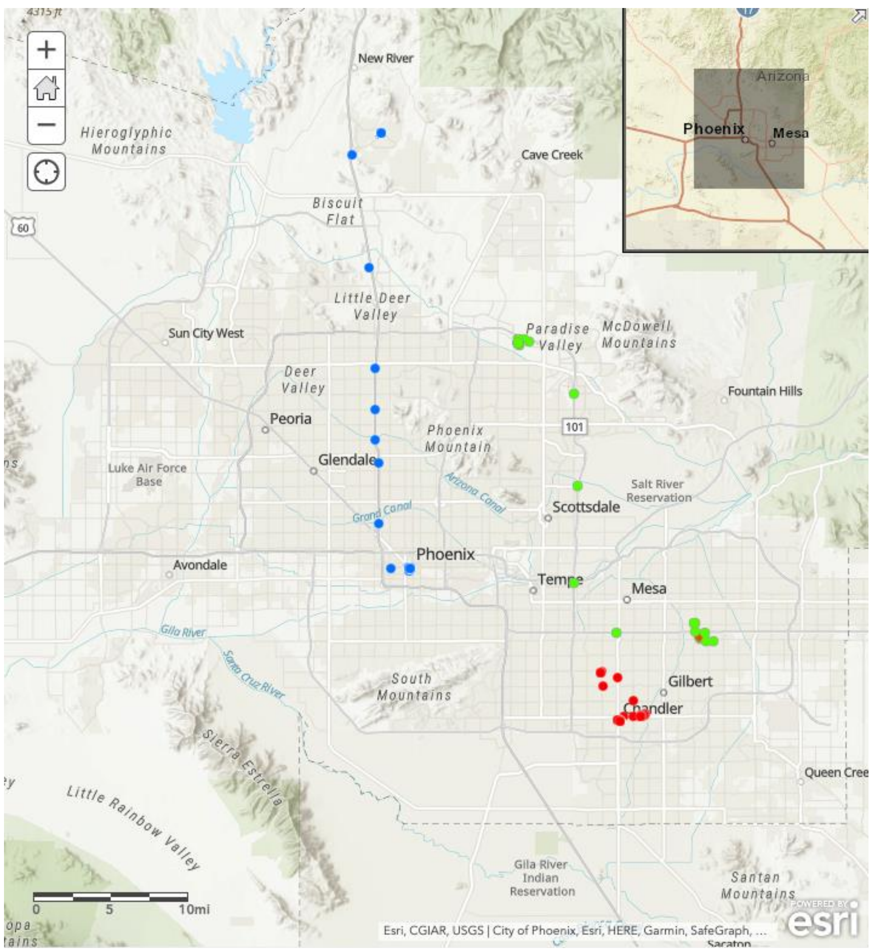

Figure 21. Map of GPS traces for three residents on Oct 15, 2018

Note: Three different colors represent three residents' traces. All of them are randomly selected from the GPS dataset. This visualization is conducted through ESRI ArcGIS Online application.

Although the above example only shows one-day GPS records of three residents randomly selected from the total 92,032 residents, it indicates the divergent nature of human mobility. If I geocode more residents' GPS records on map, how would it be? Figure 22 shows forty randomly selected residents' travel trajectories on Oct 15, 2018. These trajectory lines are generated by connecting each consecutive GPS locations recorded in the dataset using R package *trajr* (McLean & Skowron Volponi, 2018). The bolder lines represent an overlap of multiple lines in the area, suggesting a larger number of human mobilities there. Because I have such a big dataset of over 90,000 individuals which is almost impossible to map all their traces, I only use Figure 21 and Figure 22 as two demonstration examples to indicate the divergent human mobilities captured in GPS traces.

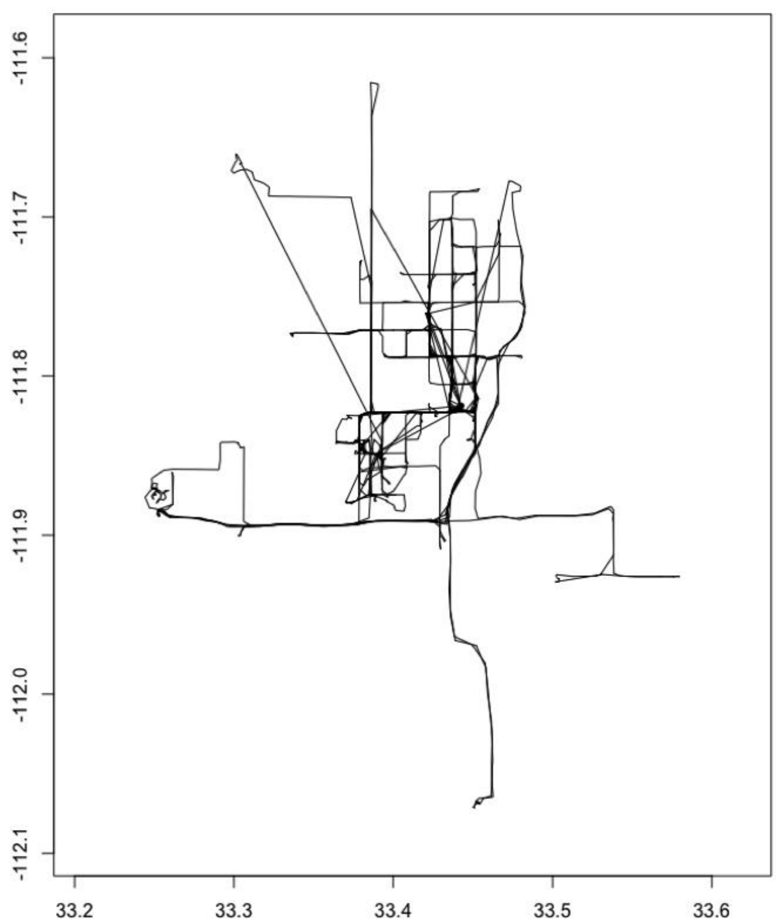

Figure 22. Visualization of travel trajectories of forty residents on Oct 15, 2018

4.5.3 Socio-demographic Characteristics in Each District

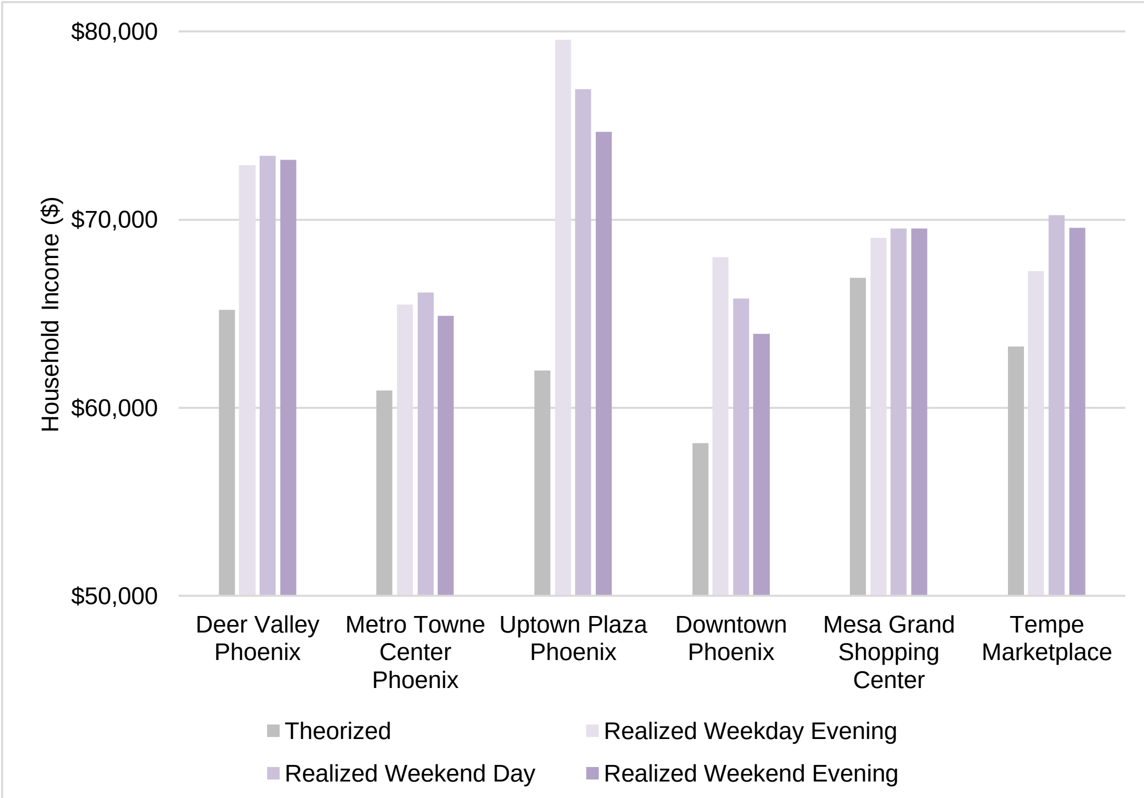

Figure 23. Median household income in each commercial district

Figure 23 shows the comparison results between the theorized median household income and realized household income with three periods of time. The grey columns represent the theorized values, and the purple columns are the realized values in weekday evenings, weekend days, and weekend evenings.

The realized income is higher than the theorized income across the districts and time periods overall. Particularly, Downtown Phoenix has the largest difference between theorized income and realized income during weekday evenings. Its theorized average median income is below $65,000, while its realized value is up to $80,000 in weekday evenings. Similarly, Deer Valley Phoenix has the similar comparative finding, but it has a consistency in its realized income across the travel time windows, which suggests that this district is always visited by the similar income groups regardless of time. On the contrary, Downtown Phoenix attracts more lower income visitors in weekends than weekday evenings. Based on the theorized accessibility, it is expected to have more lower income people going to these commercial destinations, however, travelers' behaviors indicate that more higher income people travel to non-work destinations.

Figure 24 - Figure 27 show the race and ethnicity comparison results, as indicated by the racial groups' percentages. The grey columns still represent the theorized values of six major districts, and the purple columns are the realized values in weekday evenings, weekend days, and weekend evenings, respectively. The comparison results are tested by a Pearson chi-square test method, and all of the target districts have significant p-values, which suggests strong variations in their racial groups' distributions based on two types of accessibility measures.

In Figure 24, the results show that there are much higher percentages (about 10% more) of white people go to Deer Valley and Uptown Phoenix than their theorized percentages. The realized percentages of white customers in Mero Towne Center and Mesa Grand Shopping Center are similar with the theorized values, which suggests that the actual visits by white residents are consistent with the planned visits. Downtown has the smallest percentage of white customers based on theorized accessibility, while its realized values are slightly higher, especially in weekday evenings (about 5% higher), which has more white visitors than weekends. Tempe Marketplace has an average of about 5% more white visitors than the theorized percentage.

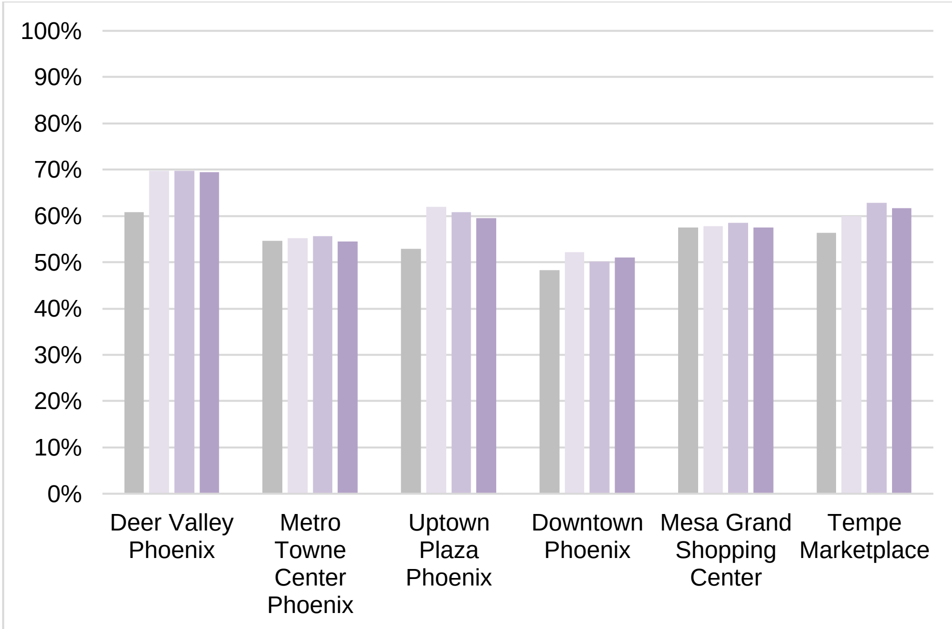

Figure 24. Percentage of White in each commercial district

Figure 25. Percentage of Latinx in each commercial district

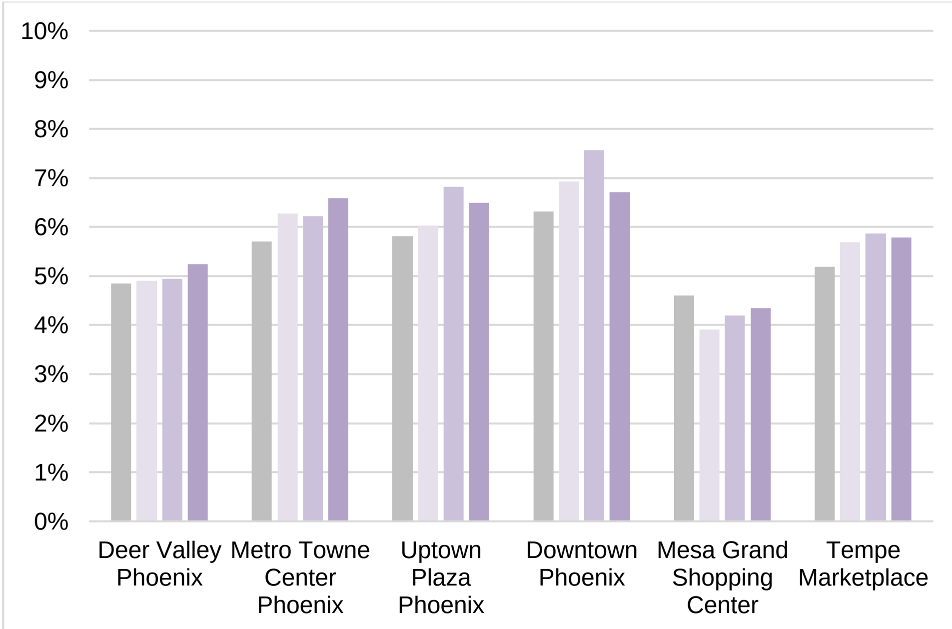

Figure 26. Percentage of Black in each commercial district

Note: This figure's scale is only from 0% to 10% due to the low value of this variable.

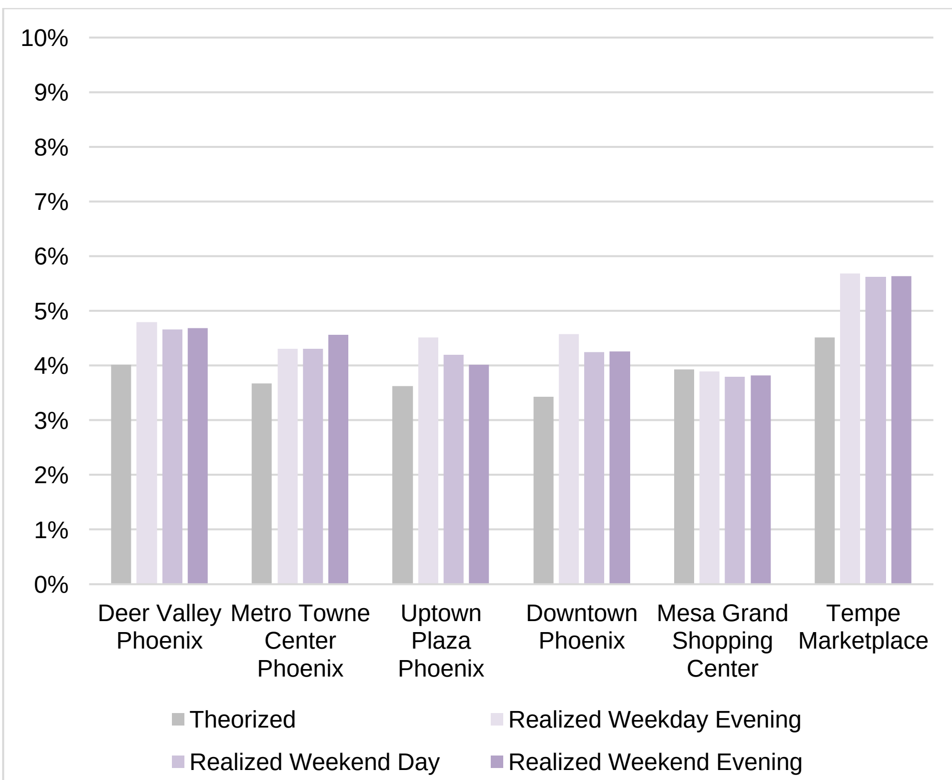

Figure 27. Percentage of Asian in each commercial district

Note: This figure's scale is only from 0% to 10% due to the low value of this variable.

Figure 25 shows the Latinx comparison results. Interestingly, significant lower percentages of Latinx people go to Downtown during weekday evenings and weekend evenings. Nevertheless, during weekend days, its realized number of Latinx visitors is as expected as the theorized number. For most districts, the realized values during weekend evenings are much larger than the theorized values, which suggests that more Latinx choose to go to commercial districts (except Downtown) during weekend evenings.

Figure 26 represent the results for black group. I find that Mesa Grand Shopping Center has a lower black population percentage than the theorized percentage, suggesting that it has the least attractiveness to black population. Most black members chose to go to other districts, especially Downtown, with a highest percentage during weekend days, for shopping, dining, and other commercial activities.

As shown in Figure 27, except Mesa Grand Shopping Center, a slightly bit more Asians visit other five districts than expected. In addition, I find that Downtown has the smallest theorized percentage of Asian compared to other districts, suggesting that Asian people have the lowest level of access to Downtown given by the place-based accessibility measures. Because this theorized accessibility is calculated based on Asian groups' dwelling locations, its lower value implies that Asian groups live farther from Downtown Phoenix.

4.5.4 Social Disparities in Accessing Major Districts over Different Neighborhoods

As shown in Figure 28 and Figure 29, I calculate travel distances of each type of neighborhoods to the six major destination districts based on the GPS data. In this section, the distance of travel to each district is calculated by the average estimated travel distances[19] of the GPS residents. Figure 28 shows comparison results between the poor and nonpoor neighborhoods. Green columns represent the poor group while red columns represent the nonpoor group. Surprisingly, the results indicate that only when the destination district is Deer Valley Phoenix, residents from poor neighborhoods experience a longer distance of travel by cars compared to residents of nonpoor neighborhoods.

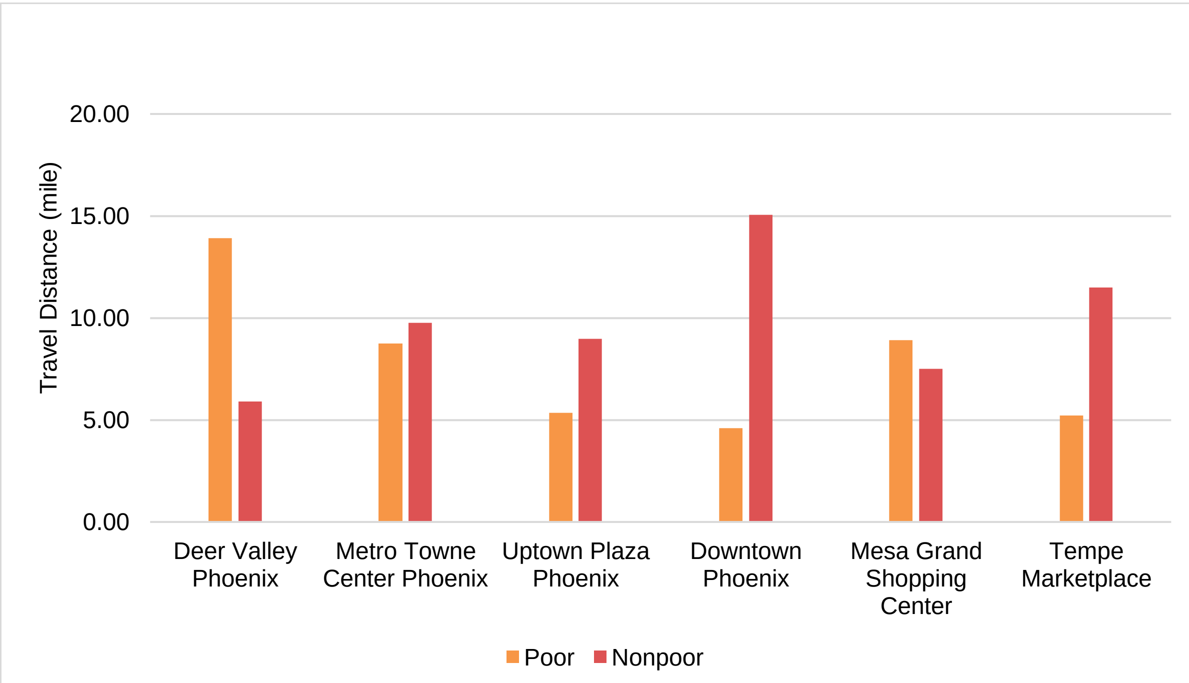

Figure 28. Travel distance across poor/nonpoor neighborhoods

Previous studies find that residents from disadvantage / vulnerable neighborhoods, such as neighborhoods with high poverty rates, probably need to use more travel time (as well as long trip distance) to access urban services (Grengs, 2015; Stanley et al., 2011; Wee et al., 2011). These studies attribute the longer travel time as a result of these disadvantaged residents

[19] Each GPS resident's travel distance is calculated based on the road network distance between the estimated home and the destination district centroid.

need to use public transit which takes more time. Although the results only focus on automobile travel, they demonstrate that individuals from different types of neighborhoods still experience different travel time.

Findings of the empirical analysis indicate that when Phoenix residents from nonpoor neighborhoods choose to go to commercial districts such as Downtown Phoenix, Uptown Phoenix, or Tempe marketplace, they have about ten miles longer driving distance than residents from poor neighborhoods. In addition, there is no significant difference in travel distance to Metro Towne Center Phoenix and Mesa Grand Shopping Center for residents from poor or nonpoor neighborhoods.

The results also reveal different auto travel costs across the racial groups. As shown in Figure 29, residents from white and Asian neighborhoods have shorter travel distance - less than ten miles of travel - for most of the districts than their black and Latinx counterparts. Intriguingly, although residents from black neighborhoods take more time to go to a majority of the six commercial districts (four of six), they actually have higher level of auto accessibility to Uptown and Downtown Phoenix. Findings also demonstrate that black neighborhoods have worse accessibility to the Mesa Grand Shopping Center in particular, where they have to take the longest travel distance (more than twenty miles). Latinx neighborhoods have a travel distance range between ten to fifteen miles when visiting a majority of the six districts and have slightly better accessibility than black neighborhoods yet still less accessible (more than five miles longer in travel distance) than white and Asian neighborhoods.

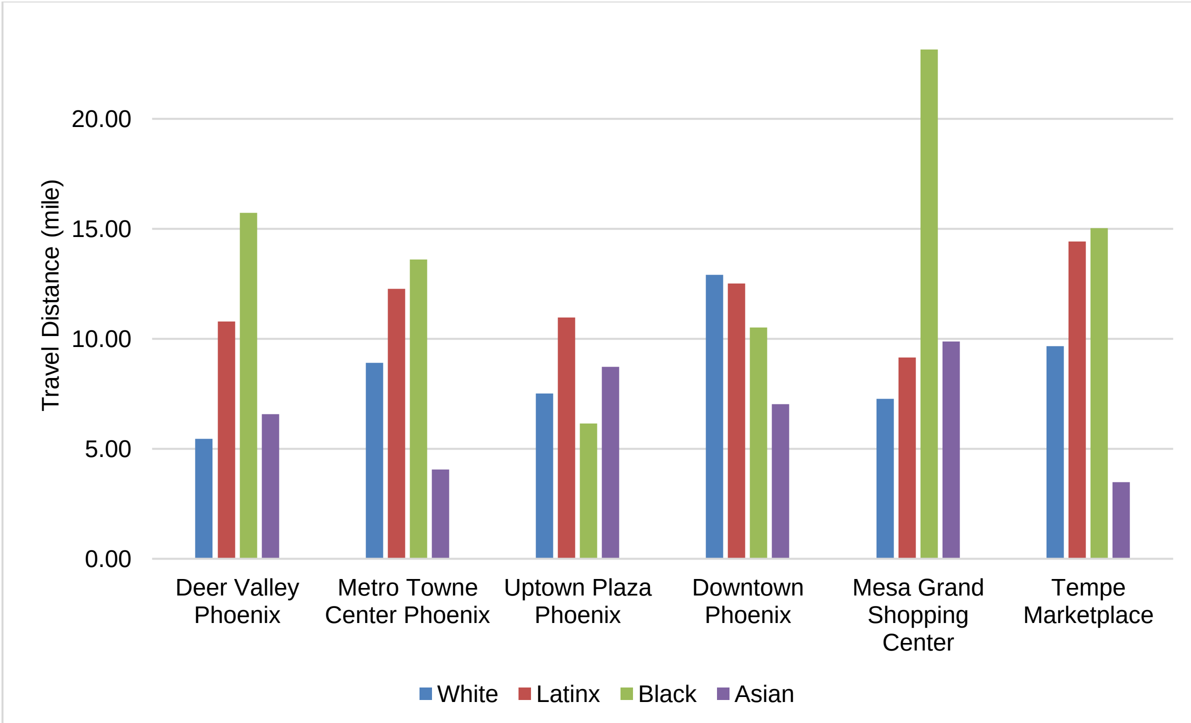

Figure 29. Travel distance across different racial neighborhoods

4.6 Limitations and Conclusions

4.6.1 Discussions

This chapter examines the residents' travel to non-work destinations using a massive mobile phone GPS tracking dataset including more than 90,000 residents in Phoenix, AZ. The results show that Phoenix residents frequently drive to six major commercial and mixed-use districts for non-work activities during weekday evenings, weekend days, and weekend evenings. This finding is aligned with previous studies suggesting that human mobility is limited to a small number of destination locations (Alessandretti et al., 2018; Song et al., 2010). In particular, for non-work trips, most people go to city centers for such activities (Li et al., 2018). This phenomenon is further confirmed by the empirical analysis using GPS data: the six frequently visited commercial districts do contain city centers (e.g., Phoenix Uptown and Downtown) in the region.

In addition, I find that non-work travel disparities exist in income and race and ethnicity, associated with different time and destinations. Individuals from different types of neighborhoods (poor/non-poor, race/ethnicity) experience different distance of driving.

4.6.2 Limitations

The analysis has some limitations as well. First, even though the vast majority of Americans (96%) owned a cellphone, and the share of smartphone ownership was up to 81% in 2019 (see Figure 7), it is still necessary to examine the potential bias issue (the representativeness) associated with the collected GPS data. The location dataset collected over 90,000 Phoenix residents' GPS traces for an entire month, but it is still a subset of the whole population. In addition, individual demographic and socioeconomic characteristics are not included in big GPS dataset, which is a common limitation of GPS data with large sample size (Çolak et al., 2015). Future research may need to include conducting a survey for some GPS participants to better address the representativeness issues associated with the GPS data - - an issue more broadly associated with all other self-selecting big data. Combining survey information with GPS data in future research can help to have to better understanding of travel behaviors.

Second, I describe the variables recorded in the GPS dataset in Table 14. After a close analysis of the timestamps from the GPS data, I find that the locations are recorded at five-to-ten-minute interval, which suggests that the GPS data don't have very high temporal resolution[20]. Although the analysis of driving behaviors in this study may not be affected by

[20] In Chapter 2, I find that some GPS data may have very high temporal resolution (less than 30s), which can help to analyze travel patterns with more accurate estimations.

this limitation, future research needs to consider to use transportation big data with more features, better structures, and higher resolutions.

In addition, this analysis only focuses on Phoenix and selects six districts of interest - where have intensive non-work activities such as shopping, dining, leisure, etc. Future research would benefit from examining additional case study cities and metropolitan areas as well as from examining other destinations of interest, such as health care, education, and other urban services.

4.6.3 Conclusions

This chapter examines residents' travel to non-work destinations using a massive mobile phone GPS tracking dataset including more than 90,000 residents in Phoenix, AZ. The investigation in this chapter demonstrates that GPS traces allow individual travel behaviors to be evaluated with fine-grained enough spatiotemporal resolutions. The analytical results illustrate that massive GPS location data could be a very useful data source for large-scale, travel behavioral pattern analysis and studies (Kwan & Weber, 2003; Mennis & Guo, 2009; Xia et al., 2019).

This investigation seeks to fill two gaps in previous studies. First, it extends research on non-work travel behavior in a large metropolitan area, providing new insights on the divergent travel behavioral patterns when going to non-work places, such as shopping centers, restaurants, bars, supermarkets, grocery stores, movie theaters etc. Results have implications in transport policy planning and design for creating equitable non-work travel access in urban environments.

In addition, it also allows the measures of travel to non-work destinations with temporal dimensions by evaluating residents' non-work travel by automobile in three time periods of interest: weekday evenings, weekend days, and weekend evenings. These three time periods are the main time periods that residents travel to non-work destinations. Results of this study will enrich our understanding of travel to non-work destinations using emerging transportation big data.

CHAPTER 5

EMPIRICAL ANALYSIS 3: INTEGRATION OF MULTIPLE TYPES OF DATA

5.1 Introduction

For decades, researchers have been studying the influence of the built environment on travel behavior. Most of these studies have found an association between people's residential location and their travel mode choice (for an overview, see Ewing & Cervero, 2001, 2010). Recent studies assume that travel behavior is guided by attitudes, especially since the publication of Ajzen's Theory of Planned Behaviour (Ajzen, 1991). Attitudes are usually defined as the degree to which the evaluation of a certain object, person or behavior is favorable or unfavorable (Van Acker et al., 2010).

A majority of existing studies found that travel-related attitudes have an important effect on travel behavior, such as the positive effects of mode-specific attitudes on the choice for that mode. Some studies even claim that travel attitudes have a stronger impact on travel behavior than the built environment does (for example, Bagley & Mokhtarian, 2002; Handy et al., 2005). In addition, as shown in the conceptual framework in Chapter 1, travel attitudes may have an impact on the built environment as well. In this sense, the direct effect of the built environment on travel behavior might be overestimated as attitudes partly explain the impact of the built environment on travel behavior.

In Chapter 3, the Yelp reviews mentioning parking, whether positive or negative in sentiment, may be an indicator that parking is a concern in the area. I use parking sentiment describing whether the traveler is satisfied with the parking environment when driving to major commercial districts as a measure of travel attitude. In Chapter 4, I use smartphone-based GPS data to examine travel to major commercial districts (travel behaviors). Based on previous travel behavior studies, it can be argued that travel attitude, behavior, and built environment are interlinked as shown in the conceptual framework in Chapter 1 (see Figure 2). It is possible that 1) travel attitude is associated with the built environment; and 2) travel attitude, built environment, and socio-demographics may have an influence on travel behavior.

Building on the literature, in this chapter, I design two data integrations. First, I examine the relationship between attitude and built environment by integrating parking attitudes and parking supply data, as shown in Figure 30 (a). Then, I study how attitude and built environment affect travel behavior by integrating geosocial media and GPS trace data, as shown in Figure 30 (b). Both of these two data integration frameworks make use of emerging transportation big data to study non-work travel.

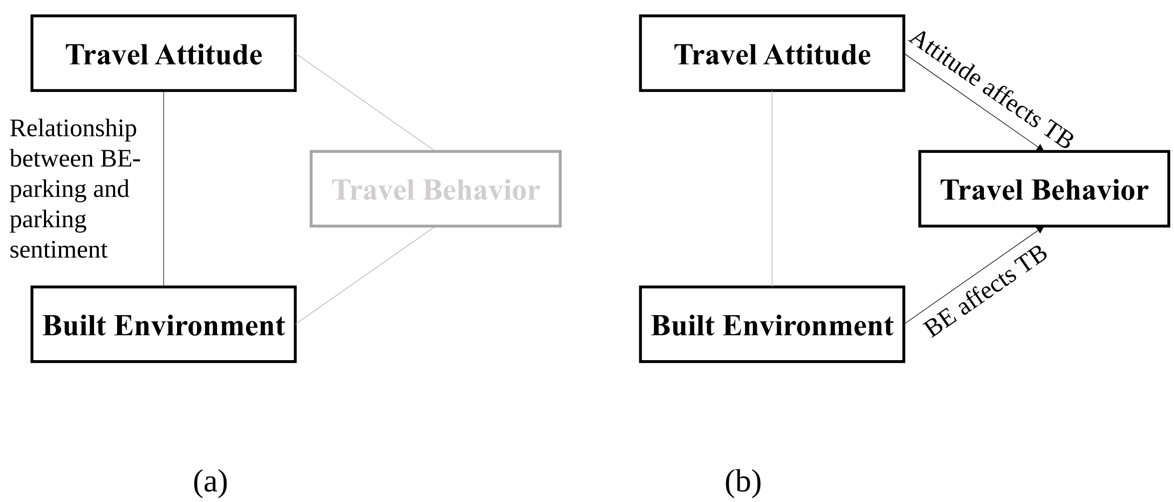

Figure 30. Data integration frameworks

Note: Built environment is short for "BE" and travel behavior is short for "TB" in this figure.

5.2 Relationship between Built Environment Parking Supply and Parking Sentiment

In this section, I examine whether parking supply modifies parking sentiment. In particular, I investigate how parking sentiment is associated with the provision of parking, using content from Yelp online reviews, a large geosocial media dataset for Phoenix.

5.2.1 Background

For transportation planners, an equally important relationship is that between transportation experiences and attitudes towards specific planning interventions. Support for road building, for example, is associated with more driving (Börjesson et al., 2015) and increasing regional congestion (Rose, 1990). Parking management strategies, such as pricing and parking maximums, also elicit public, political responses that can make or break a plan or policy (King et al., 2007; Mondschein et al., 2020). Therefore, a better understanding of the relationships between positive and negative attitudes towards parking and factors such as parking management strategies may inform planners seeking to foster multimodal, sustainable transportation and urbanization as well as help shed light not just on the behavioral effects of those strategies but their political feasibility. In addition, this approach can identify general best practices for parking management strategies and land-use approaches, as well as local variations in sentiment that can be used to identify specific issues or unexplained areas of positive or negative parking experiences.

5.2.2 Data and Research Question

As I describe in Chapter 3, most businesses in Yelp dataset have associated parking attributes, a set of binary categories (True/False) indicating the availability of five parking attributes at each business, such as “parking garage,” “parking lot,” “street parking,” “parking valet," or "validated parking." The parking attributes provide a means of ground-truthing the type of parking supplied in different commercial and mixed-use districts across the study area, though they do not indicate the absolute quantity of parking supplied (e.g., the number of parking spaces). In addition, they represent specific strategies used by businesses, business collectives such as business improvement districts, and planners to more effectively manage parking supplies in commercial and mixed-use districts.

In this section, I examine major commercial and mixed-use districts in Phoenix for answering the key empirical research question: *How do business parking management strategies shape parking sentiments?*

5.2.3 Business-level Parking Sentiments and Hotspot Analysis

(1) Parking Sentiment Score for Each Business

Using the average sentiment score for each business, the distribution of positive, neutral and negative reviews is similar across the study area (see Figure 14). I found that positive sentiments are the majority across the region. The results are consistent with the count values of sentiment classification outputs as well. These sentiment scores are normalized by the analysis model ranging from -1 to 1[21].

(2) Spatial Hotspot Analysis of Parking Sentiment in Phoenix

Geographic Information System (GIS) analysis enables a hotspot analysis of the spatial pattern of parking sentiments, in terms of the sentiments distribution of business itself and its surrounding businesses. I use the Getis-Ord Gi* statistic (Ord & Getis, 2010), a spatial

[21] If the sentiment score equals to 0, it refers to a neutral sentiment; if the score is from -1 to 0, it refers to a negative parking sentiment; if the score is from 0-1, it refers to a positive sentiment.

statistical approach, to determine the clustering pattern of parking sentiments. Getis-Ord Gi* finds where high and low sentiment ratios cluster spatially. The GIS Gi* statistic is estimated for each business with a z-score. The larger the z-score is, the more intense the hot spot clusters of high sentiment scores. The smaller the z-score is, the more intense the cold clusters of low sentiment scores.

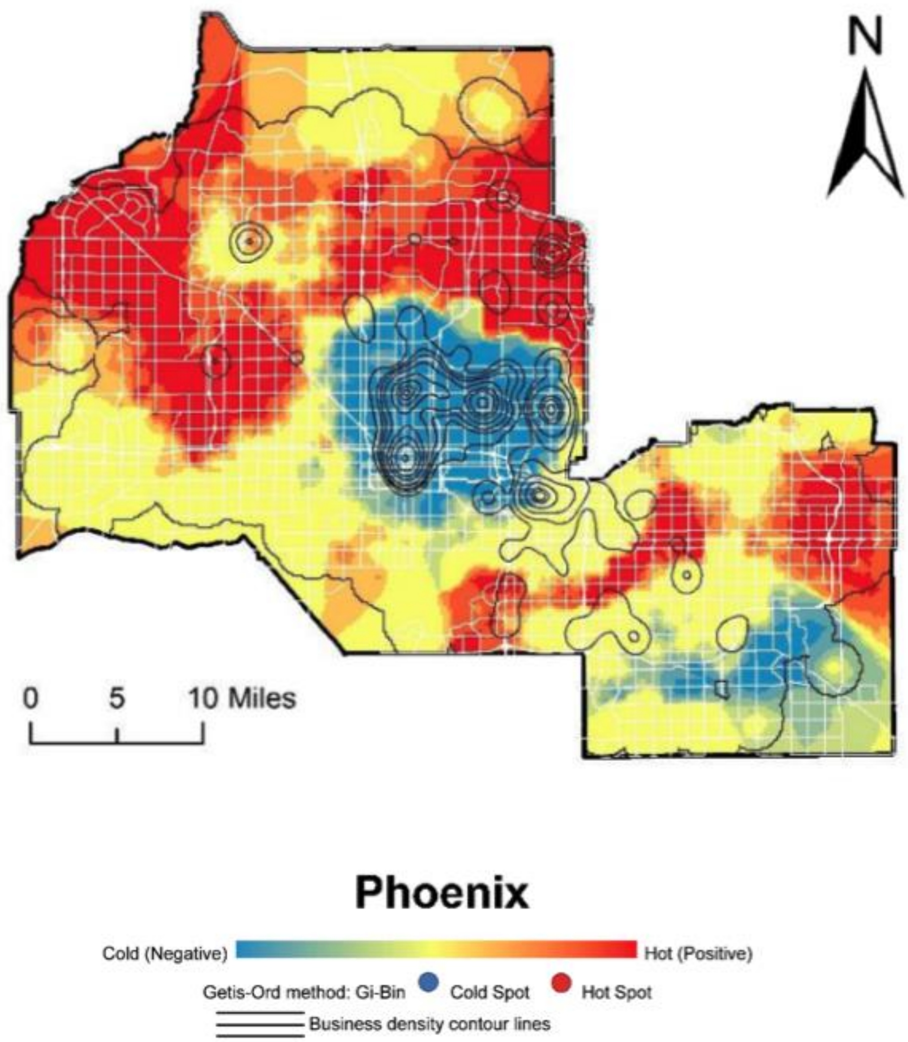

Figure 31. Spatial heatmap of “positive”, “neutral” and “negative” parking sentiments in Phoenix, AZ

In order to have a better visualization, after obtaining the z-score of each business, I use Inverse Distance Weighted interpolation (ESRI, 2018) to map the clustering patterns from Gi* hot and cold spots, shown in Figure 31. In addition, business density is also illustrated using contour lines for visual comparison, ranging from 0 to 200 businesses per sqkm, in increments of 5 businesses per sqkm (see Figure 31). Overall, the map shows distinct patterns, and negative sentiment clusters are clearly associated with the central business districts in the study area, evident from the densest business clusters. However, not all business districts are clusters of negative parking sentiment. Suburban commercial areas, such as those on the north side of Phoenix actually show clusters of positive parking sentiment.

5.2.4 Analysis of the Effect of the Provision of Parking on Parking Experiences

I use generalized linear mixed-effect (GLME) models to further evaluate how the provision of parking affects parking sentiment across districts and activity types (Mcculloch & Neuhaus, 2001; Zhang et al., 2016). A GLME model is an extension of classical linear regression models. The standard form of a GLME model is:

$$y_i|b \sim Distr(\mu_i, \frac{\sigma^2}{\omega_i})$$

$$g(\mu) = X\beta + Zb + \delta$$

where y is the response variable, the sampling unit i in the model represents the ith business (each business has a unique *business id*). The ith response variable y_i corresponds to the averaged sentiment score for this business. In this case, the response variable is the averaged sentiment score of a business, showing an overall parking experience of the parking environment of a business. β is fixed-effects, representing parking management strategies, and several built environment factors. b is random-effects, which is associated with individual experimental units drawn at random from the population and account for variations between groups. In this case, the random-effects variable refers to the destination activity types (the category of the business). $Distr$ is the distribution of y given by b, which assumes the distribution of the response variable conditioned on the random-effects variable belongs to the exponential family (Mcculloch & Neuhaus, 2001). μ is the mean of y given by b, σ^2 denotes the dispersion parameter, and ω_i represents the weight for observation i. g denotes the link function that describes the relationship between μ and a linear combination of the predictors. Therefore, the mean response μ is given by:

$$\mu = g^{-1}(\eta)$$

where g^{-1} is the inverse of the link function, and η is the linear predictor of the mixed effects. I use the function *glmer()* from the R package *lme4* (Bates et al., 2014) for fitting the generalized linear mixed-effects models. In particular, I set “family = binomial(link='logit')” in *glmer()*, which specifies the conditional distribution to be binomial. The *glmer()* allows us to fit a generalized linear mixed model incorporating both fixed-effects parameters (business parking supply) and random effect variable (business categories) in a linear predictor, via maximum likelihood (Bates et al., 2017). GLME models are useful for cross-sectional data where the response variable may be other than normally distributed (McCulloch & Neuhaus, 2005). Detailed settings of the model can be found in the following sections.

First, a detailed description of variables used in the analysis is shown in Table 16, showing two categories of independent variables: 1) business categories; and 2) business parking supply and the dependent variable - parking sentiment score.

Table 16. Description of variables

Category	Variable abbreviation	Description	Data source	Data type
Independent Variables				
Business category	Active Life	business category is Active Life	Yelp	Binary
	Arts	business category is Arts	Yelp	Binary
	Automotive	business category is Automotive	Yelp	Binary
	Health	business category is Health	Yelp	Binary
	Hotels & Travel	business category is Hotels & Travel	Yelp	Binary
	Nightlife	business category is Nightlife	Yelp	Binary
	Other	business category is Other	Yelp	Binary
	Restaurants	business category is Restaurants	Yelp	Binary
	Service	business category is Service	Yelp	Binary
	Shopping	business category is Shopping	Yelp	Binary
Business parking availability attributes	Parking valet	business parking valet is available	Yelp	Binary
	Parking lot	business parking lot is available	Yelp	Binary
	Street parking	business street parking is available	Yelp	Binary
	Parking garage	business parking garage is available	Yelp	Binary
	Validated parking	business validated parking is available	Yelp	Binary
Dependent Variable				
Parking sentiment	Sentiment towards parking	parking sentiment score of a business	Yelp	Numeric

The model estimates business-level parking sentiment scores, split the dataset 90/10 into training and testing data. In this model, I seek to understand how the types of parking supplied by a business predicts parking sentiment. Independent variables include business parking attributes and activity types of training dataset to predict the sentiment score in the testing dataset. The results of final fitted model with the minimum RMSE is shown in Table

17. The coefficients are log-odds scaled, shown with standard errors, test statistics (z values) and p-values. I observe that all of the parking attributes significantly explain business sentiment scores.

If a business reports that “street parking” is available, its parking sentiment score is significantly lower. All else equal, street parking is an indicator of a more traditional commercial environment, which results in a more challenging parking search, occurring in traffic and across a wider area (Wijayaratna & Wijayaratna, 2016).

Conversely, if a business has parking validation, nearby garage, or its own parking lot, parking experiences will be more positive. Valet parking has a negative relationship to parking sentiment, implying that drivers view valet parking as time-consuming, expensive, and risky.

Overall, in areas with shared parking facilities, parking was generally viewed more positively or mentioned less frequently. The model does not directly measure parking demand or traffic congestion, but it confirms the intertwined relationship between the type of parking available at businesses and affective experience.

Table 17. GLME modeling results

Model: The Relationship of Parking Sentiment to Parking Supply				
Dependent variables: Parking sentiment score				
Random effect: Business category				
Data source: Yelp dataset				
Weights: Review count per business				
AIC	BIC	logLik	deviance	df.resid
47170	47217	-23578	47156	6059
Scaled residuals (Model 1):				
Min	1Q	Median	3Q	Max
-12.29	-0.81	-0.2	0.69	15.13
Random effects (Model 1):				
Groups	Name	Variance	Std.Dev.	
Business Category	(Intercept)	0.00041	0.0202	
Number of obs: 14199				
Groups: 10 Business Categories				

Fixed effects in the model:

Term	Estimate (p-value)	Std.error	Statistic (z value)
(Intercept)	0.011	0.01	1.11
Parking valet TRUE	-0.065***	0.01	-6.58
Parking lot TRUE	0.012*	0.006	1.89
Street parking TRUE	-0.045***	0.007	-6.85
Parking garage TRUE	0.024**	0.01	2.44
Validated parking TRUE	0.059**	0.023	2.51

Model validation	RMSE
Best model (the current model)	0.51
Alternative Model	0.56

***Significant at the 99% level; **significant at the 95% level; *significant at the 90% level.
Alternative Model: including interaction factors of business categories and business parking availability attributes.
GLME fit by maximum likelihood (Laplace Approximation) ['glmerMod'] in R package *lme4* (Bates et al., 2017, p. 4).

5.3 Impacts of Travel Attitude and Built Environment on Non-work Travel

5.3.1 Background

Extensive studies have been conducted to explore the factors associated with people's travel behaviors. In this section, I further study attitudes in the context of the influence of the built environment on travel behavior, as shown in Figure 30 (b), by integrating geosocial media and GPS trace data. A more detailed background introduction can be found in Chapter 2 and this chapter's section 5.1. A discussion about this data fusion method can be found in the Discussions and Conclusions section.

Moreover, I also include variables such as trip characteristics in analysis in this section. Previous studies found trip characteristics and personal/household factors have an influence on travel behavior as well. Trip characteristics are often measured by level of service and trip-specific factors, such as travel time, distance of travel, and travel cost. For example, Frank et al. (2007) found that the effect of travel time on mode choice is larger than other built environment and socio-demographic variables using a Puget Sound Regional household travel data. Furthermore, the departure time of a trip is found to be related to people's travel mode choice for both work tour and non-work tour and it is correlated with the travel time of certain modes (Ye et al., 2007). Therefore, in this section, the non-work trip characteristics such as travel distance and time of travel (weekday evenings, weekend days, and weekend evenings), and travelers' home CBGs socio-demographic variables are also accounted for the analysis.

5.3.2 Data Preparation

A detailed description of variables is shown in Table 19. I use data from five categories: 1) attitude; 2) behavior; 3) demographic and socio-economic characteristics; 4) Built environment parking; and 5) land use and other built environment characteristics.

(1) Travel Attitude

As I investigate in Chapter 3, I use parking sentiment as an indicator of whether the traveler is satisfied with the parking environment when driving to major commercial and mixed-use districts, serving as a measure of non-work travel attitude.

(2) Travel Behavior

Based on the examination in Chapter 4, I quantify residents' travel behaviors to urban commercial districts using phone-based GPS trace data, which serves as the non-work travel behavior data source.

(3) Census Data

As mentioned in Chapter 1, the census data was acquired online in the US Census Bureau (2018), which covers all the census block groups in Phoenix. I collect a comprehensive list of demographic and social-economic characteristics, such as age, gender, race, income, level of education etc. in people's travel origins (home CBGs). I also obtain the geographical boundaries of the study area by using census API, which can be used for geographical analysis.

(4) BE Parking Data

Parking Availability Data from Yelp

First, I include parking availability data, as described in section 5.2.2. Yelp provides a set of binary categories (True/False) indicating the availability of five parking attributes at each business, such as "parking garage," "parking lot," "street parking," "parking valet," or "validated parking." The parking attributes provide a means of ground-truthing the type of parking supplied in different commercial districts.

Parking Supply Data at Parcel-level

I utilize a parking supply dataset which was created by cross-referencing property-use data and roadway data with minimum parking requirements in the region at parcel-level (Hoehne et al., 2019). Off-street parking was estimated for each parcel according to the required minimum parking by property type outlined in zoning codes. Minimum parking requirements were codified for all cities and towns in the Phoenix region and applied to the 1.6 million parcels of land designated by over 2000 different property types. Total parking was calculated by using the requirement in the zoning code and the size of each building, which was retrieved from the Maricopa County Assessor's Office webpage[22] (Hoehne et al., 2019). The estimation of a variety of supply variables in each commercial district is shown in Table 18.

[22] The estimated parking spaces were validated by manually counting spaces using satellite images, and in some cases, researchers visited sites to count the number of spaces in person.

Table 18. Parking supply of each commercial district

District Name	**District Area (ha)**	**Parking density**				
		Residential off-street spaces	Off-street non-residential spaces	Total off-street spaces	Total on-street spaces	Total spaces of all types
Phoenix Deer Valley	43.02	18.03	23.50	41.54	21.37	62.90
Metro Towne Center	106.75	18.70	36.66	55.36	8.05	63.41
Uptown Phoenix	82.24	32.33	44.52	76.85	14.10	90.95
Downtown Phoenix	180.93	18.15	98.23	116.37	18.52	134.89
Mesa Grand Center	44.93	7.80	34.14	41.93	13.14	55.07
Tempe Marketplace	47.91	3.24	30.89	34.14	6.02	40.16

Note: Parking supply data is from Hoehne et al., 2019.

(5) Land Use and General Built Environment Data

The land use and built environment data was acquired via US EPA (Environmental Protection Agency) smart location database (SLD) (EPA, 2014). The SLD dataset measures of a variety of built environment characteristics and destination accessibility, such as the development density, land-use mix, street and networks connectivity, availability of transit, and accessibility to destinations via car, transit, or foot. The information in the SLD database is available nationwide in the form of ArcGIS online mapping datasets. I use the SLD data at the census block group (CBG) level[23]. The SLD database summarizes over 80 attributes for every CBG in the United States.

[23] A CBG is a geographical unit used by the US Census Bureau (2014) which may have a population of 600-3,000 people.

5.3.3 Methodology

(1) Utility-based Modeling Approach: Multinomial Logit Modeling

Extensive studies have been conducted to explore the factors associated with people's travel behaviors. A primary of modeling approach in this field of analysis is utility-based modeling approach, including utility-maximizing models, activity-based models, and advanced statistical models. Bhat and Pendyala (2005) reviewed diverse methodologies used in studying mode choice and activity behavior, including discrete choice modeling, utility theory, latent class modeling, and micro-simulation approaches. Among these modeling methodologies, multinomial logit modeling is one of the most widespread tools used to study model choice decisions. The multinomial logit model is derived from consumer economics theory with a well-defined mathematical structure and interpretable results, and it was initially developed by McFadden (Domencich & McFadden, 1975; K. E. Train, 2001). The detailed model structure and applications were thoroughly discussed in the book by Ben-Akiva & Lerman (1985). The multinomial logit model captures the underlying mode choice process with utility maximization assumptions that travelers are rational decision makers who are fully informed and are able to choose the mode that has the largest utility for them.

Multinomial logit models have been employed to examine travel behaviors, such as travelers' mode choice, departure time decision, and destination choice. For example, Ashalatha et al. (2013) utilized multinomial logistic regression to analyze mode choice of commuters in India using a revealed preference study. The results showed that the preference of using a car increase with age and the preference of using a two-wheeler decrease in comparison with public transport. Results also showed that commuters tended to change to private modes from public transport if there was an increase in time and cost. Vishnu and

Srinivasan (2013) analyzed the role of individual, household, work-related, modal characteristics and transportation system attributes on departure timing decisions for work and non-work tours of workers. They found that a worker tended to undertake activities such as shopping during the a.m. pre-peak and p.m. peak periods.

In terms of the destination choice modeling, previous practice is extended to investigate the effects of trip characteristics on travelers' destination choice. For instance, Molloy and Moeckel (2017) presented a long distance destination choice model for Ontario, Canada, incorporating data from both Foursquare and traditional data sources in the utility function of a multinomial logit model. They found that travelers are more likely to take urban trips for business but head for leisure destinations outside the city.

Building on literature, I introduce multinomial logit modeling (MNL) in this section to examine how a variety of factors affect travel behavior. I seek to use MNL to analyze the impacts of parking sentiment, parking supply, and built environments on drivers' destination choice for non-work activities.

(2) Model Building

The basic assumption is that the traveler has opinions (knowledge) of the environment and preferences and basic needs, which lead to plans, programs, and schedules. The traveler processes those plans, programs, and schedules in temporal and spatial ways in the non-work destination choice (K. Train, 1986). In a basic construction of MNL, utility is an indicator value for an individual on behavioral choice. Generally, this factor can be derived from the attributes of alternatives, and the utility maximization rule states that an individual will select the alternative out of the set of available alternatives that maximize his or her utility.

The below describes the specification of the destination choice utility function in general terms. The general formulation of MNL can be found in Train (2001). Given a trip origin i, and decision-maker m, the utility of each destination choice j can be written as follows:

$$U_{j|im} = \beta_m \times TravelImpedance_{ij} + \ln(Size_{jm})$$

where the utility of a non-work district choice depends on the impedance or spatial separation between the origin i and the destination j. The impedance term is often referred to as the qualitative utility component, while the size or attraction term is referred to as the quantitative component. In this case, the impedance can be measured by distance of non-work travel (the distance between the traveler home CBG's centroid to the destination district's centroid), time of travel, cost, and other measures of spatial separation. The attraction variable is commonly referred to as the size term, which always enters in log form. It measures the activity opportunities at each destination. In this case, I consider the built environment (including parking availability, supply, and other general built environment characteristics) and parking sentiment as measures of the opportunities or attractions of the destination district.

I use the function *multinom()* from the R package *nnet* (Ripley & Venables, 2021) for fitting the multinomial logistic regression models. First, I choose the non-work travel to "Downtown Phoenix" as the baseline (reference) using the *relevel()* function. Then, I run the model using *multinom()*. The function *multinom()* does not include p-value calculation for the regression coefficients, so I use Wald tests to calculate p-values (Diggle, 2013).

First, a detailed description of variables used in the analysis is shown in Table 19, showing five categories of independent variables used in the MNL model: 1) demographics and social economics; 2) trip characteristics; 3) travel attitude represented by parking experience; 4) general built environment variables; and 5) parking environment variables, and the dependent variable - destination choice for non-work activities.

Then, I built an MNL model to examine the relationship among the behavioral choice to commercial districts by cars, the individual, household characteristics, parking sentiment, and built environment characteristics; then, I identify which of these variables have an effect on the behavioral choice. The investigation is built on the theoretical framework shown in Figure 2 and Figure 3, which is able to demonstrate whether the travel sentiment, in the context of the influence of the built environment, has an effect on the choice of travel, helping to show their effects in the non-work travel decision-making, according to an empirical analysis.

Table 19. Description of variables

Category	*Variable*	*Description*	*Type*	*Source*
Independent Variables				
Demographics and Social Economics	male	% of male population at home CBG	numeric	Census
	age	median age at home CBG	numeric	Census
	income	median household income at home CBG	numeric	Census
	children	mean number of children in a household	numeric	Census
	White	% of White population at home CBG	numeric	Census
	Black	% of Black population at home CBG	numeric	Census
	Latinx	% of Latinx population at home CBG	numeric	Census
	Asian	% of Asian population at home CBG	numeric	Census
	vehicle	% of households owning at least 1 vehicle available at home CBG	numeric	SLD
Trip Characteristics	time of travel	time of day traveling to commercial districts: weekday evening (5pm - 10pm); weekend day (9am - 5 pm); or weekend evening (5pm - 10pm)	dummy	GPS location data
	travel distance	the distance between the centroid of home CBG to the centroid of destination district	numeric	Census
Parking Experience	parking_senti	parking sentiment index at destination district, representing the relative ease to find a parking place	numeric	Yelp
Built Environment - General	o_pop den	population density at home CBG	numeric	SLD
	o_land_use	household Workers per Job Equilibrium Index at home CBG[1]	numeric	SLD
	o_auto access	jobs within 45 minutes auto travel time, time-decay (network travel time) weighted at home CBG	integer	SLD
	o_road den	road density at home CBG	numeric	SLD
	o_inter den	intersection density at home CBG	numeric	SLD
	d_pop den	population density at destination district	numeric	SLD
	d_business den	business density at destination district	numeric	Yelp
	d_land_use	household workers per job equilibrium index at destination district[1]	numeric	SLD
	d_auto access	jobs within 45 minutes auto travel time, time-decay (network travel time) weighted at destination district	integer	SLD
	d_road den	road density at destination district	numeric	SLD
	d_inter den	intersection density at destination district	numeric	SLD
Built Environment - Parking	parking valet	% of available business parking valet	numeric	Yelp
	parking lot	% of available business parking lot	numeric	Yelp
	street parking	% of available business street parking	numeric	Yelp
	parking garage	% of available business parking garage	numeric	Yelp
	validated parking	% of available business validated parking	numeric	Yelp
	r_off_street	density of residential off-street parking	numeric	parcel
	n_off_street	density of non-residential off-street parking	numeric	parcel
	t_off_street	density of total off-street parking	numeric	parcel
	t_on_street	density of total on-street parking	numeric	parcel
	t_all	density of total all types of parking	numeric	parcel
Dependent Variable: commercial district destination choice for non-work activities				

[1]: The closer to one the more balanced the resident workers and jobs in a zip code. Equation: $workers\ per\ jobs\ index\ =\ exp\ (-\left|\left(\frac{Workers}{TotEmp}\right)-1\right|)$.

5.3.4 Analysis of Factors Affecting Travel to Commercial Districts

The modeling results is shown in Table 20[24]. Results show that parking sentiment, parking supply, built environment, and other characteristics of home CBG, affect the destination choice for non-work activities.

(1) Demographics

The model results show that, controlling for other factors including parking experience and built environment, gender - the percentage of male population in traveler home CBG - is a statistically significant variable in affecting the district choice. Specifically, compared to Downtown[25], all other commercial districts are more attractive to females except the Metro Towne Shopping Center. It is possibly because the Metro Towne Mall has lots of stores that sell men's clothing, accessories, shoes, etc. This finding echoes with the previous findings: non-work activity types (trip purposes) are likely to be a factor that constructs the destination attractiveness (Molloy & Moeckel, 2017). With respect to the age effects (the median age in home CBG), the results suggest that Uptown Phoenix, Mesa Grand Center, and Tempe Marketplace attract more young people. Level of education is not a significant variable affecting travelers' trips to these major districts, suggesting that people have general travel needs for non-work tours, regardless of their levels of education attainments.

Interestingly, I don't find income has an influence on non-work travel destination choices, possibly because these six major commercial districts show pricing similarities, especially when assessing the businesses within a district as a whole, their average

[24] This table only keeps variables with significant affects.
[25] Downtown Phoenix is set as the baseline destination choice in the model.

consumption levels become close to each other as a result of they are located in the same region. Echoing the divergent non-work travel behaviors across racial groups found in Chapter 4, I find race - the percentage of White, Black, Latinx, and Asian in home CBG - is a significant variable. Districts' attractiveness to different racial groups are significantly different. For example, Metro Towne Center seems to be less attractive to all of the racial groups compared with Downtown. Phoenix Deer Valley are less attractive to White group, but more attractive to Black group. Latinx group seems to go as many as districts they can, except the Metro Towne Center. In contrast, Asian group shows a strong preference to going for non-work activities in Phoenix Deer Valley.

(2) Trip Characteristics

The distance of travel is not associated with the likelihood of a traveler's commercial activity destination choice; however, the time of travel shows an influence on that. During weekend days, an overall positive affect is found in Downtown Phoenix and Metro Towne Center, suggesting that people are more likely to go to these two *traditional* districts for non-work activities such as shopping, dining, leisure, entertainments, etc. During weekend evenings, *new* urban districts such as Deer Valley or Mesa Grand also attract visitors.

(3) Parking Sentiment

Parking sentiment in each destination district is positively associated with the destination choice to all the five commercial and mixed-used districts in the region, showing a relative hesitation of its residents in driving to Downtown when compared to going other major districts. Residents may feel stressful especially when they need to find a place to park their cars in Downtown. This suggests that travel attitude does have an influence in non-travel behavior and destination choice, controlling for other demographic and built environment factors.

(4) Built Environment

Home Built Environment

Individuals who live in areas with high population density and road density are more likely to drive to almost all other major districts rather than Downtowns, probably because their neighborhoods are very drive-centric so they get used to driving and are more likely to drive to districts with convenient and easy parking.

Destination Built Environment

With respects to the built environment in the destination, I find that a variety of variables have different influences on the residents' district choices. For example, if the road density increases in districts like Phoenix Deer Valley or Mesa Grand Center, their attractiveness will increase, maybe because the traffic near these two districts are not very congested. In Phoenix, Uptown and Downtown are common places that have more congestion and delay (City of Phoenix, 2020).

Interaction of Parking Sentiment and Built Environment

Interestingly, I do find that an interaction variable combining parking sentiment and parking lot supply is statistically significant to the districts' attractiveness, suggesting a collective effect of positive parking sentiment and increase in parking lot on the non-work travel destination choice.

Destination Parking Supply

Regarding the supplies and provisions, I find that districts with more shared parking facilities are associated with the non-work destinations' attractiveness. Construction of more parking lots will increase more mobilities there across all the five major districts. It is noticeable that not all the parking types increase will positively affect their travel attractiveness. For example, more parking garages may help some districts to attract more visits, excluding Uptown Phoenix and Tempe Marketplace. A possible reason for this is because their garage parking rates - it might be costly for drivers to park in the garage parking there. In addition, off-street parking is insignificant in these two districts as well, possible because off-street parking is less attractive to visitors in Uptown, Downtown, or Tempe Marketplace. Why does off-street parking become less attractive? One possible reason is the pricing strategy - for example, in Downtown Phoenix, off-street parking is paid, not free parking.

Table 20. Modeling results

Category	Variable	Destination district									
		Phoenix Deer Valley		**Metro Towne Center**		**Uptown Phoenix**		**Mesa Grand Center**		**Tempe Marketplace**	
		coefficient	*p-value*	*coefficient*	*p-value*	*coefficient*	*p-value*	*coefficient*	*p-value*	*coefficient*	*p-value*
Demographics and Social Economics	male	-	**	+	***	-	*	-	*	-	***
	age	+	**	+	***	-	*	-	**	-	***
	White	-	***	-	***	+	*	+	***	+	***
	Black	+	**	-	***	-	***	+	**	-	***
	Latinx	+	***	-	***	+	*	+	*	+	***
	Asian	+	***	-	***	-	***	-	***	-	***
Trip Characteristics	time of travel (base: weekday evening)										
	Weekend day	-		+	*	-	***	-	*	-	***
	Weekend evening	+	***	+	***	-	***	+	**	-	***
Parking Experience	parking sentiment	+	***	+	***	+	***	+	***	+	***
BE - home	population density	+	***	+	***	+	***	+	***	-	***
	land use mix	-		+	***	-		-	***	-	*
	road density	+	***	+	***	+	***	+	***	+	*
BE - destination	pop density	+	***	-	***	+		+	*	-	
	business density	-	***	+	***	-	***	-	***	+	*
	land use mix	+	***	-	***	+	***	+	***	-	**
	road density	+	***	-	***	-	***	+	*	-	*
BE - parking	parking lot supply	+	***	+	***	+	*	+	***	+	***
	parking garage supply	+	*	+	*	+		+	*	+	
	commercial off-street parking	+	*	+	**	+		+	*	+	
	Interaction variable: Parking sentiment * Parking lot supply	+	***	+	***	+	***	+	***	+	**

***Significant at the 99% level; **significant at the 95% level; *significant at the 90% level.

Note: this table uses plusses and minuses to show the variables' coefficient effects because it can present the results quickly and directly if having a long list of variables. Readers can quickly find whether an independent variable has a positive / negative influence (or no influence) on the destination choice.

5.4 Discussions and Conclusions

5.4.1 Discussions

In this chapter, I summarize the main findings of the prior two empirical analyses, and review the conceptual framework and literature. I start the spatial analysis by using a Gi* hotspot analysis to assess the spatial distribution of the parking sentiment in the region. The spatial clusters suggest that factors such as the type of parking supply may be influencing parking experiences in the non-work destinations. Then, I perform two "data integration" analyses. First, I unveil the relationship between parking provisions and parking sentiment by integrating the supply data and attitudes. Second, I examine travel attitudes in the context of the influence of the built environment on travel behavior by integrating the results obtained from investigations on geosocial media and GPS traces in Chapter 3 and Chapter 4.

Regarding the analysis of the effect of the provision of parking on parking experiences, I find that transportation system management and the built environment have significant impacts on how individuals experience daily travel. The spatial hotspot analysis shows that negative parking sentiment clusters are associated with central business districts (CBDs) in Phoenix region but are not always associated with business clusters outside of CBDs. The GLME model suggests that the type of parking available is also significantly associated with parking sentiment. Specifically, street parking is always linked with negative parking sentiment, and if a business has parking validation, garage, or its own parking lot, parking experiences will be more positive. The results suggest that consumers seek convenience parking and may feel more stressful or have more complains when they have to parallel park on streets.

With respect to impacts of parking sentiment, built environment, and socio-demographics on non-work travel destination choice, the multinomial model presents different effects of these variables on each district. Variables such as gender, age, race, time of travel, the built environment of home and destination locations, the parking supplies, are disproportionately affecting the district choice. This investigation provides an evidence to echo the conceptual framework for the interactions between travel attitude, built environment, and activity-travel behavior. Although it only serves as an example of many potential analyses in this area, it demonstrates the effort of explore the non-work travel by fusing multiple new transportation data.

5.4.2 Limitations

This analysis has some limitations as well. First, the data fusion analyses may face some inherent limitations: different demographic bases for multiple datasets. In this case, I have two demographic bases: Yelp users and the individuals captured in the phone-based GPS dataset. Although both of them are from the same geographical region - the Phoenix Metropolitan Area, their populations may be fundamentally different. This indicates that data fusion is a challenging task since the datasets are generated differently and there is some underlying information to which I have no access (e.g., the demographical information of each Yelp reviewer). Researchers need to consider whether different datasets can interact and inform each other before processing any data fusion tasks. Second, the GPS data used in the analysis was only collected for a month long. Future research may need to collect more data over a longer duration in order to better estimate the non-work travel behaviors.

5.4.3 Conclusions

Most transportation studies only focus on the effects of attitudes on people's residential location choice and travel behavior (referred to as residential self-selection), the focus of this chapter extends travel behavior research with a focus on the non-work travel. Building on the literature, this chapter designs two data integrations. First, it examines the relationship between attitudes and built environment by integrating parking attitudes and parking supply data, as shown in Figure 30 (a). Then it investigates travel attitudes in the context of the influence of the built environment on the generation of non-work trips (travel behaviors) by integrating geosocial media and GPS trace data as shown in Figure 30 (b).

These analyses are important, and worth further exploration, because the information embedded in either geosocial media or smartphone-based GPS trajectories are rich and compelling. In future research, planners can develop more measures for traveler information mining and behavior analysis, for example, may be a way of dynamically monitor the experience of travel.

CHAPTER 6

CONCLUSION

The advances in information and communication technology and mobile devices create new opportunities for today's transportation planners to understand cities and travel using non-survey sources of data, which are user-generated, geo-located, and contain contextual information. The emergence of such "transportation big data" has resulted in a large quantity of information documenting people's everyday movements, travel events, attitudes, perceptions, and emotions, all connected with the location and time. Why do people travel? How does travel among different groups vary? How have data and methodologies for travel behavior analysis changed in recent years? How do we understand the experience of travel and the behavior of travel using new data? What are the potential and limitations of new data-driven approaches in transportation planning? This dissertation seeks to utilize and integrate multiple types of transportation big data and analytical methods to explore these high-level questions in transportation research. Through designing a series of empirical analyses, it shows the promises of emerging data and analytics in providing useful information about non-work travel experiences and behaviors. It also enhances our understanding of the dynamic interactions between transportation systems, built environment, and people.

This chapter is organized as follows: First, I discuss some of the main challenges raised in data-driven transportation planning research. Second, I provide the major findings of this dissertation. Then, I summarize its contributions and describe my academic collaborative experiences and publications drawing on this dissertation research. Finally, I suggest a variety of problems for future research scope.

6.1 Discussions

Although new data and methods provide tremendous possibilities in transportation planning research, they raise challenges as well. Key questions regarding the applicability, reliability, and effectiveness of emerging data and methods need to be carefully considered so that they can be appropriately applied into solve practical problems. Data-driven analysis also asks for a careful selection of analytical methods. If these data are analyzed with inappropriate methods, the results can be misleading. In this section, I discuss some of the major challenges associated with this dissertation research:

6.1.1 Inherent Limitations of the Types of Data

In this dissertation, I study characteristics of both traditional data sources (survey-based) and user-generated, crowdsourced datasets (non-survey sources) in transportation planning. Although as the empirical analyses proceed, I demonstrate the use of these emerging data in data-smart transportation informatics, I find both the traditional data and big data have their inherent limitations in transportation research.

(1) Survey-based Data

Survey-based data, such as interviews, surveys, or travel diaries, though in-depth, are still limited in scope (scale) as they reflect only the small number of participants in a project. Additionally, these survey-based data collection methods can be time-consuming, expensive, and infrequently up to date. Survey participants have to remember their travel behavior and recognize their trip attributes often without any supplementary support, which may cause under-reporting of trips, imprecision or absence of locations and times. Government sourced data, e.g., census data, are commonly utilized in existing travel behavior studies, and, while informative, remain restricted by the types of data that the government has collected.

(2) Big data: Geosocial Media Data and Smartphone-based GPS Data

Emerging transportation big data, collected from either crowdsourcing, the Internet of Things (IoT), urban system records, or other novel sources, are much larger than surveys in size and scope. In recent studies, geosocial media and human GPS trace data are two widely used new data sources for transportation research. In this dissertation, I review their characteristics and applications in transportation literature. Then, I design the research framework and utilize Yelp reviews and smartphone-based GPS data to understand non-work travel experiences and behaviors. I find that geosocial media and human GPS trajectories have some common characteristics: large size (up to terabytes of data), incomplete demographics, explicit or implicit geographic information, and rapid and potential real-time updating. I summarize two main challenges that both geosocial media and GPS data may face: representativeness and missing ground truth.

Representativeness

These big data are often burdened with questions of representativeness and ethical use that must be appropriately addressed, particularly when used for public purposes. Thus, it is worth noting that the two types of data used in this dissertation may have representativeness bias, which means if the data do not represent the population, the analytical results, interpretations, and conclusions will be biased. In this sense, there is a pressing need to investigate the representativeness of data, which asks for further research on whether these data can (or cannot) be used in human mobility and transportation planning research.

For example, the demographics of Yelp users may not necessarily represent the entire population. Compared to the average U.S. adult population, Yelp users are younger, more highly educated, wealthier, and tech-savvy. In Chapter 3, I use multiple approaches to estimate Yelp data's representativeness - an issue more broadly associated with all other self-selecting geosocial media. For example, I examine gender, income, race of Yelp users in order to make sure the subset used in this dissertation is unbiased enough for conducting the following analyses. Similar to this representativeness examination, in Chapter 4, I notice that participants in the smartphone-based GPS data may not fully represent the population as well, so I use a weighting method to adjust the dataset. As these user-generated data can grow over time, addressing these issues requires a combination of empirical methods to diagnose or control for potential biases, as well as clear caveats and recognition of potential effects of biases that cannot be controlled.

Missing Ground Truth

I find that these transportation data may not have "ground-truthed" information, which refers to the accurate information that are associated with the objects in the data. For example, the smartphone-based GPS trajectories do not have ground-truthed labels for participants' travel mode choices (walk, bike, bus, transit, car, etc.), I can only use estimation methods to infer their modes. Even though the lack of ground-truthed labels is an inherent limitation and is commonly found in many types of transportation crowdsourced data, it can trick analytical models and may turn into false and meaningless results (P. R. Stopher et al., 2015).

(3) Summary

In this dissertation, I examine Yelp dataset and smartphone-based GPS traces with a variety of analytical approaches, including textual analysis, GIS spatial analysis, and statistical methods. I find that these new transportation big data cannot fully replace the traditional travel data and methods for now, but they can be utilized in parallel to shed light on travel behavior and integrated transportation and land use planning.

For example, when I utilize Yelp transportation reviews to obtain the aggregated sentiment toward shared parking, I find it important to test related parking policies' effectiveness and acceptance for a district. I recognize that new data and their analytical techniques require further investigation as to the most effective methods to provide conceptually valid and interpretable results in data-smart transportation. Therefore, this field of research needs to compare and evaluate the validity, performance, and interpretability of the results and application contexts. According to my research experiences gained from this dissertation, I find that mitigating or highlighting potential representation and accuracy biases is an essential step in transportation analysis.

6.1.2 Technical, Analytical, and Legal Challenges of Transportation Big Data

There are some technical, analytical, and legal challenges in transportation big data processing. Some of these challenges are due to the data accessibility. For example, the phone-based GPS data are collected by a location data company, risks may exist in 1) private actors' inappropriate infrastructure and data management and security systems; 2) researchers' methodological difficulties in extracting meaning from huge, complex and "noisy" volumes of data.

In fact, I find that using transportation big data to understand and assess the performance of a planning policy decision and then refine the future decisions is advocated be one of the - not necessarily new - promising directions for planning research. However, technical and analytical challenges are due to its highly interdisciplinary nature, which requires an interplay across disciplines, including computer science, information systems, transportation planning, urban design, etc. In this new context, I would argue that a combination of data science and planning expertise is needed, but major challenges lie in how to merge the imperfect, real-world datasets that urban planners must work with and the classical data / computer science methods, to solve practical problems and apply new techniques in urban planning and city management.

There are also issues of continuity of data, considering the rapid pace of technological change and innovation, and difficulties in gaining an overall picture of which big data sources or innovative methods can yield useful insights for transportation and planning policy. In this sense, the future development of "cross-section partnership" in data exchange and collaborations between the public and private could be helpful to make progress in handling this issue (Susha et al., 2017).

6.1.3 Transportation Policymaking Challenges

While a series of empirical works apply emerging data into urban transportation research, more research is still needed to establish the field of transportation planning informatics and data-driven policymaking. It is critical to concretely explore how to harness new data sources for transportation analysis and policymaking. Specifically, though promising, it faces four major challenges: 1) difficulties in combining traditional and innovative datasets and methodologies for travel analysis and its relevance for policymaking; 2) ethical issues associated with the use of data and protection of individuals' privacy; 3) difficulties in promoting and facilitating new forms or partnerships across multi-sectors, from policy to scientific communities; and 4) time consuming in sharing good practices by pilot studies in multiple cities and regions.

In addition, long range planning is more challenging than the short-term analysis because it requires continuous efforts of transportation policy interventions and expects to change residents' travel behaviors. Although the immediacy of transportation big data may make it useful and compelling for real time urban management, the long-term goal cannot be met by big transportation data analysis alone, instead, a methodological framework is needed to forecast and examine the impacts of planning interventions.

6.2 Conclusions

Cities have rapidly been shaped by the continuing advances of information technology and data. This dissertation details the process of employing new transportation big data to deliver advanced information about travel. Specifically, it integrates geosocial media data and smartphone-based GPS traces to examine travel to non-work destinations, taking Phoenix metropolitan area as a study case. It performs three empirical analyses and employs two "data integration" analyses to: 1) unveil the relationship between parking provisions and parking sentiment (attitude) by integrating the supply data and online review texts; and 2) investigate how travel attitude, the characteristics of where people live, and other built environment characteristics affect destination choices by integrating the results obtained from the prior two empirical analyses.

This dissertation finds that transportation system management and the built environment have significant impacts on how individuals experience daily non-work travel. A spatial hotspot analysis of parking sentiment shows that people's attitudes and emotions towards parking are spatially clustered in Phoenix region: traditional centers of commercial activity such as Phoenix downtown are always associated with negative parking sentiment clusters; while suburban business districts have more positive clusters. It is still stressful for most people when parking in central business districts. Statistical models help to explain how different types of parking availability are associated with parking sentiment and how parking sentiment, built environment, and socio-demographics collectively affect non-work travel destination choices. The findings suggest that if a business has parking validation, garage, or its own parking lot, parking experiences will be more positive. Specifically, street parking is always linked with negative parking sentiment, which suggests that consumers often seek convenience parking and may feel more stressful or have more complains when they have to parallel park on streets. Also, variables such as built environment characteristics of home and destination locations, parking supplies and sentiments, and time of travel, are disproportionately affecting the non-work district choice. Interestingly, the findings show that the distance of travel does not affect the destination choice, but an interaction variable of the destination parking sentiment and destination parking lot supply is statistically positively with the choice of destination district in Phoenix. An increase in parking lot capacity in districts with more positive parking sentiments will encourage more consumers to visit there.

Throughout the analyses, this dissertation emphasizes the potential of new data in transportation planning informatics. It shows that these data and methods can, when properly used, reveal new dynamics and open up new approaches to study traveler attitudes and experiences, behaviors, and mobility patterns. Collectively, these empirical analyses and findings echo the conceptual framework and demonstrate the intertwined interactions between non-work travel attitude, behavior, and built environment.

This dissertation is a piece of a broader agenda to employ big data and data science methods into transportation and urban planning research, in order to better understand the relationship between travel attitude, behaviors, and the built environment and to craft more effective transportation planning policies and interventions. The examination of Yelp reviews and personal trajectory data gives some preliminary findings for the transportation planning informatics research, which assists to deepen the understanding of travel behavior decisions and have implications for the contemporary practices of transportation planning. As people often drive to destinations for a variety of casual activities, and such urban travel is always more complex than commuting travel between home and work, this dissertation research argues the promising provided by novel datasets and analytics methods can help transportation planners to understand non-work travel and acquire useful information about transportation and cities.

6.3 Contributions

6.3.1 Enhancing the Understanding of Non-work Travel

This dissertation examines a variety of aspects of non-work travel, including travel experience, trips and travel patterns to non-work destinations, non-work travel accessibility of different income levels and race/ethnicity. The findings show that the use of emerging transportation big data can enhance the understanding of non-work travel.

First, the analysis is able to capture people's attitudes and sentiments towards parking in non-work destinations, which can be used for planners to collect such information using non-survey sources of data. Specifically, I find that geosocial media users are frequently sharing travel experiences on online review platforms such as Yelp when going to varied types of non-work destinations, and parking is of interest to them, which can be used to gauge the experience of travel in commercial districts and centers. In addition, I find the share of parking sentiment (positive VS. negative) varies across the commercial district. In general, mall-like commercial districts such as Phoenix Deer Valley and Mesa Grand Center have overall more positive parking sentiments than that in strip-like districts, possibly because these mall-like districts are located in suburban areas and often provide sufficient surface parking capacity which may provide drivers more parking convenience.

Moreover, I find Phoenix residents frequently drive to six major commercial and mixed-use districts for non-work activities during weekday evenings, weekend days, and weekend evenings. Most of them go to city centers (e.g., Phoenix Uptown and Downtown) for non-work activities. Regarding non-work travel behaviors and the realized accessibility, I find disparities exist in income and race/ethnicity, associated with different time, and locations. For example, black-dominant neighborhoods have overall longer travel distances when going to the major commercial districts so they experience worse accessibility across the racial groups. Intriguingly, I find people from non-poor neighborhoods frequently visit some old city centers (e.g., Downtown) although they need to spend more travel time, which suggests that a gentrification phenomenon - high income people choose to live far away from these old centers - but they don't mind a higher time cost in non-work travel. The analysis also indicates the association between built environment and attitude: the transportation system management and the built environment have significant impacts on how individuals experience daily non-work travel. Most people still feel stressful to find a place to park in central business districts. The findings suggest that the types of parking provisions have an influence on parking sentiments as well. The results confirm the intertwined relationship between attitude, behavior, and built environment. In specific, attitude and built environment disproportionately affect non-work travel behavior.

6.3.2 Integration of Crowdsourced Transportation Big Data

Another contribution of this dissertation to transportation planning research lies in the integration of crowdsourced transportation big data into the advanced travel information collection and analysis. In Chapter 3, using the features of the Yelp dataset, I extract the transportation content embedded in the review texts. Particularly, I analyze the sentiment towards parking by assessing satisfaction or frustration with parking at different types of businesses in six major commercial districts across the region. In Chapter 4, according to the precise spatial and temporal information of the phone-based GPS locations, I analyze the travel needs to major commercial and mixed-use districts and compare the trip characteristics (e.g., the distance of travel) and accessibility of individuals from different types of neighborhoods. In Chapter 5, I first examine associations between parking supply and parking sentiment and find that commercial districts with shared parking supplies have more positive parking sentiment. Then I present an integration of multimodal data, including geosocial media data, phone-based GPS traces, census data, land use data, and other data sources, to study impacts of the travel attitudes and built environment on behaviors.

6.3.3 Harnessing Possible Challenges associated with Emerging Transportation Data

As discussed before, there are biases inherently associated with transportation big data in their specific contexts, and there are no single or simple ways to mitigate biases. Current papers do acknowledge the bias of big data in transportation informatics, but they lack a clear finding as to why a particular method is trustful and can be leveraged to solve the problem. To fill this research gap, in this dissertation I develop three ways to mitigate representation biases through resampling, synthesizing (with locational demographics from census data), and combining the advantages of each data type.

For example, to obtain a representative sample for car users in the phone-based GPS dataset, I resample the data by filtering out a sub dataset with a distribution close enough to the time and spatial patterns of the target population. Also, human GPS trajectory data have some inherent noisy information due to the inaccurate location identification in buildings or rural areas, which may result in a precision bias. In this sense, I use a buffer of distance and time to find the adjusted path of a traveler.

6.3.4 Interdisciplinary Lens in Transportation Planning Informatics

The other significant contribution of this dissertation is the interdisciplinary lens it provides to the research in transportation planning informatics, particularly when I seek to use transportation big data to serve the public interest. New data is too valuable to only serve the commercial interests of a small number of corporations that are positioned to collect and utilize it. To ensure the success of data-driven transportation planning and management, researchers and practitioners should plan and design the cross-disciplinary work agenda ahead. This dissertation serves as an example of showing the interdisciplinary research can combine knowledges and create a "collective intelligence" (Schoder et al., 2014) to assist the evaluation of human mobilities and local planning policies.

Methodologically, Chapter 3 examines the transportation content and experience of parking via geosocial media platforms such as Yelp reviews, which suggests that geosocial media data can provide rich information to planners to detect collective public opinions towards transportation system. Human emotions, thoughts, culture, and perceptions can be connected with location and can be captured over time to understand where different sentiments manifest themselves geographically. The sentiment analysis in Chapter 3 categorizes and analyzes sentiment toward parking environment across the Phoenix region and the six major commercial districts. Spatial clustering methods and spatial statistical methods are then applied in Chapter 5 to geocode parking sentiment on the map to examine the distribution of sentiment about parking. These methods, including text mining, sentiment analysis, and spatial analysis, are brought together to extract useful travel experience information to non-work destinations from big geosocial media data.

Chapter 2 includes a review of non-work travel and its two dimensions: non-work travel patterns and measurements. There are more and more non-work trips, however, compared to work trips, only a limited number of literature focus on non-work travel. Travel patterns of non-work activities are less routinized than commuting, in this sense, the lack of fine-grained data impedes the non-work travel measurements at high resolution. With the wide adoption of smartphones, human movement information can be captured with accurate time and place via GPS-enabled smartphone apps. Chapter 4 examines two dimensions - accessibility and travel cost (distance of travel) - of non-work travel via smartphone-based GPS traces, which suggests that big human mobility data can be used to analyze trips to non-work destinations and capture such spatial interactions between people and locations. Furthermore, Chapter 5 integrates multiple types of data to quantify the impacts of built environment (parking supply) on travel attitude (parking sentiment), and the impacts of built environment, travel attitude, and socio-demographics on travel behavior. Diverse methodologies are applied to these analyses, including big data processing and analytics and statistical modeling, which provides an application example of data-driven innovative solutions in non-work travel research from a cross-disciplinary perspective.

Contemporary cities are drowning in data, if planners can collaborate with data scientists, urban scientists, and computer scientists, the outcome could overcome the challenges of sustainability and urbanization. Each type of new data has its specific advantages, limitations and application scopes for transportation informatics, and one data type alone is not likely to be capable of providing multi-dimensional information about travel patterns and the overall state of the system. Therefore, the investigation throughout this dissertation demonstrates the potential of an integration approach which can combine multimodal types of new data with different features, structures, resolutions, and precision for understanding, imagining and shaping the future of data-smart transportation functioning, planning, design, development, and management.

6.4 Collaborations and Publications

As my dissertation research focuses on studying transportation planning and informatics through a data-driven approach, from a highly interdisciplinary perspective, thus, along with my doctoral path, I actively collaborate with academic scholars from diverse departments (e.g., civil engineering, digital humanities, computer science, environmental science, etc.) and institutions (e.g., Rutgers, Virginia Tech, etc.) I also present my collaborated work in academic conferences, such as ACSP (Association of Collegiate Schools of Planning), TRB (National Academies Transportation Research Board), CUPUM (Computational Urban Planning and Management), Women in Data Science, Diversifying Scholarship Conference, ASCE (American Society of Civil Engineers) Transportation and Development conference, and so forth.

During my doctoral studies, I have co-authored, published ten peer-reviewed journal articles, one book chapter, and two conference proceedings. Table 21 shows a list of publications and presentations that are only related to this dissertation research, either conceptually, methodologically, or empirically.

Table 21. Selected publications and presentations

Jiang, Z., Mondschein, A. 2021. Analyzing Parking Sentiment and its Relationship to Parking Supply and the Built Environment Using Online Reviews. *Journal of Big Data Analytics in Transportation. 3, 61–79. https://doi.org/10.1007/s42421-021-00036-1*

Jiang, Z., Zheng, M. 2021. Public Perceptions and Attitudes Towards Driverless Technologies in the United States: A Text Mining of Twitter Data. *Conference proceedings in the CUPUM (Computational Urban Planning and Urban Management) 2021 Conference: "Urban Informatics for Future Cities".*

Jiang, Z., Mondschein, A. 2020. Examining Transportation Accessibility to Commercial Areas using Phone-based GPS Traces. *Presented at the Association of Collegiate Schools of Planning Conference, October 2020, Virtual.*

Mondschein, A., King, D. A., Hoehne, C., Jiang, Z., & Chester, M. 2020. Using social media to evaluate associations between parking supply and parking sentiment. *Transportation Research Interdisciplinary Perspectives, 4, 100085. https://doi.org/10.1016/j.trip.2019.100085*

Jiang, Z., Mondschein, A. 2019. Examining the effects of proximity to rail transit on travel to non-work destinations: Evidence from Yelp data for cities in North America and Europe. *Journal of Transport and Land Use, 12(1), 303-326. https://doi.org/10.5198/jtlu.2019.1409*

Jiang, Z., Mondschein, A. 2019. The Effect of the Built Environment on Parking Experiences: Evidence from Sentiment Analysis of Yelp Reviews. *Presented at the Annual Meeting of the Transportation Research Board, January 2019, Washington, DC. https://trid.trb.org/view/1572914*

Jiang, Z., Mondschein, A. 2018. Examining Non-Work Accessibility in Commercial Areas: Evidence of Parking and Nonmotorized Experiences using Sentiment Analysis of Yelp Reviews. *Presented at the Association of Collegiate Schools of Planning Conference, October 2018, Buffalo, NY.*

For example, in the 2019 publication (Z. Jiang & Mondschein, 2019), I first examined the rich content about transportation mentioned in Yelp reviews, so I asked, how does the public perceive their accessibility opportunities and their relationships to the local contexts? To answer this question, I examined the behavioral geography of transit-adjacent development from the travel reviews about public transit, and found that reported rail use for diverse activity purposes is highly but variably sensitive to proximity to stations. In the 2021 publication (Z. Jiang & Mondschein, 2021), I integrated the derived parking reviews and analyzed the parking sentiment and its associations with parking supplies and built environment characteristics. The findings indicated transportation policies such as shared parking may shape how individuals perceive parking availability in commercial districts. These publications and presentations are inspired by this dissertation research.

Overall, my dissertation research seeks to deepen the knowledge of using emerging datasets in human mobility and travel behavior analysis research more broadly. Most of the current research on social media or other crowdsourced datasets are descriptive and do not often address whether limitations in these data can be mitigated conceptually or methodologically to give insight on transportation planning interventions through experiential and empirical processes. In the near future, I expect to derive more pieces from this dissertation work and publish them in journals or conferences.

6.5 Future Work

Some important extensions are worth mentioning at the end of this dissertation and they will suggest more possibilities for future research. These extensions may include advanced analytical methods with higher data dimensionality, travel behavior studies with more ground-truthed datasets, and data fusion for intelligent transportation planning and management.

Below I describe some of the extensions for future research in this area, guided by the findings of the dissertation. They will allow me to extend the current data integration research for non-work travel to a boarder scope of data-driven transportation planning research, and also address some of the shortcomings and limitations not covered in this dissertation.

6.5.1 Multiple Urban Data Fusion

Ever increasing sources, quality, and types of data translate into a new opportunity to leverage data science and analytics to help city planners, infrastructure engineers, and policy-makers both understand and engage with cities in entirely new manners. Future research can further integrate multiple types of urban data for city planning and management.

(1) Travel Attitude of Underrepresented Population

Multiple data fusion, especially for disparate types of "urban data", is still at the preliminary research stage. Future research needs to further combine multiple data sources to collectively solve problems. For example, in Chapter 3, I find some of the inherent limitations associated with geosocial media data: the demographics of Yelp users may not necessarily represent the entire population. Compared to the average U.S. adult population, Yelp users are younger, more highly educated, wealthier, and tech-savvy. Thus, some research questions may continue to require survey efforts that reach populations that do not participate as readily in Yelp, or where Yelp reviews does not supply critical information to answer those questions. E.g., senior people's opinions may not be fully represented on Yelp, future research should try to have a better grasp of the senior population's view and perceptions. Similar to the population bias of Yelp users, smartphones are not fully adopted to all population, especially for seniors. Thus, using big data for travel studies need to mitigate bias since they may under-represent older segments of the population. I would state that the presence of bias in big transportation data does not preclude its usefulness, but suggests that importance of recognizing and addressing the issue.

(2) Collective Dynamics of Travel Behavior

Extending from this dissertation, future research can match individual-level travel behavior with place of interests on geosocial media, rather than focusing on district-level only. For example, each destination of interest on geosocial media such as Yelp can be connected with consumers' travel behaviors to the destination. I would argue that matching individual-level behavior with place of interests is a powerful approach to understanding the collective dynamics of human behavior. There are multiple types of non-work destinations, future research can reveal what types of non-work activities are more attractive and what are less attractive. More importantly, this pairing approach can unveil travel origins (dwelling locations), travel path, mode of choice, non-work activities (destinations), and time use allocation (duration), which are valuable to understand urban mobility patterns at high resolution. In addition, such dynamics of the travel-activity patterns retrieved from the pairing approach are useful to develop travel demand models. Planners can make use of the dynamics to see how changes in travel and time variables lead to changes in non-work activity participation in places of interests. These interactions also have implications for evaluating economic outcomes and land use plans.

(3) Sensing and Detecting Transportation Events

In Chapter 3, I have demonstrated that text mining analytics can be used to collect the public's embedded attitudes, perceptions, and emotions towards the mode of transportation to various destinations in Yelp reviews. There are other useful geosocial media datasets generated by users that can be integrated with transportation geographic information (e.g., road networks, infrastructure) for planners to sense and detect transportation events. One example is Twitter, a popular social networking site, which can be utilized by the public to share information about their daily lives through micro-blogging. "Traffic" is a popular topic people would like to discuss in their daily lives. Twitter provides a free approach to acquire public tweets through open API to mine out Tweets (Twitter, 2018). Future research can integrate volunteered geographic information and Twitter data to detect traffic events. An integration of GIS data and geosocial media data such as Twitter can be used to generate a wide range of traffic related timely, on-the-ground information to detect traffic events to support traffic planning and management.

(4) Mitigating Data Bias

In addition, more future empirical analyses on multi-data fusion can help to evaluate the observations from one dataset alone. Introducing the government GIS data is useful to validate spatial observations from geosocial media analysis. Conducting analysis of other types of geosocial media (Facebook, Twitter, TripAdvisor, Foursquare, etc.) can be used to reduce sampling bias. Also, multiple urban data fusion may help to provide a better explanation of why travelers have such attitudes and behaviors. Although in Chapter 3 the sentiment analysis can uncover the overall sentiment and ratios, it cannot clearly explain the latent reasons that dominate these sentiments. Further survey-based research can be used as a comparison to test whether the sentiments carried by the survey results agree with that of geosocial media. In Chapter 4, I can observe non-work travel behaviors by using the location traces, but I cannot infer the motivations behind the behavior patterns. Non-work travel may be related to personal tastes and lifestyle choices, and such factors could vary across different cultures. Therefore, the survey-based data are still powerful and promising for planners to explain the observational phenomena of travel. Future research can include such “small data” into the data fusion framework so that to support planners’ critical thinking on multiple dimensions of urban data (human mobility data, land use / built environment data, housing / neighborhood data, etc.)

and extend their insights for planning interventions and policies.

6.5.2 Impacts of Geosocial Media on Travel Behavior

This dissertation mainly focuses on extracting useful information about travel from geosocial media datasets. In fact, people not only use social media platforms to share their experiences of travel (post-trip), they also use the information shared on social media for trip plans and ideas (pre-trip). Social media sites are increasingly being used by travelers to get information regarding their route plan or get reviews about destinations. This way makes travelers feel that the information they get from these sources will be unbiased and trustworthy.

It has been observed that the public is actively using social media to exchange opinions and experiences about the trip before, during and after the trip as well. Existing studies have shown that even a novice user can now easily share photos, videos, blogs, etc. so as to contribute ratings and recommendations for a destination, and another traveler can easily get information about any destination he is planning to travel to, get reviews about the place (restaurants, movie theaters, grocery stores, hotels, etc.), which makes it much easier for the traveler to decide whether or not to visit that particular destination. Thus, social media can affect trip plans, mode choice, and travel behavior. This interaction between the social media data and travelers could be a worthwhile future study. Future research can further explore the impact of social media on pre-travel planning and decision making of travelers.

6.5.3 Feasibility, Reliability, and Generalization Power of New Transportation Data

The analyses and discussions throughout the dissertation present an increasing need in systematically address issues that are associated with the feasibility, reliability, and capability of generalization of new transportation data. According to recent literature, although new transportation data have some strengths, e.g., longer collection duration, large number of participants, large spatial coverage, ease of access, low cost, and accurate location information, rigorous cross-validation of the use of emerging data in transportation planning is needed. Future research can further explore the validity and robustness of the new transportation data across spatial and temporal scales.

For example, travel demand estimation is essential for urban planning and management of transportation networks. The time series of visits to various locations by individuals are aggregated to study the flows of people between different zones/regions. Traditionally, the estimation of the OD matrices relies on input data from household travel surveys, census data, and other traffic surveys. However, these surveys are not often available across different geographical levels. Thus, new transportation crowdsourced data analyzed in this dissertation can be constructed to analyze the origins and destinations of all trips if they can be validated for their feasibility, reliability, and generalization power in approximating the travel demand. One future direction is to examine the robustness of the travel demand estimation for different spatial scales by comparing the trips estimated from new data and the surveys.

This direction of future research is inspired by the fact that technologies and urban data science methodologies has become a fundamental element in the development of a better, smarter future for our cities, and transportation planners can further develop their own methods to make use of new data in a rigorous way. Empowered with the methods, planners can better understand, imagine and shape the future of cities and make urban life more sustainable and equitable.

Appendix A Trend of VMT for All Trips in the United States

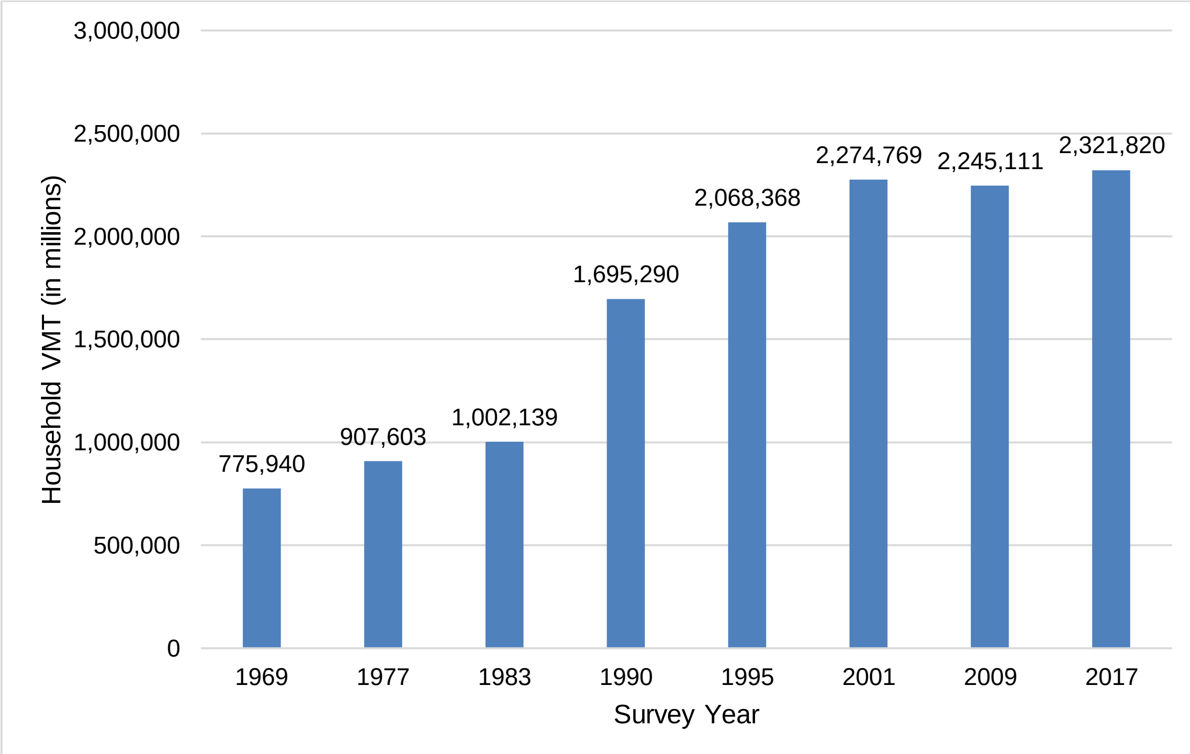

Figure 32. Trends in VMT estimates

Source: data is from table 1d, p8, McGuckin & Fucci, 2018.

Note:

1. 2017 NHTS were adjusted to make them more comparable with later surveys.

BIBLIOGRAPHY

Adams, D. (2012). *Urban Planning and The Development Process*. Routledge. https://doi.org/10.4324/9780203857007

Aggarwal, C. C., & Zhai, C. (2012). A Survey of Text Clustering Algorithms. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 77–128). Springer US. https://doi.org/10.1007/978-1-4614-3223-4_4

Aichner, T., & Jacob, F. (2015). Measuring the Degree of Corporate Social Media Use. *International Journal of Market Research*, *57*(2), 257–276. https://doi.org/10.2501/IJMR-2015-018

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alessandretti, L., Sapiezynski, P., Lehmann, S., & Baronchelli, A. (2017). Multi-scale spatio-temporal analysis of human mobility. *PLOS ONE*, *12*(2), e0171686. https://doi.org/10.1371/journal.pone.0171686

Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., & Baronchelli, A. (2018). Evidence for a conserved quantity in human mobility. *Nature Human Behaviour*, *2*(7), 485–491. https://doi.org/10.1038/s41562-018-0364-x

Alsger, A., Assemi, B., Mesbah, M., & Ferreira, L. (2016). Validating and improving public transport origin–destination estimation algorithm using smart card fare data. *Transportation Research Part C: Emerging Technologies*, *68*, 490–506. https://doi.org/10.1016/j.trc.2016.05.004

Andrienko, G., Andrienko, N., Bosch, H., Ertl, T., Fuchs, G., Jankowski, P., & Thom, D. (2013). Thematic Patterns in Georeferenced Tweets through Space-Time Visual Analytics. *Computing in Science Engineering*, *15*(3), 72–82. https://doi.org/10.1109/MCSE.2013.70

Arribas-Bel, D., & Bakens, J. (2019). Use and validation of location-based services in urban research: An example with Dutch restaurants. *Urban Studies*, *56*(5), 868–884. https://doi.org/10.1177/0042098018779554

Ashalatha, R., Manju, V. S., & Zacharia, A. B. (2013). Mode Choice Behavior of Commuters in Thiruvananthapuram City. *Journal of Transportation Engineering*, *139*(5). https://trid.trb.org/view/1248034

Ayeh, J. K., Au, N., & Law, R. (2013). "Do We Believe in TripAdvisor?" Examining Credibility Perceptions and Online Travelers' Attitude toward Using User-Generated Content. *Journal of Travel Research*, *52*(4), 437–452. https://doi.org/10.1177/0047287512475217

Bagley, M. N., & Mokhtarian, P. L. (2002). The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *The Annals of Regional Science*, *36*(2), 279–297. https://doi.org/10.1007/s001680200083

Bamberg, S., Kühnel, S. M., & Schmidt, P. (1999). The Impact of General Attitude on Decisions. *Rationality and Society*, *11*(1), 5–25. https://doi.org/10.1177/104346399011001001

Barbier, G., & Liu, H. (2011). Data Mining in Social Media. In C. C. Aggarwal (Ed.), *Social Network Data Analytics* (pp. 327–352). Springer US. https://doi.org/10.1007/978-1-4419-8462-3_12

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv:1406.5823 [Stat]*.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., & Green, P. (2017). *lme4: Linear Mixed-Effects Models using “Eigen” and S4*. https://cran.r-project.org/web/packages/lme4/index.html

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. The MIT Press.

Ben-Akiva, M., & Lerman, S. R. (1979). Disaggregate Travel and Mobility-Choice Models and Measures of Accessibility. In *Behavioural Travel Modelling*. Routledge.

Bhat, C. R., & Pendyala, R. M. (2005). Modeling intra-household interactions and group decision-making. *Transportation*, *32*(5), 443–448. https://doi.org/10.1007/s11116-005-6789-x

Boarnet, M. G. (2011). A Broader Context for Land Use and Travel Behavior, and a Research Agenda. *Journal of the American Planning Association*, *77*(3), 197–213. https://doi.org/10.1080/01944363.2011.593483

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, *17*(3), 285–297. https://doi.org/10.1016/j.trc.2008.11.004

Börjesson, M., Hamilton, C. J., Näsman, P., & Papaix, C. (2015). Factors driving public support for road congestion reduction policies: Congestion charging, free public transport and more roads in Stockholm, Helsinki and Lyon. *Transportation Research Part A: Policy and Practice*, *78*, 452–462. https://doi.org/10.1016/j.tra.2015.06.008

Bureau of Transportation Statistics. (2018). *Transportation Statistics Annual Report*. Bureau of Transportation Statistics. https://www.bts.dot.gov/sites/bts.dot.gov/files/docs/browse-statistical-products-and-data/transportation-statistics-annual-reports/Preliminary-TSAR-Full-2018-a.pdf

Caragea, C., Squicciarini, A., Stehle, S., Neppalli, K., & Tapia, A. H. (2014). Mapping moods: Geo-mapped sentiment analysis during hurricane sandy. *ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management*, 642–651. https://pennstate.pure.elsevier.com/en/publications/mapping-moods-geo-mapped-sentiment-analysis-during-hurricane-sand

Carmon, N., & Fainstein, S. S. (Eds.). (2013). *Policy, planning, and people: Promoting justice in urban development* (1st ed). University of Pennsylvania Press.

Cascetta, E., Cartenì, A., & Montanino, M. (2013). A New Measure of Accessibility based on Perceived Opportunities. *Procedia - Social and Behavioral Sciences*, *87*, 117–132. https://doi.org/10.1016/j.sbspro.2013.10.598

Chakirov, A., & Erath, A. (2012). *Activity identification and primary location modelling based on smart card payment data for public transport* [Application/pdf]. 23 p. https://doi.org/10.3929/ETHZ-A-007328823

Chaniotakis, E., Antoniou, C., Aifadopoulou, G., & Dimitriou, L. (2017). Inferring Activities from Social Media Data. *Transportation Research Record: Journal of the Transportation Research Board*, *2666*(1), 29–37. https://doi.org/10.3141/2666-04

Charles-Edwards, E., Bell, M., & Corcoran, J. (2014). Integrating Undergraduate Fieldwork into the Study of Human Mobility. *Australian Geographer*, *45*(4), 505–519. https://doi.org/10.1080/00049182.2014.953734

Chatman, D. G. (2008). Deconstructing development density: Quality, quantity and price effects on household non-work travel. *Transportation Research Part A: Policy and Practice*, *42*(7), 1008–1030. https://doi.org/10.1016/j.tra.2008.02.003

Chen, B. Y., Wang, Y., Wang, D., Li, Q., Lam, W. H. K., & Shaw, S.-L. (2018). Understanding the Impacts of Human Mobility on Accessibility Using Massive Mobile Phone Tracking Data. *Annals of the American Association of Geographers*, *108*(4), 1115–1133. https://doi.org/10.1080/24694452.2017.1411244

Chen, B. Y., Yuan, H., Li, Q., Wang, D., Shaw, S.-L., Chen, H.-P., & Lam, W. H. K. (2017). Measuring place-based accessibility under travel time uncertainty. *International Journal of Geographical Information Science*, *31*(4), 783–804. https://doi.org/10.1080/13658816.2016.1238919

Chen, C., Bian, L., & Ma, J. (2014). From traces to trajectories: How well can we guess activity locations from mobile phone traces? *Transportation Research Part C: Emerging Technologies*, *46*, 326–337. https://doi.org/10.1016/j.trc.2014.07.001

Chen, C., Gong, H., Lawson, C., & Bialostozky, E. (2010). Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. *Transportation Research Part A: Policy and Practice*, *44*(10), 830–840. https://doi.org/10.1016/j.tra.2010.08.004

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, *68*, 285–299. https://doi.org/10.1016/j.trc.2016.04.005

Chen, J., Shaw, S.-L., Yu, H., Lu, F., Chai, Y., & Jia, Q. (2011). Exploratory data analysis of activity diary data: A space–time GIS approach. *Journal of Transport Geography*, *19*(3), 394–404. https://doi.org/10.1016/j.jtrangeo.2010.11.002

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, *19*(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0

Choe, Y., Kim, J. (Jamie), & Fesenmaier, D. R. (2017). Use of social media across the trip experience: An application of latent transition analysis. *Journal of Travel & Tourism Marketing*, *34*(4), 431–443. https://doi.org/10.1080/10548408.2016.1182459

Christiansen, P., Engebretsen, Ø., Fearnley, N., & Usterud Hanssen, J. (2017). Parking facilities and the built environment: Impacts on travel behaviour. *Transportation Research Part A: Policy and Practice*, *95*, 198–206. https://doi.org/10.1016/j.tra.2016.10.025

Chung, N., & Koo, C. (2015). The use of social media in travel information search. *Telematics and Informatics*, *32*(2), 215–229. https://doi.org/10.1016/j.tele.2014.08.005

City of Phoenix. (2020). *City of Phoenix Open Data*. https://www.phoenixopendata.com/

Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiratta, S. R., & González, M. C. (2015). Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations, and Opportunities. *Transportation Research Record: Journal of the Transportation Research Board*, *2526*(1), 126–135. https://doi.org/10.3141/2526-14

Collins, C., Hasan, S., & Ukkusuri, S. (2013). A Novel Transit Rider Satisfaction Metric: Rider Sentiments Measured from Online Social Media Data – National Center for Transit Research. *Journal of Public Transportation*, *16*(2).

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: Situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, *40*(2), 130–139. https://doi.org/10.1080/15230406.2013.777137

Crane, R. (2000). The Influence of Urban Form on Travel: An Interpretive Review. *Journal of Planning Literature*, *15*(1), 3–23. https://doi.org/10.1177/08854120022092890

Dabiri, S., & Heaslip, K. (2018). Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transportation Research Part C: Emerging Technologies*, *86*, 360–371. https://doi.org/10.1016/j.trc.2017.11.021

D'Andrea, E., Ducange, P., Lazzerini, B., & Marcelloni, F. (2015). Real-Time Detection of Traffic From Twitter Stream Analysis. *IEEE Transactions on Intelligent Transportation Systems*, *16*(4), 2269–2283. https://doi.org/10.1109/TITS.2015.2404431

Dang, T. A., Chiam, J., & Li, Y. (2018). A Comparative Study of Urban Mobility Patterns Using Large-Scale Spatio-Temporal Data. *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 572–579. https://doi.org/10.1109/ICDMW.2018.00089

DataReportal. (2021). *Global Social Media Stats*. https://datareportal.com/social-media-users

Diggle, P. (Ed.). (2013). *Analysis of longitudinal data* (Second Paperback Edition). Oxford University Press.

Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. *Proceedings of the International Conference on Web Search and Web Data Mining - WSDM '08*, 231. https://doi.org/10.1145/1341531.1341561

Dittmar, H., & Ohland, G. (2012). *The new transit town: Best practices in transit-oriented development*. Island Press.

DMR Business Statistics. (2021). *Yelp Statistics, User Counts and Facts (2021)*. DMR Business Statistics. https://expandedramblings.com/index.php/yelp-statistics/

Domencich, T. A., & McFadden, D. (1975). *Urban travel demand: A behavioral analysis: a Charles River Associates research study*. North-Holland Pub. Co. ; American Elsevier.

Dunkley, B., Helling, A., & Sawicki, D. S. (2004). Accessibility Versus Scale: Examining the Tradeoffs in Grocery Stores. *Journal of Planning Education and Research*, *23*(4), 387–401. https://doi.org/10.1177/0739456X04264890

Efthymiou, D., & Antoniou, C. (2012). Use of Social Media for Transport Data Collection. *Procedia - Social and Behavioral Sciences*, *48*, 775–785. https://doi.org/10.1016/j.sbspro.2012.06.1055

El-Geneidy, A. M., & Levinson, D. M. (2006). Access to Destinations: Development of Accessibility Measures. *University of Minnesota Digital Conservancy, Saint Paul, US*. https://hdl.handle.net/11299/638

ESRI. (2018). *How IDW works*.

Evans, L., & Saker, M. (2017). *Location-Based Social Media: Space, Time and Identity*. Springer.

Ewing, R., & Cervero, R. (2001). Travel and the Built Environment: A Synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, *1780*(1), 87–114. https://doi.org/10.3141/1780-10

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment: A Meta-Analysis. *Journal of the American Planning Association*, *76*(3), 265–294. https://doi.org/10.1080/01944361003766766

Federal Highway Administration. (2017). *2017 National Household Travel Survey*. https://nhts.ornl.gov/

Feinerer, I. (2018). *Introduction to the "tm" Package Text Mining in R*.

Feuerriegel, S., & Proellochs, N. (2018). *Package "SentimentAnalysis."*

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley Pub. Co.

Forrest, T., & Pearson, D. (2005). Comparison of Trip Determination Methods in Household Travel Surveys Enhanced by a Global Positioning System. *Transportation Research Record: Journal of the Transportation Research Board*, *1917*, 63–71. https://doi.org/10.3141/1917-08

Frank, L., Bradley, M., Kavage, S., Chapman, J., & Lawton, T. K. (2007). Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transportation*, *35*(1), 37–54. https://doi.org/10.1007/s11116-007-9136-6

Garber, M. D., Watkins, K. E., & Kramer, M. R. (2019). Comparing bicyclists who use smartphone apps to record rides with those who do not: Implications for representativeness and selection bias. *Journal of Transport & Health*, *15*, 100661. https://doi.org/10.1016/j.jth.2019.100661

Ge, Q., & Fukuda, D. (2016). Updating origin–destination matrices with aggregated data of GPS traces. *Transportation Research Part C: Emerging Technologies*, *69*, 291–312. https://doi.org/10.1016/j.trc.2016.06.002

Geurs, K. T., & van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: Review and research directions. *Journal of Transport Geography*, *12*(2), 127–140. https://doi.org/10.1016/j.jtrangeo.2003.10.005

Ghorpade, A., Pereira, F. C., Zhao, F., Zegras, C., & Ben-Akiva, M. (2015). *An Integrated Stop-Mode Detection Algorithm for Real World Smartphone-Based Travel Survey* (No. 15–6021). Article 15–6021. Transportation Research Board 94th Annual MeetingTransportation Research Board. https://trid.trb.org/view/1339538

Gonzalez, P. A., Weinstein, J. S., Barbeau, S. J., Labrador, M. A., Winters, P. L., Georggi, N. L., & Perez, R. (2010). Automating mode detection for travel behaviour analysis by using global positioning systems-enabled mobile phones and neural networks. *IET Intelligent Transport Systems*, *4*(1), 37. https://doi.org/10.1049/iet-its.2009.0029

Graham, S. (2002). *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition* (1st ed.). Routledge. https://doi.org/10.4324/9780203452202

Graham, S. (Ed.). (2010). *Disrupted cities: When infrastructure fails*. Routledge.

Grengs, J. (2010). Job accessibility and the modal mismatch in Detroit. *Journal of Transport Geography*, *18*(1), 42–54. https://doi.org/10.1016/j.jtrangeo.2009.01.012

Grengs, J. (2015). Nonwork Accessibility as a Social Equity Indicator. *International Journal of Sustainable Transportation*, *9*(1), 1–14. https://doi.org/10.1080/15568318.2012.719582

Griffin, G. P., & Jiao, J. (2015). Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *Journal of Transport & Health*, *2*(2), 238–247. https://doi.org/10.1016/j.jth.2014.12.001

Griffioen-Young, H. J., Janssen, H. J. W., Van Amelsfoort, D. J. C., & Langefeld, J. J. (2004). The psychology of parking. *In Proceedings of the ECOMM 2004 Conference.*

Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment*, *10*(6), 427–444. https://doi.org/10.1016/j.trd.2005.05.002

Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, *25*(2), 73–76. https://doi.org/10.1080/01944365908978307

Hardy, A., Hyslop, S., Booth, K., Robards, B., Aryal, J., Gretzel, U., & Eccleston, R. (2017). Tracking tourists' travel with smartphone-based GPS technology: A methodological discussion. *Information Technology & Tourism*, *17*(3), 255–274. https://doi.org/10.1007/s40558-017-0086-3

Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing - UrbComp '13*, 1. https://doi.org/10.1145/2505821.2505823

Heesch, K. C., & Langdon, M. (2016). The usefulness of GPS bicycle tracking data for evaluating the impact of infrastructure change on cycling behaviour: GPS bicycle tracking data in evaluating cycling behaviour. *Health Promotion Journal of Australia*, *27*(3), 222–229. https://doi.org/10.1071/HE16032

Hess, D. B. (2012). Walking to the bus: Perceived versus actual walking distance to bus stops for older adults. *Transportation*, *39*(2), 247–266. https://doi.org/10.1007/s11116-011-9341-1

Higuchi, K. (2012). *Quantitative Content Analysis or Text Mining by KH Coder*. https://sourceforge.net/p/khc/wiki/KWIC%20Concordance/

Higuchi, K. (2014). *KH Coder (Version 2.00 beta. 32)*. Japan: Koichi Higuchi.

Hoehne, C. G., Chester, M. V., Fraser, A. M., & King, D. A. (2019). Valley of the sun-drenched parking space: The growth, extent, and implications of parking infrastructure in Phoenix. *Cities*, *89*, 186–198. https://doi.org/10.1016/j.cities.2019.02.007

Hu, X., & Liu, H. (2012). Text Analytics in Social Media. In *Mining Text Data* (pp. 385–414). Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-3223-4_12

Huang, Q., & Wong, D. W. S. (2015). Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data. *Annals of the Association of American Geographers*, *105*(6), 1179–1197. https://doi.org/10.1080/00045608.2015.1081120

Huang, S., & Hsu, C. H. C. (2009). Effects of Travel Motivation, Past Experience, Perceived Constraint, and Attitude on Revisit Intention. *Journal of Travel Research*, *48*(1), 29–44. https://doi.org/10.1177/0047287508328793

Ignatow, G., & Mihalcea, R. (2016). *Text Mining: A Guidebook for the Social Sciences*. SAGE Publications.

Immergluck, D., & Smith, G. (2005). Measuring the Effect of Subprime Lending on Neighborhood Foreclosures: Evidence from Chicago. *Urban Affairs Review*, *40*(3), 362–389. https://doi.org/10.1177/1078087404271444

Inci, E. (2015). A review of the economics of parking. *Economics of Transportation*, *4*(1–2), 50–63. https://doi.org/10.1016/j.ecotra.2014.11.001

Infogroup. (2016). *Infogroup US Historical Business Data* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/PNOFKI

INRIX. (2017). *The Impact of Parking Pain in the US, UK and Germany*. http://www2.inrix.com/research-parking-2017

Iqbal, Md. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, *40*, 63–74. https://doi.org/10.1016/j.trc.2014.01.002

Jain, J., & Lyons, G. (2008). The gift of travel time. *Journal of Transport Geography*, *16*(2), 81–89. https://doi.org/10.1016/j.jtrangeo.2007.05.001

Jiang, S., Ferreira, J., & Gonzalez, M. C. (2017). Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. *IEEE Transactions on Big Data*, *3*(2), 208–219. https://doi.org/10.1109/TBDATA.2016.2631141

Jiang, Z., & Mondschein, A. (2019). Examining the effects of proximity to rail transit on travel to non-work destinations: Evidence from Yelp data for cities in North America and Europe. *Journal of Transport and Land Use*, *12*(1). https://doi.org/10.5198/jtlu.2019.1409

Jiang, Z., & Mondschein, A. (2021). Analyzing Parking Sentiment and its Relationship to Parking Supply and the Built Environment Using Online Reviews. *Journal of Big Data Analytics in Transportation*, *3*(1), 61–79. https://doi.org/10.1007/s42421-021-00036-1

Jockers, M. (2014). *Text Analysis with R for Students of Literature*. Springer International Publishing.

Jones, K. G., & Simmons, J. W. (1993). *Location, location, location: Analyzing the retail environment* (2nd ed.). Nelson Canada.

Jones, T., Baxter, M., & Khanduja, V. (2013). A quick guide to survey research. *The Annals of The Royal College of Surgeons of England*, *95*(1), 5–7. https://doi.org/10.1308/003588413X13511609956372

Kang, C., Gao, S., Lin, X., Xiao, Y., Yuan, Y., Liu, Y., & Ma, X. (2010). Analyzing and geo-visualizing individual human mobility patterns using mobile call records. *2010 18th International Conference on Geoinformatics*, 1–7. https://doi.org/10.1109/GEOINFORMATICS.2010.5567857

Kaplan, J. (2020). *predictrace: Predict the Race of a Given Surname Using Census Data* (1.2.1) [Computer software]. https://CRAN.R-project.org/package=predictrace

Keadle, S. K., McKinnon, R., Graubard, B. I., & Troiano, R. P. (2016). Prevalence and trends in physical activity among older adults in the United States: A comparison across three national surveys. *Preventive Medicine*, *89*, 37–43. https://doi.org/10.1016/j.ypmed.2016.05.009

Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). *Mastering the Information Age Solving Problems with Visual Analytics*. Eurographics Association. https://doi.org/10.2312/14803

Khanna, K. (2020). *Wru: Who are You? Bayesian Prediction of Racial Category Using Surname and Geolocation* (0.1-10) [Computer software]. https://CRAN.R-project.org/package=wru

Khetarpaul, S., Chauhan, R., Gupta, S. K., Subramaniam, L. V., & Nambiar, U. (2011). Mining GPS data to determine interesting locations. *Proceedings of the 8th International Workshop on Information Integration on the Web in Conjunction with WWW 2011 - IIWeb '11*, 1–6. https://doi.org/10.1145/1982624.1982632

Kimpton, A. (2017). A spatial analytic approach for classifying greenspace and comparing greenspace social equity. *Applied Geography*, *82*, 129–142. https://doi.org/10.1016/j.apgeog.2017.03.016

Kimpton, A. (2020). Explaining the Railheading Travel Behaviour with Home Location, Park 'N' Ride Characteristics, and the Built Environment to Strengthen Multimodalism. *Applied Spatial Analysis and Policy*. https://doi.org/10.1007/s12061-020-09361-4

King, D., Manville, M., & Shoup, D. (2007). The political calculus of congestion pricing. *Transport Policy*, *14*(2), 111–123. https://doi.org/10.1016/j.tranpol.2006.11.002

Kovacs-Györi, A., Ristea, A., Kolcsar, R., Resch, B., Crivellari, A., & Blaschke, T. (2018). Beyond Spatial Proximity—Classifying Parks and Their Visitors in London Based on Spatiotemporal and Sentiment Analysis of Twitter Data. *ISPRS International Journal of Geo-Information*, *7*(9), 378. https://doi.org/10.3390/ijgi7090378

Krippendorff, K. (2012). *Content Analysis: An Introduction to Its Methodology*. SAGE.

Kung, K. S., Greco, K., Sobolevsky, S., & Ratti, C. (2014). Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data. *PLoS ONE*, *9*(6), e96180. https://doi.org/10.1371/journal.pone.0096180

Kurkcu, A., Ozbay, K., & Morgul, E. F. (2016). Evaluating the Usability of Geo-located Twitter as a Tool for Human Activity and Mobility Patterns: A Case Study for New York City. *TRB 95th Annual Meeting Compendium of Papers*.

Kwan, M.-P., & Weber, J. (2003). Individual Accessibility Revisited: Implications for Geographical Analysis in the Twenty-first Century. *Geographical Analysis*, *35*(4), 341–353. https://doi.org/10.1111/j.1538-4632.2003.tb01119.x

Lee, J., Benjamin, S., & Childs, M. (2020). Unpacking the Emotions behind TripAdvisor Travel Reviews: The Case Study of Gatlinburg, Tennessee. *International Journal of Hospitality & Tourism Administration*, 1–18. https://doi.org/10.1080/15256480.2020.1746219

Li, W., Li, Y., Ban, X. (Jeff), Deng, H., Shu, H., & Xie, D. (2018). Exploring the Relationships between the Non-Work Trip Frequency and Accessibility Based on Mobile Phone Data. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(42), 91–102. https://doi.org/10.1177/0361198118774170

Lin, J., & Mele, C. (Eds.). (2013). *The urban sociology reader* (Second edition). Routledge, Taylor & Francis Group.

Lin, T., Wang, D., & Guan, X. (2017). The built environment, travel attitude, and travel behavior: Residential self-selection or residential determination? *Journal of Transport Geography*, *65*, 111–122. https://doi.org/10.1016/j.jtrangeo.2017.10.004

Litman, T. (2006). *Parking management best practices*. American Planning Association.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Lockwood, P., & Demetsky, M. (1994). *Methodology for nonwork travel analysis in suburban communities*. Virginia Transportation Research Council.

Ma, X., Wu, Y.-J., Wang, Y., Chen, F., & Liu, J. (2013). Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, *36*, 1–12. https://doi.org/10.1016/j.trc.2013.07.010

Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., & Woodward, J. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science*, *27*(4), 662–684. https://doi.org/10.1080/13658816.2012.696649

Manovich, L. (2012). *Trending: The Promises and the Challenges of Big Social Data*. 460–475. https://doi.org/10.5749/minnesota/9780816677948.003.0047

Manville, M., & Shoup, D. (2005). Parking, people, and cities. *Journal of Urban Planning and Development*, *131*(4), 233–245.

Markov, Z., & Larose, D. T. (2007). *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage*.

Marra, A. D., Becker, H., Axhausen, K. W., & Corman, F. (2019). Developing a passive GPS tracking system to study long-term travel behavior. *Transportation Research Part C: Emerging Technologies*, *104*, 348–368. https://doi.org/10.1016/j.trc.2019.05.006

Marsden, G. (2006). The evidence base for parking policies—A review. *Transport Policy*, *13*(6), 447–457. https://doi.org/10.1016/j.tranpol.2006.05.009

McCulloch, C. E., & Neuhaus, J. M. (2005). Generalized Linear Mixed Models. In P. Armitage & T. Colton (Eds.), *Encyclopedia of Biostatistics* (p. b2a10021). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470011815.b2a10021

Mcculloch, C., & Neuhaus, J. (2001). *Generalized linear mixed models*. John Wiley & Sons, Ltd.

McFadden, D. (2000). Disaggregate behavioral travel demand's RUM side: A 30-year retrospective. *INTERNATIONAL CONFERENCE ON TRAVEL BEHAVIOUR RESEARCH, 9TH, 2000, GOLD COAST, QUEENSLAND, AUSTRALIA, VOL 1*. https://trid.trb.org/view/788681

McGuckin, N., & Fucci, A. (2018). *Summary of Travel Trends – 2017 National Household Travel Survey*. US Department of Transportation, Federal Highway Administration. https://nhts.ornl.gov/assets/2017_nhts_summary_travel_trends.pdf

McLean, D. J., & Skowron Volponi, M. A. (2018). trajr: An R package for characterisation of animal trajectories. *Ethology*, *124*(6), 440–448. https://doi.org/10.1111/eth.12739

Mennis, J., & Guo, D. (2009). Spatial data mining and geographic knowledge discovery—An introduction. *Computers, Environment and Urban Systems*, *33*(6), 403–408. https://doi.org/10.1016/j.compenvurbsys.2009.11.001

Millard-Ball, A., Weinberger, R. R., & Hampshire, R. C. (2014). Is the curb 80% full or 20% empty? Assessing the impacts of San Francisco's parking pricing experiment. *Transportation Research Part A: Policy and Practice*, *63*, 76–92. https://doi.org/10.1016/j.tra.2014.02.016

Miller, H. (2007). Place-Based versus People-Based Geographic Information Science: Place-based versus people-based geographic information science. *Geography Compass*, *1*(3), 503–535. https://doi.org/10.1111/j.1749-8198.2007.00025.x

Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLoS ONE*, *8*(5), e64417. https://doi.org/10.1371/journal.pone.0064417

Molloy, J., & Moeckel, R. (2017). Improving Destination Choice Modeling Using Location-Based Big Data. *ISPRS International Journal of Geo-Information*, *6*(9), 291. https://doi.org/10.3390/ijgi6090291

Mondschein, A. (2015). Five-star transportation: Using online activity reviews to examine mode choice to non-work destinations. *Transportation*, *42*(4), 707–722. https://doi.org/10.1007/s11116-015-9600-7

Mondschein, A., King, D. A., Hoehne, C., Jiang, Z., & Chester, M. (2020). Using social media to evaluate associations between parking supply and parking sentiment. *Transportation Research Interdisciplinary Perspectives*, 100085. https://doi.org/10.1016/j.trip.2019.100085

Monzon, A. (2015). Smart cities concept and challenges: Bases for the assessment of smart city projects. *2015 International Conference on Smart Cities and Green ICT Systems (SMARTGREENS)*, 1–11.

Mora, H., Gilart-Iglesias, V., Pérez-del Hoyo, R., & Andújar-Montoya, M. (2017). A Comprehensive System for Monitoring Urban Accessibility in Smart Cities. *Sensors*, *17*(8), 1834. https://doi.org/10.3390/s17081834

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, *37*(1/2), 17. https://doi.org/10.2307/2332142

Mullen, L. (2020). *gender: Predict Gender from Names Using Historical Data* (R package version 0.5.4) [Computer software]. https://github.com/ropensci/gender

Munar, A. M., & Jacobsen, J. Kr. S. (2013). Trust and Involvement in Tourism Social Media and Web-Based Travel Information Sources. *Scandinavian Journal of Hospitality and Tourism*, *13*(1), 1–19. https://doi.org/10.1080/15022250.2013.764511

Nassir, N., Hickman, M., Malekzadeh, A., & Irannezhad, E. (2016). A utility-based travel impedance measure for public transit network accessibility. *Transportation Research Part A: Policy and Practice*, *88*, 26–39. https://doi.org/10.1016/j.tra.2016.03.007

Nelson, D., & Niles, J. (1999). Essentials for Transit-Oriented Development Planning: Analysis of Non-Work Activity Patterns and a Method for Predicting Success. *Proceedings of the 7th TRB Conference on the Application of Transportation Planning Methods*.

Nelson, D., Niles, J., & Hibshoosh, A. (2001). *A New Planning Template for Transit-Oriented Development*. Mineta Transportation Institute.

Neuburger, H. (1971). User Benefit in the Evaluation of Transport and Land Use Plans. *Journal of Transport Economics and Policy*, *5*(1), 52–75.

Neutens, T., Schwanen, T., Witlox, F., & De Maeyer, P. (2010). Equity of Urban Service Delivery: A Comparison of Different Accessibility Measures. *Environment and Planning A: Economy and Space*, *42*(7), 1613–1635. https://doi.org/10.1068/a4230

Nikšič, M., Campagna, M., Massa, P., Caglioni, M., & Nielsen, T. (2017). Opportunities for Volunteered Geographic Information Use in Spatial Planning. In *Mapping and the Citizen Sensor* (pp. 327–349). Ubiquity Press. https://www.ubiquitypress.com/site/chapters/10.5334/bbf.n/

Obar, J. A., & Wildman, S. (2015). Social media definition and the governance challenge: An introduction to the special issue. *Telecommunications Policy*, *39*(9), 745–750. https://doi.org/10.1016/j.telpol.2015.07.014

Ord, J. K., & Getis, A. (2010). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, *27*(4), 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x

Pandhe, A., & March, A. (2012). Parking availability influences on travel mode: Melbourne CBD offices. *Australian Planner*, *49*(2), 161–171. https://doi.org/10.1080/07293682.2011.616177

Parkany, E., Gallagher, R., & Viveiros, P. (2004). Are Attitudes Important in Travel Choice? *Transportation Research Record: Journal of the Transportation Research Board*, *1894*(1), 127–139. https://doi.org/10.3141/1894-14

Pearson, K. (1900). X. *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157–175. https://doi.org/10.1080/14786440009463897

Pelletier, M.-P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, *19*(4), 557–568. https://doi.org/10.1016/j.trc.2010.12.003

Pew Research Center. (2019). *Pew Research Center: Mobile Fact Sheet*. Pew Research Center. https://www.pewresearch.org/internet/fact-sheet/mobile/

Phoenix City Council. (2015). *PlanPHX General Plan Update*.

Polzin, S., Chu, X., & Rey, J. (1999). Mobility and Mode Choice of People of Color for Non-Work Travel. *Transportation Research Board. Personal Travel: The Long and Short of It*. /paper/MOBILITY-AND-MODE-CHOICE-OF-PEOPLE-OF-COLOR-FOR-Polzin-Chu/2ceacd1a1b3b104bbac4b4981497baea3f1ef095

Purifoye, G. Y. (2015). Nice–Nastiness and Other Raced Social Interactions on Public Transport Systems. *City & Community*, *14*(3), 286–310. https://doi.org/10.1111/cico.12116

Quantcast. (2017). *Yelp Audience Insights and Demographic Analytics.* https://www.quantcast.com/yelp.com/demographics/WEB?country=US

Rahim Taleqani, A., Hough, J., & Nygard, K. E. (2019). Public Opinion on Dockless Bike Sharing: A Machine Learning Approach. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(4), 195–204. https://doi.org/10.1177/0361198119838982

Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, *75*, 197–211. https://doi.org/10.1016/j.trc.2016.12.008

Rathore, M. M., Paul, A., Hong, W.-H., Seo, H., Awan, I., & Saeed, S. (2018). Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data. *Sustainable Cities and Society*, *40*, 600–610. https://doi.org/10.1016/j.scs.2017.12.022

Reddy, S., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2008). Determining transportation mode on mobile phones. *2008 12th IEEE International Symposium on Wearable Computers*, 25–28. https://doi.org/10.1109/ISWC.2008.4911579

Ren, F., Tong, D., & Kwan, M. (2014). Space–time measures of demand for service: Bridging location modelling and accessibility studies through a time-geographic framework. *Geografiska Annaler: Series B, Human Geography*, *96*(4), 329–344. https://doi.org/10.1111/geob.12055

Ripley, B., & Venables, W. (2021). *nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models* (7.3-16) [Computer software]. https://CRAN.R-project.org/package=nnet

Roberts, H., Sadler, J., & Chapman, L. (2019). The value of Twitter data for determining the emotional responses of people to urban green spaces: A case study and critical evaluation. *Urban Studies*, *56*(4), 818–835. https://doi.org/10.1177/0042098017748544

Rose, M. H. (1990). *Interstate: Express highway politics, 1939-1989* (Rev. ed). University of Tennessee Press.

Rybarczyk, G., Banerjee, S., Starking-Szymanski, M. D., & Shaker, R. R. (2018). Travel and us: The impact of mode share on sentiment using geo-social media and GIS. *Journal of Location Based Services*, *12*(1), 40–62. https://doi.org/10.1080/17489725.2018.1468039

Sabab Zulfiker, Md., Kabir, N., Ali, H. M., Haque, M. R., Akter, M., & Uddin, M. S. (2020). Sentiment Analysis Based on Users’ Emotional Reactions About Ride-Sharing Services on Facebook and Twitter. In M. S. Uddin & J. C. Bansal (Eds.), *Proceedings of International Joint Conference on Computational Intelligence* (pp. 397–408). Springer. https://doi.org/10.1007/978-981-15-3607-6_32

Sapiezynski, P., Stopczynski, A., Gatej, R., & Lehmann, S. (2015). Tracking Human Mobility Using WiFi Signals. *PLOS ONE*, *10*(7), e0130824. https://doi.org/10.1371/journal.pone.0130824

Sarriera, J. M., Álvarez, G. E., Blynn, K., Alesbury, A., Scully, T., & Zhao, J. (2017). To Share or Not to Share: Investigating the Social Aspects of Dynamic Ridesharing. *Transportation Research Record: Journal of the Transportation Research Board*, *2605*(1), 109–117. https://doi.org/10.3141/2605-11

Saxena, A., Chaturvedi, K. R., & Rakesh, S. (2018). Analysing Customers Reactions on Social Media Promotional Campaigns: A Text-mining Approach. *Paradigm*, *22*(1), 80–99. https://doi.org/10.1177/0971890718759163

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Schoder, D., Putzke, J., Metaxas, P. T., Gloor, P., & Fischbach, K. (2014). Information Systems for "Wicked Problems"—Research at the Intersection of Social Media and Collective Intelligence. *Business & Information Systems Engineering*, *6*(1), 3–10.

Scott, D., & Horner, M. (2008). Examining the role of urban form In shaping people's accessibility to opportunities: An exploratory spatial data analysis. *Journal of Transport and Land Use*, *1*(2). https://doi.org/10.5198/jtlu.v1i2.25

Sedera, D., Lokuge, S., Atapattu, M., & Gretzel, U. (2017). Likes—The key to my happiness: The moderating effect of social influence on travel experience. *Information & Management*, *54*(6), 825–836. https://doi.org/10.1016/j.im.2017.04.003

Sekar, A., Chen, R. B., Cruzat, A., & Nagappan, M. (2017). Digital Narratives of Place: Learning About Neighborhood Sense of Place and Travel Through Online Responses. *Transportation Research Record: Journal of the Transportation Research Board*, *2666*(1), 10–18. https://doi.org/10.3141/2666-02

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (Muki). (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, *31*(1), 139–167. https://doi.org/10.1080/13658816.2016.1189556

Shin, D., Aliaga, D., Tunçer, B., Arisona, S. M., Kim, S., Zünd, D., & Schmitt, G. (2015). Urban sensing: Using smartphones for transportation mode classification. *Computers, Environment and Urban Systems*, *53*, 76–86. https://doi.org/10.1016/j.compenvurbsys.2014.07.011

Shoup, D. C. (2006). Cruising for parking. *Transport Policy*, *13*(6), 479–486. https://doi.org/10.1016/j.tranpol.2006.05.005

Shoup, D. C. (2011). *The high cost of free parking* (Updated). Planners Press, American Planning Association.

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., & Fotheringham, A. S. (2016). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, *30*(5), 881–906. https://doi.org/10.1080/13658816.2015.1100731

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In A. Sattar & B. Kang (Eds.), *AI 2006: Advances in Artificial Intelligence* (Vol. 4304, pp. 1015–1021). Springer Berlin Heidelberg. https://doi.org/10.1007/11941439_114

Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, *6*(10), 818–823. https://doi.org/10.1038/nphys1760

Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2019). Quantifying international human mobility patterns using Facebook Network data. *PLOS ONE*, *14*(10), e0224134. https://doi.org/10.1371/journal.pone.0224134

Stanley, J., Hensher, D. A., Stanley, J., Currie, G., Greene, W. H., & Vella-Brodrick, D. (2011). Social Exclusion and the Value of Mobility. *Journal of Transport Economics and Policy*, *45*(2), 197–222.

Statista. (2019). *Statista Digital Market Outlook: Number of smartphone users in the United States from 2018 to 2024 (in millions)*. Statista. https://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/

Statistics Canada. (2016). *Data products, 2016 Census.*

Stead, D., & Marshall, S. (2001). The Relationships between Urban Form and Travel Patterns. An International Review and Evaluation. *European Journal of Transport and Infrastructure Research*, Vol 1 No 2 (2001). https://doi.org/10.18757/EJTIR.2001.1.2.3497

Stevens, M. R. (2017). Does Compact Development Make People Drive Less? *Journal of the American Planning Association*, *83*(1), 7–18. https://doi.org/10.1080/01944363.2016.1240044

Stone, P. J., Bales, R. F., Namenwirth, J. Z., & Ogilvie, D. M. (2007). The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science*, *7*(4), 484–498. https://doi.org/10.1002/bs.3830070412

Stopher, P., FitzGerald, C., & Zhang, J. (2008). Search for a global positioning system device to measure person travel. *Transportation Research Part C: Emerging Technologies*, *16*(3), 350–369. https://doi.org/10.1016/j.trc.2007.10.002

Stopher, P., Jiang, Q., & FitzGerald, C. (2005). Processing GPS data from travel surveys. *Australasian Transport Research Forum (ATRF), 28th, 2005, Sydney, New South Wales, Australia*, *28*.

Stopher, P. R., & Greaves, S. P. (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, *41*(5), 367–381. https://doi.org/10.1016/j.tra.2006.09.005

Stopher, P. R., Shen, L., Liu, W., & Ahmed, A. (2015). The Challenge of Obtaining Ground Truth for GPS Processing. *Transportation Research Procedia*, *11*, 206–217. https://doi.org/10.1016/j.trpro.2015.12.018

Susha, I., Janssen, M., & Verhulst, S. (2017). Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development. *Hawaii International Conference on System Sciences 2017 (HICSS-50)*. https://aisel.aisnet.org/hicss-50/eg/open_data_in_government/4

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267–307. https://doi.org/10.1162/COLI_a_00049

Tang, J., Liu, F., Wang, Y., & Wang, H. (2015). Uncovering urban human mobility from large scale taxi GPS data. *Physica A: Statistical Mechanics and Its Applications*, *438*, 140–153. https://doi.org/10.1016/j.physa.2015.06.032

Thomas, J. J., & Cook, K. A. (2006). A visual analytics agenda. *IEEE Computer Graphics and Applications*, *26*(1), 10–13. https://doi.org/10.1109/MCG.2006.5

Torre-Bastida, A. I., Del Ser, J., Laña, I., Ilardia, M., Bilbao, M. N., & Campos-Cordobés, S. (2018). Big Data for transportation and mobility: Recent advances, trends and challenges. *IET Intelligent Transport Systems*, *12*(8), 742–755. https://doi.org/10.1049/iet-its.2018.5188

Train, K. (1986). *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand*. MIT Press.

Train, K. E. (2001). *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511805271

Traunmueller, M. W., Johnson, N., Malik, A., & Kontokosta, C. E. (2018). Digital footprints: Using WiFi probe and locational data to analyze human mobility trajectories in cities. *Computers, Environment and Urban Systems*, *72*, 4–12. https://doi.org/10.1016/j.compenvurbsys.2018.07.006

Tsai, C.-W., Lai, C.-F., Chao, H.-C., & Vasilakos, A. V. (2015). Big data analytics: A survey. *Journal of Big Data*, *2*(1), 21. https://doi.org/10.1186/s40537-015-0030-3

Tumlin, J. (2012). *Sustainable transportation planning: Tools for creating vibrant, healthy, and resilient communities*. Wiley.

Tung, E. (2015, September 2). *Automatically Categorizing Yelp Businesses*.

Twitter. (2018). *Twitter API*. https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets

United Nations. (2018). *Population Division*. World Urbanization Prospects: The 2018 Revision. https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html

United States Environmental Protection Agency. (2014). *Smart Location Database*. https://edg.epa.gov/data/Public/OP/SLD/SmartLocationDB.zip

U.S. Census Bureau. (2014). *Understanding Geographic Relationships: Counties, Places, Tracts and More*. The United States Census Bureau. https://www.census.gov/newsroom/blogs/random-samplings/2014/07/understanding-geographic-relationships-counties-places-tracts-and-more.html

U.S. Census Bureau. (2018). *U.S. Census Bureau: American community survey, 2018 5-year estimates.*

USDOT. (1995). *Nationwide Personal Transportation Survey*. Federal Highway Administration. https://www.fhwa.dot.gov/ohim/sept97.html

Van Acker, V., Van Wee, B., & Witlox, F. (2010). When Transport Geography Meets Social Psychology: Toward a Conceptual Model of Travel Behaviour. *Transport Reviews*, *30*(2), 219–240. https://doi.org/10.1080/01441640902943453

Vazquez-Prokopec, G. M., Bisanzio, D., Stoddard, S. T., Paz-Soldan, V., Morrison, A. C., Elder, J. P., Ramirez-Paredes, J., Halsey, E. S., Kochel, T. J., Scott, T. W., & Kitron, U. (2013). Using GPS Technology to Quantify Human Mobility, Dynamic Contacts and Infectious Disease Dynamics in a Resource-Poor Urban Environment. *PLoS ONE*, *8*(4), e58802. https://doi.org/10.1371/journal.pone.0058802

Verplanken, B., & Aarts, H. (1999). Habit, Attitude, and Planned Behaviour: Is Habit an Empty Construct or an Interesting Case of Goal-directed Automaticity? *European Review of Social Psychology*, *10*(1), 101–134. https://doi.org/10.1080/14792779943000035

Vinithra, S. N., Selvan, S. J. A., Kumar, M. A., & Soman, K. P. (2015). Simulated and Self-Sustained Classification of Twitter Data based on its Sentiment. *Indian Journal of Science and Technology*, *8*(24). https://doi.org/10.17485/ijst/2015/v8i24/80205

Vishnu, B., & Srinivasan, K. K. (2013). Tour-based Departure Time Models for Work and Non-work Tours of Workers. *Procedia - Social and Behavioral Sciences*, *104*, 630–639. https://doi.org/10.1016/j.sbspro.2013.11.157

Walle, S., & Steenberghen, T. (2006). Space and time related determinants of public transport use in trip chains. *Transportation Research Part A: Policy and Practice*, *40*(2), 151–162. https://doi.org/10.1016/j.tra.2005.05.001

Wang, H., Calabrese, F., Di Lorenzo, G., & Ratti, C. (2010). Transportation mode inference from anonymized and aggregated mobile phone call detail records. *13th International IEEE Conference on Intelligent Transportation Systems*, 318–323. https://doi.org/10.1109/ITSC.2010.5625188

Wang, Q., Phillips, N. E., Small, M. L., & Sampson, R. J. (2018). Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences*, *115*(30), 7735–7740. https://doi.org/10.1073/pnas.1802537115

Wee, B. V., Wee, B. V., & Geurs, K. (2011). Discussing Equity and Social Exclusion in Accessibility Evaluations. *European Journal of Transport and Infrastructure Research*, Vol 11 No 4 (2011). https://doi.org/10.18757/EJTIR.2011.11.4.2940

Weinberger, R., Kaehny, J., & Rufo, M. (2010). U.S. parking policies: An overview of management strategies. *The TRIS and ITRD Database*.

Wijayaratna, S., & Wijayaratna, K. P. (2016). Quantifying the Impact of On-Street Parking on Road Capacity: A Case Study of Sydney Arterial Roads. *TRB 95th Annual Meeting Compendium of Papers*.

Williams, N. E., Thomas, T. A., Dunbar, M., Eagle, N., & Dobra, A. (2015). Measures of Human Mobility Using Mobile Phone Records Enhanced with GIS Data. *PLOS ONE*, *10*(7), e0133630. https://doi.org/10.1371/journal.pone.0133630

Woldeamanuel, M. G. (2016). *Concepts in urban transportation planning: The quest for mobility, sustainability and quality of life*. McFarland & Company, Inc., Publishers.

World Bank. (2020). *World Development Indicators*. World Bank. worldbank.org

Wu, C., Yang, Z., Xu, Y., Zhao, Y., & Liu, Y. (2015). Human Mobility Enhances Global Positioning Accuracy for Mobile Phone Localization. *IEEE Transactions on Parallel and Distributed Systems*, *26*(1), 131–141. https://doi.org/10.1109/TPDS.2014.2308225

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 97–107. https://doi.org/10.1109/TKDE.2013.109

Xia, T., Song, X., Zhang, H., Song, X., Kanasugi, H., & Shibasaki, R. (2019). Measuring spatio-temporal accessibility to emergency medical services through big GPS data. *Health & Place*, *56*, 53–62. https://doi.org/10.1016/j.healthplace.2019.01.012

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, *31*(2), 179–188. https://doi.org/10.1016/j.tourman.2009.02.016

X-Mode. (2021). *X-Mode Social: About Us*. https://xmode.io/about-us/

Yang, C., Xiao, M., Ding, X., Tian, W., Zhai, Y., Chen, J., Liu, L., & Ye, X. (2019). Exploring human mobility patterns using geo-tagged social media data at the group level. *Journal of Spatial Science*, *64*(2), 221–238. https://doi.org/10.1080/14498596.2017.1421487

Yang, F., Jin, P. J., Cheng, Y., Zhang, J., & Ran, B. (2015). Origin-Destination Estimation for Non-Commuting Trips Using Location-Based Social Networking Data. *International Journal of Sustainable Transportation*, *9*(8), 551–564. https://doi.org/10.1080/15568318.2013.826312

Yap, M., Cats, O., & van Arem, B. (2020). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A: Transport Science*, *16*(1), 23–42. https://doi.org/10.1080/23249935.2018.1537319

Ye, X., Pendyala, R. M., & Gottardi, G. (2007). An exploration of the relationship between mode choice and complexity of trip chaining patterns. *Transportation Research Part B: Methodological*, *41*(1), 96–113. https://doi.org/10.1016/j.trb.2006.03.004

Yelp. (2018). *API 2.0: All Category List. Yelp for Developers.*

Yelp. (2019a). Yelp Dataset Challenge: Round 13. *Yelp.Com*. https://www.yelp.com/dataset_challenge

Yelp. (2019b). *Yelp Factsheet*. https://www.yelp.com/factsheet

Zhang, X., Zhou, Y., Ma, Y., Chen, B.-C., Zhang, L., & Agarwal, D. (2016). *GLMix: Generalized Linear Mixed Models For Large-Scale Response Prediction*. 363–372. https://doi.org/10.1145/2939672.2939684

Zhao, K., Tarkoma, S., Liu, S., & Vo, H. (2016). Urban human mobility data mining: An overview. *2016 IEEE International Conference on Big Data (Big Data)*, 1911–1920. https://doi.org/10.1109/BigData.2016.7840811

Zhou, C., Jia, H., Gao, J., Yang, L., Feng, Y., & Tian, G. (2017). Travel mode detection method based on big smartphone global positioning system tracking data. *Advances in Mechanical Engineering*, *9*(6), 168781401770813. https://doi.org/10.1177/1687814017708134

Zhou, X., Wang, M., & Li, D. (2017). From stay to play – A travel planning tool based on crowdsourcing user-generated contents. *Applied Geography*, *78*, 1–11. https://doi.org/10.1016/j.apgeog.2016.10.002

Zhou, Y., Lau, B. P. L., Yuen, C., Tuncer, B., & Wilhelm, E. (2018). Understanding Urban Human Mobility through Crowdsensed Data. *IEEE Communications Magazine*, *56*(11), 52–59. https://doi.org/10.1109/MCOM.2018.1700569

Zhu, L., Yu, F. R., Wang, Y., Ning, B., & Tang, T. (2019). Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, *20*(1), 383–398. https://doi.org/10.1109/TITS.2018.2815678