FINAL REPORT

Project B

December 2019

Technology Influence on Travel Demand and Behaviors

Dr. Virginia P. Sisiopiku | University of Alabama at Birmingham

Dr. Mohammed Hadi | Florida International University

Dr. Noreen McDonald | University of North Carolina at Chapel Hill

Dr. Ruth Steiner | University of Florida

Dr. Ossama E. Ramadan | University of Alabama at Birmingham

STRIDE | Southeastern Transportation Research, Innovation, Development and Education Center

UF | Transportation Institute UNIVERSITY of FLORIDA

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGEMENT OF SPONSORSHIP AND STAKEHOLDERS

This work was sponsored by a grant from the Southeastern Transportation Research, Innovation, Development, and Education Center (STRIDE) and the Alabama Department of Transportation (ALDOT).

Funding Agreement Number - 69A3551747104

LIST OF AUTHORS

Lead PI:

Virginia P. Sisiopiku, Ph.D.
University of Alabama at Birmingham
vsisiopi@uab.edu
ORCID 0000-0003-4262-8990

Co-PIs:

Mohammed Hadi, Ph.D.
Florida International University (FIU)
hadim@fiu.edu
ORCID 0000-0003-2233-8283

Noreen McDonald, Ph.D.
University of North Carolina at Chapel Hill
noreen@unc.edu
ORCID 0000-0002-4854-7035

Ruth Steiner, Ph.D.
University of Florida
rsteiner@ufl.edu
ORCID 0000-0001-7276-3742

Ossama E. Ramadan, Ph.D.
University of Alabama at Birmingham
oramadan@uab.edu
ORCID 0000-0002-2765-7278

Additional Researchers:

Allie Thomas, Ph.D.
University of North Carolina at Chapel Hill
alainna@email.unc.edu
ORCID 0000-0002-8892-2493

Da Yan, Ph.D.
University of Alabama at Birmingham
yanda@uab.edu
ORCID 0000-0002-4653-0408

Furat Salman, MS
University of Alabama at Birmingham
furat@uab.edu

Sahila Sarjana
University of Alabama at Birmingham
sahila17@uab.edu

Guimu Guo
University of Alabama at Birmingham
guimuguo@uab.edu

Taniya Sultana
University of Alabama at Birmingham
tans834@uab.edu

Jesse Anderson
University of Florida
jdf65@ufl.edu

Gedaliah (Gadi) Dreyfuss
University of Florida
gadidreyfuss@ufl.edu

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Over the last decade, the popularity of Transportation Network Companies (TNCs) as a mode of travel has been increasing at a steady pace, even in medium size cities. However, the determinants that influence transportation users to adopt TNCs as a preferred mode choice are still not well understood, nor are the impacts of such preferences on their travel patterns and transportation network operation.

This study used a mixed methods approach to examine and document technology influence on travelers' attitudes, preferences, and choices and their potential impact on transportation services in the Southeast. More specifically, the study investigated the influence of Transportation Network Companies (TNCs) such as Uber and Lyft, on travelers' behavior in two medium size cities in the Southeast based on three distinct but interrelated case studies, in addition to a comprehensive literature review and synthesis.

The first case study was a survey of 600 millennials (born 1981-1996) in North Carolina that was used to understand their travel behavior in a market where ride-hailing services have taken off in terms of use and coverage in the past 5 years. The results from the study showed that most millennials surveyed had used ridehailing services—with 66% having used Lyft, Uber, or both; many on a fairly regular basis. Detailed analysis also showed significant differences in use or familiarity amongst ethnic or racial groups. The case study findings demonstrate that even in states with small urban areas and lower densities, millennials are aware of and are taking advantage of ridehailing, carsharing, and ridesharing services.

The second case study focused on factors that influence transportation users to select TNCs (such as Uber/Lyft) for completing typical day trips. A questionnaire survey was developed and used to survey over 450 transportation users in the Birmingham Metro area on their current travel preferences and practices and document their attitudes toward TNC use as a travel mode of choice. The survey participants provided detailed trip information for a typical 24-hr day along with demographic data and travel preference information. The results revealed that Birmingham travelers are aware of TNC services and 45% of those surveyed have used TNC services. The most important determinants that make TNCs a preferable mode to Birmingham travelers included convenience of use and reduction of concerns for traffic safety (especially for late night trips to bars and eating establishments). Lack of parking availability at the destination was also listed as a reason for selecting TNCs as a mode of travel along with lack of vehicle availability. Overall, the study provided valuable insights on the leading reasons and conditions that drive people towards the use of TNC services in the Birmingham Metro Area.

The third case study evaluated the feasibility of building an agent-based simulation model of the Birmingham Metro Area in order to study the impact of shifts in travel demand due to applications of shared-use economy on local and regional congestion. Due to the fact that commonly used traffic simulation models lack the ability to simulate shared modes in detail, the Birmingham prototype model was developed using the Multi-Agent Transport Simulation (MATSim) modeling platform and was a major undertaking in itself. The case study identified

data needs and requirements for model development and adopted a data-driven approach for addressing data sparsity issues encountered. Future research by the research team will extend this work by expanding the prototype Birmingham MATSim model to incorporate public transit and quantifying the impacts from the integration of TNCs and transit on travel demand and congestion.

The overall findings documented in this report provide fresh insights on the links between technology and driving choices among transportation users in the Southeast in the presence of TNC services. Additionally, the study developed a prototype agent-based model that will provide the basis quantify influences of these technologies on urban and regional congestion.

This study is also significant for providing transportation agencies the means to better-plan mobility as a service (MaaS) where car/ridesharing platforms are active. Moreover, study findings can inform TNC- and other shared-mode services about the needs and opportunities of the local market and enable them to better understand how the travel behavior, mode-choice, and travel demand might affect the use of TNCs in the future.

Keywords: Transportation Network Companies, travel behavior, mode choice, survey.

1.0 INTRODUCTION

Advanced technologies influence travel behaviors; however, there is currently no consensus on what this influence will be for recently available and upcoming technologies and new transportation modes and options. For example, it is still not clear how travelers respond to smartphone-based traveler information services and how such information may influence their mode-choice or travel behavior. Understanding these attitudes becomes even more important when traveler information is utilized to plan, design, and manage transportation as a service.

The rapid technological developments in the 21st century created new opportunities for shared-use economy applications around the globe. Examples of existing technology-based car-sharing and ridesharing solutions in USA include:

1. Ride-hailing (or dynamic car-sharing) platforms like *Uber*, *Lyft*;
2. Organization-based car-sharing platforms like *Zipcar*, *Enterprise CarShare*, *Hertz On Demand*, *Car2Go*, and *DriveNow*;
3. Peer-to-peer car-sharing platforms like *Getaround* and *Turo*,
4. Ridesharing dedicated social networks like *Carticipate*; and
5. Secure ridesharing for companies and universities like *Zimride*.

Despite the recent growth of technology-based ridesharing options in many markets across the U.S., the impacts of shared-use economy on urban congestion and transportation system efficiency is still not systematically measured and far from being well understood. Apparent increase in market penetration by Transportation Network Companies (TNCs) offering dynamic ridesharing platforms like Uber and Lyft, raised the question of whether or not such platforms are any different from hailing a taxi with respect to operational impacts. Sun and Edara (2015) examined (Sun & Edara, 2015) this question independently from platform and/or jurisdiction and reached a negative answer based on their examination of the history of ridesharing and modal characteristics. Nevertheless, their analysis was qualitative and did not provide evidence-based conclusions as to the anticipated contributions to congestion relief or increasing occupancy of vehicles from the use of TNC services.

Recent literature advocates the use of agent-based simulation models to study ridesharing and how it impacts traffic demand ((Ciari, Balac, & Axhausen, 2016); (Ronald, Thompson, & Winter, 2015a); (Ronald, Thompson, & Winter, 2015b); (Ronald, Yang, & Thompson, 2016)). Yet, available literature lacks clear quantification of dynamic ridesharing impacts on traffic congestion. In addition, most of the studies that investigated emerging ridesharing technologies were performed in Europe or Australia. This background information emphasizes a clear and pressing need for U.S.-based studies that investigate the influences of technologies, specifically smartphone applications, on travel behavior, mode choices, and ultimately urban congestion.

1.1 Objective

The objective of this research is to develop an understanding of the influence of recently available technologies and Transportation Network Services on transportation users' choices and behaviors. The study documents users' attitudes, choices, and behavioral trends and investigates the feasibility of using simulation modeling to quantify the influence of traveler information use and Transportation Network Services availability on travel behavior and demand. The ultimate goal is to inform policy on the links between technology and driving choices in the southeastern region, where the auto-oriented built environment likely influences these links.

1.2 Scope

This study investigates the influence of TNCs such as Uber and Lyft, on travelers' behavior and traffic demand in medium size cities in the southeast and uses Alabama and North Carolina as case studies to understand transportation users' adaption to TNC services and related impacts. TNC services are provided by ride hailing platforms (such as Lyft and Uber), and arrange one-time rides on an on-demand basis. Over the last decade, the popularity of TNCs as a mode of travel has been increasing at a steady pace, even in medium size cities. Thus, it is important to identify the determinants that influence transportation users to adopt TNCs as a preferred mode choice and the impacts of such preferences on their travel patterns and transportation network operation.

In addressing the aforementioned study objectives, this study performed five tasks, including a comprehensive literature review and synthesis; a survey of millennials travel behavior in North Carolina and users' adoption to TNC services; a travel diary questionnaire survey of TNC-aware travelers in Birmingham, Alabama; a case study to evaluate the feasibility of building an agent-based simulation model of the Birmingham Metro Area; and final reporting.

1.3 Study Structure

This report begins with a comprehensive literature review and synthesis in Chapter 2 focusing on shared mobility options and associated technologies that enables them, and their impacts on travel behavior, travel demand, and congestion.

Chapter 3 documents the methodology and findings from an online survey of millennials in North Carolina that investigated the impact of ride-hailing services on travel behavior and choices.

Chapter 4 presents a study undertaken in Alabama to document factors that influence transportation users' decisions to choose TNCs for completing typical day trips. The travel diary questionnaire survey developed as part of this effort provided valuable

insights on awareness and use of technology for mode choice of 450+ transportation system users in Birmingham, AL. Moreover, data collected through this survey were used as inputs to the population synthesis subtask of the agent-based model development for the Birmingham region discussed in Chapter 5.

Chapter 5 discusses the steps taken for the development of a pilot model for the Birmingham region in MATSim that accounts for ridesharing. Using Birmingham as a testbed, the purpose of this effort was to a) test the feasibility of modeling non-traditional modes of transportation (like ridesharing) using an agent-based simulation platform and b) highlight associated model requirements and challenges.

2.0 LITERATURE REVIEW

Over the last two decades, urban centers have seen major changes in response to population growth, changing demographics, and the associated economic activities. These urban centers have traditionally served a variety of users – pedestrians, bicyclists, and transit users – but planning for all modes of travel, instead of emphasizing the automobile, has only recently been incorporated into traditional planning practices. Accordingly, traffic congestion and other problems related to auto-dependency are evident in urban centers and undermine the efficiency of the nation's transportation system. The undesirable impacts of traffic, environment, and fuel consumption are well documented. According to the 2015 Urban Mobility Scorecard (Schrank, Eisele, Lomax & Bak 2015), the financial cost of congestion nationwide is $160 billion annually or $960 per commuter. A comprehensive analysis of traffic conditions in 471 urban areas across the United States shows that travel delays, due to traffic congestion, caused drivers to waste more than 3.1 billion gallons of fuel and kept travelers stuck in their cars for nearly 7 billion extra hours – 42 hours per rush-hour commuter (Schrank et al. 2015). The report also predicts that urban roadway congestion will continue to get worse unless more assertive approaches manifest on the project, program, and policy fronts (Schrank et al. 2015).

These figures highlight the importance of considering strategies and transportation improvements that reduce the urban congestion problem. One such strategy is ridesharing. The traditional ridesharing strategy aims at sharing of a vehicle for the purpose of reducing vehicle trips. Individuals participating in a traditional rideshare arrangement, share a ride with a common origin and/or destination and typically share the travel costs as well. The idea is to better utilize the vehicle occupancy potential in order to reduce the number of vehicles on the transportation network and thus the undesirable consequences of traffic congestion, pollution, and fuel consumption. In recent years, a service commonly referred to as dynamic ridesharing (or real-time ridesharing) has emerged. Such a service is provided by ride-hailing platforms (such as Lyft and Uber) and arranges one-time rides on an on-demand basis. The dynamic ridesharing is characterized by great flexibility and less interdependence compared to traditional ridesharing.

In this review, we focus on car-sharing and ride-sharing services, the associated technology that enables them, and their impacts on various characteristics of the transportation industry, such as impacts on travel behavior, travel demand, as well as recurring and non-recurring congestion. With an emphasis on mobile software applications, as well as wearable technologies, this document provides a comprehensive review of the literature, as well as new and emerging technologies for car-sharing and ridesharing applications. Additionally, this review identifies available simulation platforms capable of simulating demand responsive transportation (DRT), car and ridesharing modes, dynamic ridesharing (DRS), and peer-to-peer ridesharing. Furthermore, applications of identified platforms are identified, including

successful applications, barriers to and challenges for simulating DRT, lessons learned, costs, and benefits of the proliferation of car and ridesharing platforms.

Car-sharing and ride-sharing services have recently entered the spotlight of the transportation field, specifically as people become more economically, socially, and environmentally conscious. In 2016, the World Economic Forum estimated that the technological advances of the transportation industry up to the year 2025 will have a societal value of $3.1 trillion (Weinelt et al. 2016). As the technology and services are still in the early adopter phase, it is important to assess the attraction factor and potential paradigm shift that could occur in the transportation field, due to their inception. Moreover, it is necessary to investigate travel behavior changes and the determinants that initiate such changes, specifically through technological advances under the notion of shared mobility.

2.1 Shared Mobility and Car/Ridesharing Applications

Shared mobility is defined as the shared use of a motor vehicle, bicycle or other forms of low-speed transportation and is just one of many characteristics of the emerging sharing economy (Shaheen, Cohen & Zohdy, 2016; Shaheen, Chan, Bansal & Cohen, 2015). Shared mobility is typically characterized by enabling users to have short-term access to modes of transportation as needed, as opposed to actual ownership (Shaheen, Cohen & Zohdy, 2016; Shaheen, Chan, Bansal & Cohen, 2015). The umbrella term, shared mobility, includes various forms of car-sharing, bike sharing, ridesharing and on-demand transport services. However, for the purposes of this review, shared mobility will concentrate on the services using automobiles as the transportation vehicle.

Focusing specifically on ridesharing, Chan and Shaheen (2012) studied its past, present, and future in North America. They presented a comprehensive review of the evolution of ridesharing since the 1990s focusing on strategies that ride-matching systems employ to create what they defined as "critical mass" (Chan & Shaheen, 2012). Chan and Shaheen (2012) categorized those strategies as:

1. Regional and large employer partnerships,
2. Financial incentives,
3. Social networking to younger population, and
4. Real-time ride-matching through smartphone applications.

Furthermore, Chan and Shaheen (2012) developed a comprehensive ridesharing classification scheme that organizes ridesharing platforms into three categories, namely: acquaintance-based, organization-based, and ad-hoc (see Figure 1). Their classification scheme accounts for all available platforms despite its development in 2012 (Chan & Shaheen, 2012). Through such a classification scheme it is clear that the rapid technological developments in the 21st century is creating new opportunities for shared-

use economy applications around the globe. According to Chan and Shaheen (2012), current examples of existing technology-based car-sharing and ridesharing solutions in the United States include:

1. Ride-hailing (or dynamic car-sharing) platforms like Uber, and Lyft;
2. Organization-based car-sharing platforms like Zipcar, Enterprise CarShare, Hertz On Demand, Car2Go, and DriveNow;
3. Ridesharing dedicated social networks like Carticipate;
4. Secure ridesharing for companies and universities like Zimride; Peer-to-peer car-sharing platforms like Getaround and Turo; and
5. Solutions for automated car passenger counting for HOV toll discounts like Carma (formerly known as Avego).

Although a proliferation of technology-based applications is occurring, ridesharing has increased slightly in recent years to about 8-11% in the Canada and the United States, respectively. The authors emphasized, in their conclusion, the key importance of marketing and public education to raise awareness about ridesharing and car-sharing services, and their potential to reduce traffic congestion and other negative impacts of automobile dependence (Chan & Shaheen, 2012).

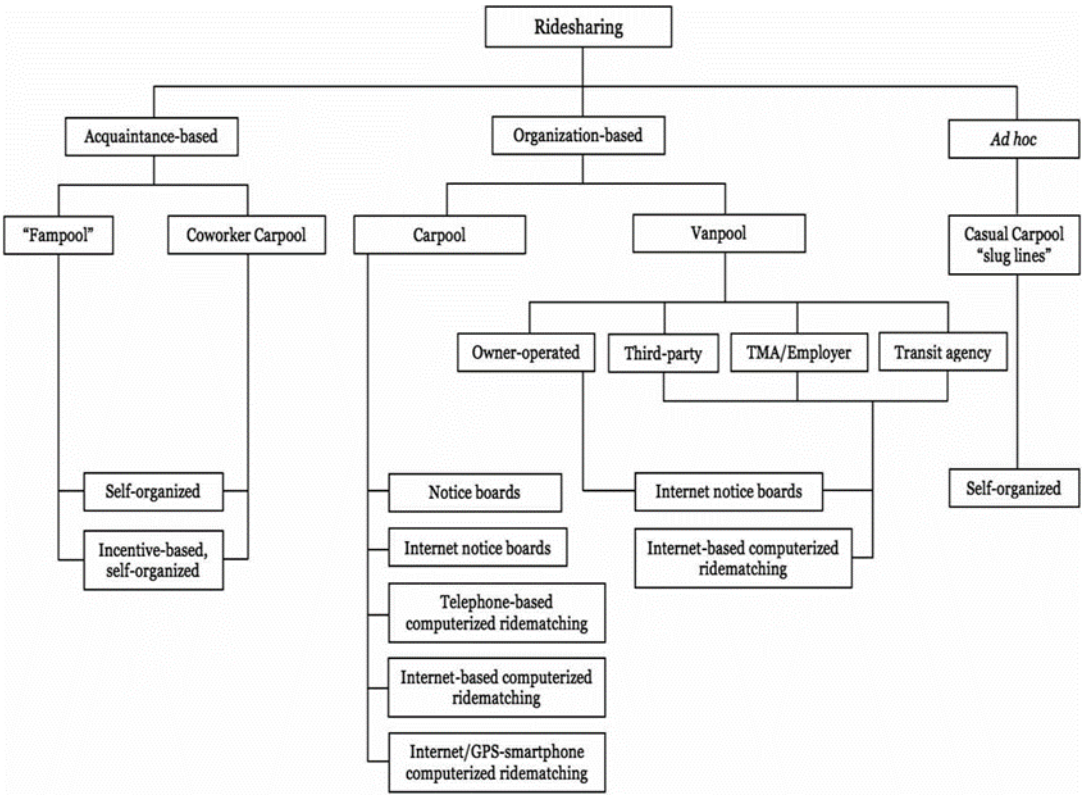


FIGURE 2-1. RIDESHARING CLASSIFICATION SCHEME AS DEVELOPED BY CHAN AND SHAHEEN (2012).

In a complementary study, Siuhi and Mwakalonge (2016) presented opportunities and challenges of smart mobile applications in transportation. Their summary of available mobile applications revealed applications for route planning, car/ridesharing, traffic safety, parking information, transportation data collection, fuel consumption and emissions, and travel information applications (Siuhi & Mwakalonge, 2016). Their study concluded that opportunities for smartphone-based applications could provide important benefits both individually and collectively in reducing travel times and cost, congestion, and vehicle and emissions (Siuhi & Mwakalonge, 2016). However, Siuhi and Mwakalonge (2016) raised a concern about the associated additional cognitive load and relevant risks arising from unnecessary distractions.

2.2 Car/Ridesharing Applications and Adoption Rates

While a proliferation of opportunities has arisen, adoption of these services has occurred at a much slower rate. According to Shaheen, Cohen & Zohdy (2016), the first car-sharing and bike-sharing programs in North America launched in 1994. However, for the first fifteen years and until Uber's 2009 market entry, these services were relatively limited to niche markets, such as college campuses and high-density urban areas (Clewlow & Laberteaux, 2016). The advent of Uber spurred a proliferation of ride-hailing

and on-demand transport companies, as well as the ensuing technological advances that accompany and are the hallmark of the vast majority of these services.

Shaheen, Cohen, and Zohdy (2016) found that by July of 2015, twenty-two active car-sharing programs and over 600 ride-matching services existed in the United States alone. However, the adoption of these services differs greatly. The rate of adoption for ride-hailing services is significantly higher than its earlier introduced counterpart, car-sharing services (Clewlow, 2017). For example, over the first fifteen years of their existence, car-sharing services garnered only five million global users, whereas 250 million global users were accrued over just the first five years of ride-sharing services (Clewlow, 2017). Moreover, half of car-sharing service (e.g., Zipcar, Car2Go, DriveNow) users have dropped their memberships, with 23% citing ride-sharing services as the main reason for this occurrence (Clewlow, 2017).

Clewlow and Laberteaux (2016) further found that ride-hailing services are expanding into new markets previously untouched by the car-sharing industry. However, some are concerned about a potential national monopoly by Uber with its command of market shares for ride-hailing services. Uber has largely dominated the market since its 2009 inception, accounting for over 80% of its shares, though recently this proportion has dropped below 75% (Cortright, 2017). While Uber's total market share has only slightly decreased across the nation, several urban markets show Lyft as a worthy competitor, accruing up to 45% of shares in certain markets (Cortright, 2017). The surge in Lyft's shares in individual urban markets has reduced some of the conversation surrounding Uber's monopoly and provides customers with more competitive pricing (Cortright, 2017). Although Lyft has been able to change the conversation about market share into a discussion of a duopoly, as opposed to a monopoly, local startups have had difficulty competing with the branding power and financial backing of the two national scale ride-hailing services (Cortright, 2017).

In addition to branding and finances, consumer loyalty is also a significant factor in the ride-hailing service industry, with Cortright (2017) finding that 90% of ride-hailing customers exclusively using either Uber (79%) or Lyft (19%) and only 10% using both. This makes it considerably harder for local startups to compete in the market. Nevertheless, the increase in smartphone and wearable technology, as well as prevailing social, economic and environmental conditions have acted as catalysts to these car-sharing and ride-sharing programs (Shaheen, Cohen & Zohdy, 2016). In fact, because of these conditions and advances in technology, other forms of the transportation sharing economy have emerged, such as peer-to-peer and ride-matching services. While technology has enabled the emergence of such services, these services largely rely on and require the use of technology.

The use of ridesharing has been found to be contingent upon the choice of modes, their cost and the length of the trip. Hampshire, Simek, Fabusuyi, Di and Chen (2017)

completed a stated and revealed preference study of 1,840 former users of Uber and/or Lyft in Austin, Texas, when the national ridesharing companies temporarily suspended service in 2016. They found that in response to the disruption, trip frequency declined significantly, and 45% switched to the use of personal vehicle, 41% to another (local) TNC, and 3% shifted to public transit. Among those who switched to a personal vehicle, 8.9% reported purchasing a vehicle in response to the service disruption. Individuals who lived in city center were more likely to switch to another TNC than to purchase a vehicle (Hampshire et al. 2017). Chavis and Gayah (2017) developed a mode choice model that can be used to describe how transit users select between emerging competitive transit options. The results revealed that monetary costs, expected in-vehicle waiting time, expected waiting time and walking time were found to be statistically significant predictors of the type of flexible transit option selected (Chavis & Gayah, 2017). Martinez, Correia, Moura and Lopes (2017) developed a detailed agent-based model to simulate a one-way carsharing systems in Lisbon. The simulation incorporates a stochastic demand model discretized in time and space and a detailed environment characterization with realistic travel times. Their results suggest that carsharing performs worse than private cars in terms of time and costs, it outperforms taxis in terms of cost, and outperforms buses, metro, and walking in terms of travel time. They concluded that "the competitiveness of carsharing is highly determined by trip length, becoming more competitive than other modes, (travel time wise) as trips become longer" (p. 148). Cuevas, Estrada and Salanova (2016) compared one-way carsharing to taxi stand service in Barcelona and found that "although the performance of both systems is very similar, the taxi service is up to three times more expensive" (p. 155).

2.3 Travel Behavior

Through the inherent interactions between technology and car/ridesharing applications, knowledge and utility of these services is found to vary greatly in accordance to a vast array of socio-demographic variables, as with most new technology (Smith, 2016). Before discussing the various socio-demographic variables that influence the use of these services, it is important to note that in a national Pew Research Center survey of 4,787 American adults in 2015, Smith (2016) found only 15% of Americans had used ride-hailing apps, whereas one third had yet to even hear of them. Nonetheless, socio-demographics are pivotal in explaining the adoption and use of these services. While Smith (2016) did not find race or gender as influential factors in the use of these apps; age, education, income level and type of locale (i.e. urban, suburban, or rural) were all found to be strongly explanatory. According to Smith (2016), 29% of college graduates had used the services compared to only 6% of people with educational attainment levels of a high school graduate or lower. Similarly, 26% of people with an annual income in excess of $75,000 had used ride-hailing apps, while only 10% of people with an income

less than $30,000 had. Moreover, 28% of 18-29-year-olds and 19% of 30-49-year-olds have used these services, whereas only 4% of 65+-year-olds have (Smith, 2016). Clewlow and Mishra (2017) found similar results in a study using "comprehensive travel and residential surveys deployed in seven major U.S. cities, in two phases from 2014 to 2016, with a targeted, representative sample of their urban and suburban populations" (p. 1). They found that 21% of adults in major cities personally use ride-hailing services; and an additional 9% use ride-hailing with a friend even if they have not personally installed the app (Clewlow & Mishra, 2016).

It is important to note, however, that non-mutually exclusive socio-demographic factors have an overlapping influence that appear to explain significant variation in the adoption of the ride-hailing programs. For instance, while 7% of all 18-29-year-olds use these apps on a daily or weekly basis, this proportion increases to 10% for urban 18-29-year-olds (Smith, 2016). Nevertheless, some factors are more influential than others. This is best exemplified through the geographical concentration of these services, with most being more available in urban and suburban areas, leaving rural parts largely unserved. "One-in-five urban Americans (21%) have used ride-hailing services, as have 15% of those living in suburban areas. By contrast, just 3% of rural residents have used these services and more than half (54%) have never heard of ride-hailing apps (Smith, 2016). While the explanatory aspect of Smith's (2016) article is weak, due to the lack of quantitative analysis, several findings from the survey data are helpful in understanding the adoption, and usage and demand for these services.

The demand for these services is further explored by other researchers. Agatz et al. (2011) used 2008 travel demand data for metropolitan Atlanta to develop a optimization-based approach that attempted to minimize the total system-wide vehicle miles traveled (VMT) incurred by system users. They concluded that "even with relatively low participation rates, it appears that sustainable populations of dynamic ride-sharing participants may be possible even in relatively sprawling urban areas with many employment centers" (Agatz, Erera, Savelsbergh & Wang, 2011: p. 532). In contrast, Martinez et al. (2017) concluded that uncertainty remains on the economic viability of the carsharing given the complex relation between supply and demand, and how this may influence the level of service provided. Wielinski, Trepanier and Morency (2015) explore the use of a free-floating car sharing compared traditional, station-based service and found that more women were members of the free-floating carsharing and the trip distances and durations were much shorter, and trip ends were concentrated near the central business district in midday. "When asked what mode users would have used in the absence of the free-floating service, people mentioned public transit, taxis and walking; the popularity of these alternatives varied, probably in relation to seasonal changes (Wielinski et al. 2014: p. 28).

In addition to the geographical aspect noted above, the adoption rates among different generations also foretell changes in these trends. Clewlow and Mishra (2017) found that, among a representative sample of population in seven major urban areas, only 4% of those aged 65 and older have used ride-hailing services, compared with 36% of those 18 to 29. Millennials are often said to have differing travel behaviors and lifestyles at the same life stage as proceeding generations (Circella et al. 2017). Examples of these differences range from further urbanization to delaying driving licensure, among others (Cirella et al., 2017). However, the impacts on the national transportation sector from these changes are still largely unknown. In response to this gap in knowledge, Cirella et al. (2017) examine the differences in travel mode choice between Millennials and Generation Xers in California. From a data set including 2,155 young adults and members from the proceeding generation, Cirella et al. (2017) found that compared to Gen X'ers, Millennials are two times more likely to ride a bike, three times more likely to use Uber or Lyft and five times more likely to use a work or school shuttle. Although Millennials are more likely to adopt alternative modes of transport, it should be noted that no more than 4% of those surveyed adopted any mode of choice other than a car (Cirella et al. 2017). Nonetheless, Millennials are not only more likely to adopt alternative modes of transport, they are also more likely to engage in multimodal and intermodal trip behaviors (Cirella et al. 2017). Although Feigon and Murphy (2016) do not analyze users by age, they define the concept of supersharers, who are travelers who use several shared modes – bikesharing, carsharing and ridesourcing. They conclude that "greater use of shared modes is associated with greater likelihood to use transit frequently, own fewer cars, and have reduced transportation spending" (Feigon & Murphy, 2016: p. 1). "Approximately 57% of supersharers said public bus or train was the single shared mode they use most often, followed by bikesharing, ridesourcing, and carsharing" (p. 7) and they own half as many cars as people who use transit alone (Feigon & Murphy, 2016).

One aspect that is overlooked in Smith's (2016) article and only slightly touched upon by Cirella et al. (2017) is the use of ride-hailing programs as ride-sharing services. The real social, environmental, and even economic value stems from the potential of these programs to be used as ride-sharing applications, thus having multi-consumer occupancy, as opposed to single-consumer occupancy trips. Ride-hailing and car-sharing services are surrounded by ongoing debates concerning their impacts on congestion, VMT and other transportation-related performance measures that impact the social, economic, and environmental nature of a geographical area. In this respect, it is more important to assess the type of trips made, as well as the occupancy levels of such ride-hailing trips to determine their impact on the aforementioned variables.

2.4 Congestion and Ride-Hailing Services

In 2016, the U.S. Energy Information Administration reported that transportation had become the economy's largest polluting sector (Tomer, 2017). This is unsurprising and consistent with the results of the 2015 Urban Mobility Scorecard, as previously mentioned (Schrank et al. 2015). Transportation is also the second highest average individual expense, yet 76.3% of all commuters drive alone to work (Tomer, 2017). This behavior results in the average rate of occupancy for motor vehicles in the U.S of 1.6 persons per vehicle mile and even lower for commute trips (Huang, Jin, Bastani & Wang, 2014). Yet, according to Tomer's (2017) analysis of the American Community Survey (ACS) (American Community Survey) estimates for 2007-2016, there was less than a 1.5% change in the share of total commuters for any mode of transportation. As such, there has yet to be a substantial paradigm shift towards ride-sharing as a dominant mode of transportation, even in urban environments.

Nevertheless, Sivak (2015) found that various characteristics related to the transportation sector, such as vehicles-miles driven and gallons consumed, all reached their peak by 2008. The data used by Sivak (2015) to assess these changes extends from 1984 to 2012. While it could be suggested that some of these "peaks" are due to the recession, the author asserts that these changes "reflect fundamental, noneconomic changes in society, such as increased telecommuting, increased use of public transportation, increased urbanization of the population, and changes in the age composition of drivers" (Sivak, 2015, p.1). Moreover, with the technology being in the early adopter's phase and still undergoing mass technological advancements, an ongoing debate has emerged about whether these ride-hailing services will actually reduce congestion and other negative transportation characteristics, such as VMT.

Using a two-phase survey of seven metro areas in the United States, Clewlow and Mishra (2017) found that 49%-61% of ride-hailing trips are trips that would have never been made or would have been made by walking, biking or using public transit. As a result of this mode substitution, Clewlow and Mishra (2017) conclude that ride-hailing services are increasing VMT in major cities across the U.S. Therefore, congestion and emissions are likely to grow through the adoption of ride-hailing services (Clewlow & Mishra, 2017). In response to these findings, Clewlow and Mishra (2017) suggest that policymakers need to address the issue of added VMT through congestion pricing and by means of prioritizing high-occupancy vehicles (HOVs). Thus, the authors advocate for the use of ride-sharing systems, as opposed to the current norm of single-consumer occupancy ride-hailing services.

These findings are supported by Bliss (2017) who, from a study of 4,000 users in seven U.S. metropolitan areas, claims that ride-hailing services are not reducing vehicle ownership or VMT. Instead, they are helping car-sharing members transition into ride-

sharing services (Bliss, 2017). Bliss (2017) further found that the top two reasons cited for using ride-sharing services were parking and drinking. Bliss's results are reinforce by findings in other studies. Clewlow and Mishra (2017) found parking represented the top reason that urban ride-hailing users use ride-hailing instead of driving themselves. Avoiding driving while drinking was another top reason users substitute ride-hailing for driving themselves (Clewlow and Mishra, 2017). Balac, Ciari and Axhausen (2017) found that "free floating vehicles are able to use parking spaces more efficiently than private vehicles" (p.207) and can avoid spatial parking pressure peaks. Greenwood and Wattal (2015) looked at total alcohol-related deaths associated with the introduction of two Uber services in California – Uber Black, a premium car service, and Uber S, the discount. They found that while the introduction of premium Uber Black service did not impact on alcohol related deaths, Uber X correlated with a 3.6 – 5.6 percent reduction per quarter in 540 townships in California (Greenwood & Wattal, 2015). Feigon and Murphy (2016) in a survey of 4,500 shared mobility users in seven cites - Austin, Boston, Chicago, Los Angeles, San Francisco, Seattle, and Washington, DC found that "ridesourcing is most commonly used for recreation and social trips, late at night, and often when alcohol is involved" (p. 13).

The Schaller (2017b) similarly found evidence of added VMT from ride-hailing services due to increases in deadheading from ride-hailing services traveling without passengers. Using data from the NYC Taxi and Limousine Commission, which is collected for both taxi cab and ride-hailing services, Schaller (2017b) examined the unoccupied time between trips for typical ride-hailing platforms. In doing so, Schaller (2017b) found that 45% of operation time is spent unoccupied, thus, adding to VMT, congestion and emissions by increasing the number of vehicles on the transportation network. These results are supported by Santi et al. (2014) who conducted an analysis of taxi sharing in New York City using the notion of a shareability network, which allowed them to model the collective benefits of sharing as a function of passenger inconvenience. Using a dataset of millions of taxi rides with increasing but still relatively low passenger inconvenience, they concluded that cumulative trip length could be reduced by 40% or more with the associated reductions in costs (due to split fares) and emissions (Santi et al. 2014). Martin and Shaheen (2011) analyzed the greenhouse gas emissions from individuals participating in carsharing organizations and concluded that the relationships is complex. The majority of households joining carsharing are slightly increasing their travel and associated emissions by gaining access to automobiles (Martin & Shaheen, 2011). At the same time, other carsharing families are driving less and reducing emissions by shedding vehicles (Martin & Shaheen, 2011). Xu, Ordonez and Dessouky (2015) combined a ridesharing market model with a classic elastic demand Wardrop traffic equilibrium model to show that "(1) ridesharing base price influences the congestion level; (ii) within a certain price range, an increasing in price may reduce the traffic congestion; and (iii) the utilization of ridesharing increases as congestion increases" (p. 793).

In a study using the San Francisco Bay Area activity-based model, Rodier, Alemi, and Smith (2016) found that the relationship is complex. They simulated business-as-usual, transit-oriented development and auto pricing scenario with and without high, medium and low ridesharing participation. They found that relatively large VMT reductions are possible from moderate and high participation levels, but at low participation levels, VMT reductions are negligible. Moderate dynamic ridesharing alone compares favorably, with a 9% reduction in VMT, to transit-oriented development and auto pricing scenarios. The analysis also suggests a potentially promising policy combination: a moderately used regional dynamic ridesharing system with a 10- to 30-cent increase in per mile cost of auto travel, which together may reduce VMT on the order of 11% to 19% (Rodier, Alemi & Smith, 2016: p. 120).

In contrast to the findings of increased congestion, Li, Hong and Zhang (2016a) found that the entry of Uber in U.S metropolitan areas resulted in reduced congestion. Their analysis spanned an 11-year time interval and included 87 urban environments. Where Bliss (2017) and Clewlow and Mishra (2017) rely solely on survey data to suggest that congestion is actually increasing, Li, Hong, and Zhang (2016a) employed regression analysis with log transformations and a difference-in-difference approach, using 957 observations, to investigate whether the traffic congestion before and after Uber entry is different across different urban area. Thus, Li, Hong and Zhang (2016a) were able to infer Uber's positive impact in alleviating congestion across U.S metropolitan areas as a ride-hailing service. These findings are supported by Alexander and Gonzàlez (2015) in a study exploring different adoption rates to estimate ride-sharing demand through smart-phone based technology, which similarly concluded that moderate to high adoption rates are likely to reduce congestion and travel times. Although the findings suggest Uber does reduce VMT and congestion, Li, Hong and Zhang (2016a) similarly advocate for the ride-sharing aspect of services like Uber for realizing the full impacts of such programs on the aforementioned transportation characteristics.

While Li, Hong, and Zhang (2016a) assessed the impacts of Uber's entry in various urban markets to assess changes in congestion, the study did not take into account mode substitution, which Clewlow and Mishra (2017) identified as a major component to measuring reduced VMT and congestion. Between 1995 and 2015 public transit ridership in the United States increased by nearly 3 billion trips (Lewyn, 2018). However, public transit ridership has declined over the last couple of years (Lewyn, 2018). According to Bliss (2017), this has a direct relationship with the proliferation of ride-sharing services, as the author suggests that a 6% and 3% net reduction in the use of buses and light rail, respectively, are due to shared mobility's attraction. This finding was further supported by Rayle et al. (2014), who found, from a survey in San Francisco, that 33% of riders would have otherwise used the bus or rail if ride-hailing services were unavailable. Sperling and Brown (2018) concludes that a shift in mass transit should be expected, due to ride-hailing services, as they offer many of the same advantages as

mass transit. Nevertheless, ride-hailing services still have the potential to reduce congestion and VMT by facilitating multi-passenger pooling (Sperling & Brown, 2018).

2.5 Ride-sharing and Technology

The pooling aspect of ride-sharing services is crucially important in their ability to produce the social and environmental value envisioned with their adoption. However, carpooling in the United States has been on downward trend from 20% in the 1980's to under 10% currently (Sperling & Brown, 2018). While carpooling has lessened in the United States over the last few decades, the most recent available data suggest that the proportion of total commuters carpooling is double that of mass transit (Sperling & Brown, 2018). As such, it is an attractive market for companies, like Uber and Lyft, that simultaneously seek to expand their market through their dynamic ride-sharing services (DRS), UberPool and Lyft Line, respectively, which match drivers and riders in real time.

Although ride-sharing is suggested to reduce congestion by the likes of Uber and Lyft, Schaller (2017a) cites mode substitution as a primary reason as to why DRS, even with its ride-sharing component, will lead to increased VMT in New York. Adoption is also another factor, with exclusive-rides accounting for the majority of ride-hailing trips (Schaller, 2017a). Henao (2017) further identifies barriers to DRS through passenger willingness and added inconvenience for drivers, when compared with exclusive-ride trips. While ride-sharing and more specifically, DRS, have yet to gain widespread adoption, a significant optimism exists towards their environmental and social benefits, with Quarles (2017) stating “DRS is one of the few ways the world’s transportation future becomes environmentally sustainable” (Quarles 2017, p. 11).

More specifically, ride-sharing services are said to have entered the fifth phase in their development, which is characterized by technology-enabled ride-matching (Chan & Shaheen, 2012). While traditional ride-sharing services were fixated around casual car-pooling, HOV lanes, and park-and-ride efforts, the fifth phase is most notable for its utilization of the internet, mobile phones and social networking (Chan & Shaheen, 2012). Through this added dimension, Chan and Shaheen (2012) suggest that the greatest contribution may be helping to overcome the critical mass barrier, which is the ability to provide “enough users to consistently create a successful instant ridesharing match” (Chan & Shaheen, 2012, p. 13). Furthermore, they conclude that interoperability among ride-sharing databases and open source data sharing across ride-matching services could be influential towards this goal. Chan and Shaheen (2012) took their suggestions one step further to include multimodal integration, which they suggest would be “the seamless connection of ridesharing with other transportation modes, such as public transit and carsharing” (Chan & Shaheen, 2012, p. 18).

Such a service, called UbiGo was launched in Sweden. This innovative service combines transit, car-sharing, rental car service, taxi and bicycle systems on one platform. The

service is used through a smartphone application with a comprehensive invoicing system, 24/7 support, and bonus credits for sustainable choices. Sochor et al. (2015) conducted a field operational test using mixed data collection methods to investigate barriers to and opportunities for implementing mobility as a service. They relied on responses from UbiGo customers to questionnaires, interviews, focus groups, travel diaries, and workshops. Their analysis of results concluded that 93% of participants were satisfied with the service with 97% wanting to continue using UbiGo. However, no data were collected regarding the impacts on VMT, traffic congestion, or other impacts of such a service on the existing transportation network. Uber also seems to want to extend its operations in such a direction, with the CEO stating he wants to run a bus company for a city (Brasuell, 2018). Making his ambitions clearer, Khosrowshahi was quoted saying: "I want you to be able to take an Uber and get into the subway — if the trains are running on time, you've got real-time data — get in the subway, get out and have an Uber waiting for you for right now. Or know that there's a bike right there for you that gets you where you're going in the fastest manner" (Brasuell, 2018).

Zhu et al. (2016) propose another form of ride-sharing application that they identify as public vehicles (PV), which can act as a public on-demand ride-sharing option. In their article, Zhu et al. (2016) use Shanghai as a cases study to propose a PV system that could replace buses, private cars and taxis within urban environments. Through the use of simulations of the PV platform, it is estimated that the number of vehicles operating in an urban environment that employs PV can be reduced by 90%. Furthermore, simulation results suggest that the PV system would be even more efficient than conventional ride-sharing programs, such as Uber and Lyft, reducing vehicles operating in the urban areas by an additional 57% (Zhu et al. 2016). While the estimated reductions are quite large, it is important to recognize that the Shanghai region is characterized by extremely high density development compared to most Western cities. Furthermore, the PV system modeled by Zhu et al. is highly centralized and, as such, would be significantly more efficient than a decentralized system of taxis and buses.

In addition to reducing vehicles in the urban system, which in turn reduces congestion, the simulation results also suggest that VMT or total travel distance would be reduced by 34% compared to urban environments without a ride-sharing platform and by 14% compared to those that employ a traditional version of these services (Zhu et al. 2016). The PV system proposed by Zhu et al. (2016) uses a cloud-based algorithm that schedules passengers by means of the origin and destination locations through smartphone apps. The scheduling and ride-matching are conducted through the cloud, which matches riders with drivers. While these would be public vehicles, Zhu et al. (2016) suggest that the system could be run by a government entity or a private company, which brings to question such a system could be more efficient than already established private companies, such as, Uber and Lyft.

2.6 Ride-sharing, Regulations, and Safety

While a proliferation of ride-sharing services, such as the PV system of Zhu et al. (2016) has been proposed and suggested to tackle issues of increased congestion and VM, new issues regarding regulations and safety are coming into the spotlight. For example, Beer et al. (2017) evaluated six driver-related and three company-related types of ride-hailing regulations in 24 major American cities and concluded that ride-hailing companies are less likely to operate in cities where there are stricter background checks and where fingerprinting of drivers is required. Such issues with regulations surround the taxi industry and excessive concern about regulation can be seen as counterproductive. This is especially so with the findings from a survey conducted by Rayle et al. (2014), who concluded through a paired analysis that occupancy levels for taxis were just at 1.1 passengers per vehicle, whereas ride-sharing services average 1.8 (Rayle et al. 2014; Li, Hong & Zhang, 2016b). This difference is extremely impactful, as adding nearly one additional person to every commute trip has the potential of achieving a savings of 7.7% on fuel consumption and a reduction of 12.5% on VMT (Li, Hong & Zhang, 2016b).

Nevertheless, discussion of regulations leads to the question of whether ride-sharing, which is largely regulated by private companies, is safe. Examining this question, Feeney (2015) actually found that Uber and Lyft users are demonstrating their dissatisfaction with taxi services, which he suggests is due to the shoddy and unreliable services with inflated prices. Moreover, the author concludes that the cash-free transactions and self-identified customers of ride-sharing services have reduced the risk of violent crime (Feeney, 2015). Feeney (2015) further suggests that in some respects the screening processes by private companies, such as Uber and Lyft, are superior to screenings for taxi drivers.

Safety is just one concern of adopting ride-sharing as a typical mode of transportation. Other psycho-social variables related to determining the willingness of a person to use car-sharing or ride-sharing services also factor into adoption rates. Some of these other considerations include the nature and purpose of the trip, the distance and time of day (Chowdhury and Ceder, 2016; Malodia and Singla, 2016; Merat, Madigan & Nordhoff, 2016), as well as personality of riders (Roy, 2016; Merat, Madigan & Nordhoff, 2016). As for willingness to use these services, it was found that while safety (both physical and identity wise) are influential factors, the two main factors affecting a person's willingness to carpool are time and cost (Malodi & Singla, 2016).

2.7 Shared Autonomous Vehicles (SAVs) and Ride-Sharing

Fagnant et al. (2015) went a step further than current ride-sharing services and predicted that once autonomous vehicles (AVs) are commercially available, a new transportation mode for personal travel will be set to arrive. They named that anticipated mode as "shared autonomous vehicles." Fagnant et al. (2015) portrayed

SAVs as a merger of car-sharing, ridesharing, and taxi services. The authors developed a simulation model of SAVs with a market penetration rate of 1.3% of regional trips in the regional core of Austin, Texas. Their simulation results suggested that SAVs could replace conventional vehicles at a rate of 1:9.3 assuming all other modes are available. In their conclusion, they stated that they anticipate positive impacts once the technology stabilizes at a higher market penetration rate. Other simulation studies have also supported this optimistic outlook, with Alonso-Mora et al. (2017) finding, through simulation, that it is possible for 98% of taxi rides in New York, currently served by over 13,000 taxis, to be served by just 3,000 vehicles in an SAV DRS model. Moreover, these trips would only incur a mean waiting time of 2.8 minutes and a mean trip delay of 3.5 minutes, thus minimizing inconveniences for drivers and riders, alike (Alonso-Mora, Samaranayake, Wallar, Frazzoli, & Rus, 2017).

Nevertheless, SAVs bring further regulatory, safety and resulting adoption rates into question. Safety concerns surrounding technology and automotive vehicles should come as no surprise, as the World Health Organization estimates that 1.35 million people die each year worldwide as a result of car crashes, with millions more being injured (WHO, 2018). Moreover, humans are very good at and trust themselves in recognizing other drivers, their intentions and their best driving actions (Althoff, 2010). In addition to these traits, Althoff (2010) further suggests that humans have the ability to focus on relevant information, handle unexpected situations and learn from them. As such, Althoff (2010) states that

Aprerequisite of autonomous driving is the equipment of vehicles with sensors for the detection of their environment. More importantly, relevant information such as the position and velocity of other traffic participants has to be correctly extracted from the raw data streams of the sensors. This has to work properly in different weather conditions and even when unknown or unexpected objects are present (Althoff, 2010, p. 122).

In addition to these prerequisites, Althoff (2010) also recommends that providers of SAVs need to gauge intentions of other travelers so that the optimal behavior for AVs can be deduced (Althoff, 2010). In response to the issues with safety surrounding SAVs and AVs in general, Althoff (2010) reminds us that one of the main objectives of AVs is to exclude human error and thus envision accident-free driving.

Fagnant and Kockelman (2015) similarly communicate the conceivable decrease in crashes through the inception of AVs and also identify barriers to their implementation. The authors first identify sensor recognition as a primary concern that has been shared by other academics in response to the implementation of AVs on to the roadways (Fagnant & Kockelman, 2015). Adding to this, Fagnant and Kockelman (2015) identify two other major concerns surrounding the adoption and implementation of this new technology – evasive decisions made by the AVs in response to various objects in their

path (i.e. cardboard box vs concrete block) and the liability associated with crashes involving AVs. The latter is of crucial importance as it is suggested to be a potentially "substantial impediment" (Fagnant & Kockelman, 2015, p. 4).

From the abovementioned concerns, rapid adoption of such future technology remains in question. In a pursuit to address this gap in the literature, Schoettle and Sivak (2016) examine preference for various levels of automation. Using the SurveyMonkey platform, the authors recruited 618 respondents to their survey, all licensed and aged 18 years or older (Schoettle & Sivak, 2016). From this survey, Schoettle and Sivak (2016) found that only 15.5% of drivers preferred completely self-driving vehicles. An AAA survey (2016) found that 75% of drivers report fears of riding in self-driving vehicles, with over 66% percent of drivers being at least moderately concerned. However, other findings from Schoettle and Sivak (2016) point to a potential for automation to continue catching on, with 37.8% of drivers preferring partially self-driving cars. Yet, acceptance of full automation will require features that make riders feel comfortable. For example, having the ability to take control at will also seems to be a factor in adoption with 94.5% of drivers still wanting the potential to have steering, gas and brake pedal control (Schoettle & Sivak, 2016).

Because of the demand for such a type of feature, several states have passed laws requiring drivers of AVs tested on public roads to be licensed (Fagnant & Kockelman, 2015). Therefore, AV licensure presents another barrier to implementation. This stems from the nature of AV enacted legislation occurring at the state level rather than the national level, causing regulatory uncertainty for manufacturers and consumers, alike (Fagnant & Kockelman, 2015). Other barriers to implementation, are vehicle costs, issues of privacy and security, as well as concerns about litigation and liability (Fagnant & Kockelman, 2015).

2.8 Simulation Platforms

Through such aforementioned barriers, demand responsive transportation systems have yet to benefit from widespread adoption and AV technology has yet to establish itself as a viable mode of transport. Consequently, little is known about the value each can bring to an urban environment, especially in terms of traffic and congestion-related impacts. As such, most studies that focus on this topic rely on mathematical simulations to extrapolate the potential impacts related to these technological transitions in transportation. According to Ronald et al. (2015a) "to adequately model DRT, a combination of optimization (incorporating allocation of requests to vehicles) and simulation (incorporating the movement of people and vehicles) is required" (Ronald et al. 2015a, p. 405-406). The authors refer to these required pieces as demand, supply, and algorithms, respectively, where demand and supply are inputs into algorithms to produce an output (Ronald et al. 2015a). In this respect, recent literature advocates the

use of agent-based simulation models to study car-sharing and how it impacts traffic demand (Ciari et al. 2016; Ronald et al. 2015a; Ronald et al. 2015b; Ronald et al. 2016).

Gilbert (2008) defines agent-based simulation models as "a computational method that enables a researcher to create, analyze, and experiment with models composed of agents that interact within an environment" (Gilbert, 2008, p. 2). Agent-based modeling is said to have advantages over traditional mathematical approaches "by providing insight into the operation of a system, especially taking passenger preference into account and catering for a wider range of scenarios" (Ronald et al., 2015a, p. 409).

It is important to note that two different approaches can be taken to simulating demand responsive transportation, namely fixed route and ad hoc (Ronald et al., 2015b). Ronald et al. (2015b) investigate the performance changes between each. In their work, they model demand data from a real-world scheme currently running in rural Victoria, Australia, using a fixed scheme and an ad hoc scheme. Ronald et al. (2015b) classifies fixed schemes as those with fixed routes or times, which typifies most DRT schemes, while the removal of these constraints is characterized as ad hoc simulation. The authors further use two optimization schemes; one for operators and one for passengers (Ronald et al. 2015b). Through the two simulation schemes, the authors found that each produced different outcomes for passengers and operators, with the Flexiride (fixed route) producing better outcomes for the operator-directed optimization scheme, whereas the ad hoc favoring the passenger scheme (Ronald et al. 2015b). The following table represents the characteristics of each scheme.

Characteristic	Flexiride	Ad Hoc
Scheduling type	Demand responsive	Unscheduled
Route type	Flexible route	Flexible route
Vehicle type	Taxi	Taxi–minibus
Origin and destination relationship	Many to many	Many to many
Origin and destination service	Checkpoint	Checkpoint

FIGURE 2-2. CLASSIFICATION OF SCHEMES UNDER INVESTIGATION AS DEVELOPED BY RONALD ET AL. (2015B).

To conduct such simulations one can use various platforms that are readily available. According to a 2013 Federal Highway Administration (FHWA) report, numerous agent-based modeling platforms are already available with some of the most widely-known being Transportation Analysis and Simulation System (TRANSIMS), Multi-Agent Transport Simulation Toolkit (MATSim), Sacramento Activity-Based Travel Demand Simulation

Model (SACSIM), Simulator of Activities, Greenhouse Emissions, Networks, and Travel (SimAGENT), Open Activity-Mobility Simulator (OpenAMOS), and Integrated Land Use, Transportation, Environment (ILUTE) (Zheng et al., 2013). Zheng et al. (2013) conclude that most of these agent-based models are individual-based, with nearly all exhibiting similar architecture. The authors found that microsimulation of agent activities is exhibited across all platforms in their review, though the authors did note that there are differences in characteristics, such as in design processes, functionalities, agent activities and data structures, among others (Zheng, et al., 2013). In addition to the similar architecture present within the platforms reviewed, Zheng et al. (2013) found that, in general, current agent-based systems incorporate three components similar to those identified by Ronald et al. (2015a). These are: travels' activity decisions, travelers' route decisions and microsimulation.

Although most agent-based models are individual-based, the authors do separate agent-based models into two "methodological domains: individual-based models that study personal transportation-related activities and behavior, and system and computational methods, known as *MAS* (multi-agent system), to study a collaborative and reactive transportation system by modeling autonomous decision making by a collection of subsystem entities called *agents"* (Zheng et al., 2013, p. 59). The latter of the two methodological domains is described as a method in system modeling, where the common feature found in related studies is "that the inherent distribution allows for a natural decomposition of the complex system into multiple subsystems" (Zheng et al., 2013, p. 39). One crucial difference between individual-based models and MAS is the agent in question; humans act as agents in the former, while the latter incorporates the term agent to signify autonomous operators (Zheng et al., 2013).

MAS models have been applied to various transportation problems, such as traffic management (TRYS/TRYSA2, CTMRGS, CLAIRE, and CARTESIUS), dynamic route guidance (TRACK-R) and signal control (aDAPTS, HUTSIG and Botelho) (Zheng et al., 2013). To demonstrate the differences between the individual-based models and the MAS models, the authors concisely state "that the former is closely related to models for activity-based travel demand and land use, whereas the latter is typically scoped as a powerful technique for simulating dynamic complex systems to observe emergent behavior" (Zheng et al., 2013, p. 59). The hopes of this paradigm shift towards agent-based modeling systems is that these models can discern and flexibly predict traveler's behaviors and actions through real-time information with sudden changes in the transportation network, in addition to identifying emergent travel behaviors for newly established transportation environments (Zheng et al. 2013).

In a similar study, Saidallah et al. (2016) compared 11 urban road traffic simulators, choosing the most mature and most used. In their list, the authors include ARCHISIM, SUMO, MATSim, MITSIMlab, AIMSUN, CORSIM, Paramics, SimTraffic, TransModeler,

VISSIM, and TRANSIMS. In comparing the different simulation platforms, the authors use nine different criteria, namely type of simulation model (microscopic, mesoscopic, macroscopic), software category (open-source, commercial), system (discrete or continuous), visualization (2D, 3D or both), infrastructure (difficulty and flexibility in coding), vehicles and pedestrians (type, dimensions, priority), scope area (maximum area simulator can simulate), detectors (sensors, cameras, etc.), and geographic information systems (GIS) (importing maps categorized as yes, partially or not) (Saidallah et al. 2016). The following chart depicts the categorization scheme used by Saidallah et al. (2016).

Simulators	Model			Category		System		Visualization		Infrastructure						Vehicles and pedestrians					Scope Area			Detectors		GIS		
										Difficulty			*Flexibility*															
	I	*E*	*A*	*O*	*C*	*D*	*C*	*2D*	*3D*	*E*	*M*	*D*	*F*	*L*	*VL*	*T*	*D*	*R*	*E*	*O*	*I*	*R*	*O*	*WD*	*WL*	*Y*	*P*	*N*
AIMSUN	√	√	√		√		√	√	√			√	√					√	√	√		√	√	√	√	√		
ARCHISIM	√				√	√		√				√			√						√			√				√
CORSIM	√				√	√		√	√						√			√	√		√	√		√				√
MATSim	√			√			√	√						√							√	√		√		√		
MITSIMLab	√			√				√						√							√			√				√
Paramics	√				√	√		√	√				√					√	√	√	√	√		√	√			
SimTraffic	√				√			√	√	√				√				√	√		√	√		√				√
SUMO	√			√			√	√				√			√						√	√		√				
TRANSIMS	√	√		√		√		√							√							√	√	√				
TransModeler	√	√	√		√			√	√					√					√			√	√	√		√		
VISSIM	√				√		√	√	√	√			√					√	√	√	√	√		√	√	√		

FIGURE 2-3. COMPARATIVE TABLE OF TRAFFIC SIMULATORS AS DEVELOPED BY SAIDALLAH ET AL. (2016).

Saidallah et al. (2016) conclude that only four simulation platforms simulate traffic continuously (VISSIM, SUMO, MATSim and AIMSUN), whereas the other reviewed simulations use discrete systems. Ease of coding also differed with VISSIM and SimTraffic providing easier coding processes, whereas others, such as AIMSUN, ARCHISIM and SUMO, requiring significant coding processes (Saidallah et al. 2016). AIMSUN, Paramics and VISSIM are also considered more flexible than the other simulators (Saidallah et al. 2016). However, one of the major conclusions from their article was the ability of most commercial simulators to "support the type and size of the vehicle, as well as taking into consideration the pedestrians and emergency vehicles, such as ambulances and police cars. They have the opportunity to simulate the public transport vehicles such as buses and trams, in contrast to open-source simulators" (Saidallah et al. 2016, p. 5). Nevertheless, only AIMSUN, MATSim, TransModeler and VISSIM support GIS, with MATSim is the only open-source simulator that uses GIS (Saidallah et al. 2016).

Through a review of literature implementing simulation platforms, as well as through the two reviews above, one of the most well-known and seemingly widely used open-source simulation platforms appears to be MATSim, which is an activity-based multi-agent microscopic simulation of transport developed in Zurich (MATSim Community, n. d.). The model has been used both by its developers (Balac, Ciari & Axhausen, 2015; Balac, Ciari, & Waraich, 2016; Ciari, Balac & Axhausen, 2016; Dubernet, Rieser-Schussler & Axhausen, 2016) and other researchers (Alemi & Rodier, 2016; Ayed, Khadraoui & Aggoune, 2015; Fagnant, Kockelman & Bansal, 2015; Ronald, Thompson & Winter, 2015; Ronald, Yang & Thompson, 2016). In fact, because of its notoriety Ciari, Balac and Axhausen (2016) provide a comprehensively review invoking the use of this platform for car-sharing modeling, while also summarizing its current limitations and on-going developments. Discussing the limitations, Ciari Balac and Axhausen (2016) suggest that MATSim's behavioral model assumes homogeneity in evaluation criteria for travelers regarding car-sharing and all other modes, thus, not capturing the individual or average preferences.

An additional limitation that is acknowledged by the authors is the extent to which MATSim is able to differentiate between different activities, which may need further refinement to sufficiently model car-sharing usage (Ciari, Balac & Axhausen, 2016). Ciari, Balac, and Axhausen (2016) also explain that car-sharing is known to fluctuate throughout the week, yet MATSim is limited to only single day simulations. Thus, the authors conclude that "the properties of agent-based modeling are particularly suitable to assess hypothetical scenarios on which limited previous knowledge is available, yet long-term effects of car-sharing are beyond the scope of the simulation" (Ciari, Balac & Axhausen, 2016, p. 19). In other words, the MATSim simulation environment enables the evaluation of a scenario, providing a snapshot, rather than what the author's term as "a time-dependent path view of things" (Ciari, Balac & Axhausen, 2016, p. 19).

While the authors lament some of the limitations of the platform, they provide a positive outlook by suggesting that MATSim is ideally situated to evaluate future circumstances through assumed behavioral changes, although work is needed on the behavioral model (Ciari, Balac & Axhausen, 2016). Ciari, Balac, and Axhausen, (2016) suggest that it could also be an ideal complement to other modeling techniques through its understanding of mobility behavior. In addition to this future value, the authors also show that MATSim is a very detailed platform and go so far as to state "in its current form [it] can already be used to obtain insight into how different operation strategies would work and to gain a feeling on how demand would be modified" (Ciari, Balac & Axhausen, 2016, p. 19). Furthermore, it can already account for mode substitution based on supply characteristics (Ciari, Balac & Axhausen, 2016).

2.9 Simulations, Ride-sharing, SAVs and Congestion

In lieu of some of the noted limitations, the use of modeling for traffic and congestion through simulation platforms is now widely used. This is mainly due to the technological paradigm shift in the field of transportation that necessitates a better understanding of future impacts of the implementation of new technologies, such as SAVs and optimal ride-matching services. Various academics and other researchers have employed the use of these simulation platforms to assess implementation of new technology, such as AVs and SAVs while additionally attempting to simulate optimal ride-sharing systems (Fagnant & Kockelman, 2018; Fagnant, Kockelman & Bansal, 2015; Shen & Lopes, 2015; Bischoff & Maciejewski, 2016).

Many simulation efforts are also geared towards understanding congestion issues and their ensuing negative externalities. One such example comes from Bischoff and Maciejewski (2016), who study congestion effects of autonomous taxi fleets through a multi-agent simulation of Berlin and Brandenburg. In their study, the authors model both real-time autonomous taxi operation and mixed autonomous/conventional vehicle traffic flow using MATSim (Bischoff & Maciejewski, 2016). To provide a comprehensive analysis of impacts from autonomous taxi fleets, the authors use various replacement ratios to estimate potential effects for different stages of inception (Bischoff & Maciejewski, 2016). From their simulation results, Bischoff and Maciejewski (2016) suggest potential positive traffic benefits from large-scale AV taxi fleets in cities with one autonomous taxi replacing between 10 and 12 conventional vehicles. The authors further found that proximity to the city center shows more significant positive benefits than moving further away (Bischoff & Maciejewski, 2016). However, the authors do qualify these benefits by stating that they are dependent on more fluent traffic flow for application at the large scale (Bischoff & Maciejewski, 2016).

Fagnant, Kockelman, and Bansal (2015) likewise employed the simulation platform MATSim to assess autonomous vehicle fleets in Austin, Texas. Using a low market penetration level of 1.3% of regional trips, the authors found that such an autonomous fleet could produce replacement rates of 1:9.3 of SAVs for conventional vehicles (Fagnant, Kockelman & Bansal, 2015). However, Fagnant et al. (2015) qualify these findings by suggesting that VMT from deadheading is likely to increase, especially in the early deployment stages of SAV fleets. Nevertheless, the authors provide an optimistic outlook in the long term through more expansive SAV fleets leading to greater efficiency.

In a similar study, Fagnant and Kockelman (2018) use MATSim to further the understanding of advancing technology in the transportation field by examining shared-autonomous dynamic ridesharing and fleet sizing for Austin, Texas. Building off of the work of Fagnant et al. (2015), Fagnant and Kockelman (2018) attempt to address the issue of added VMT through deadheading. Acknowledging that without any ride-sharing,

autonomous vehicle fleets will result in an additional 8.7% in VMT, the authors communicate the importance of dynamic ride-sharing in avoiding new congestion related issues (Fagnant & Kockelman, 2018). In fact, using conservative parameters for modeling the DRS, Fagnant and Kockelman (2018) found that pooling trips could cut added VMT from 8.7% to 4.5%. The authors went one step further to suggest that "as trip-making intensity rises and DRS parameters are loosened, greater ride-sharing and less relocation may actually reduce net VMT" (Fagnant & Kockelman, 2018, p. 157).

As consumer utility is of crucial importance to the success of autonomous mobility-on-demand services (AMOD), Shen and Lopes (2015) use the MobilityTestbed simulation platform to test and improve management of such services for the betterment of passenger experiences. The authors introduce their own algorithm, *Expand and Target* (EAT), which "dynamically expands the search space and targets the autonomous vehicles, in which the expansion and targeting can be viewed as a multi-agent, self-adaptive process" (Shen & Lopes, 2015, p. 5). Additionally, the authors develop three different scheduling strategies (no scheduling, static scheduling and online scheduling) to test the performance of their algorithm. Through their EAT algorithm, Shen and Lopes (2015) conclude that in all three scheduling strategies the ability of AMOD performance was significantly improved. Specifically, the authors found that the EAT algorithm reduced average passenger waiting time by nearly 30% and increased trip success rate by almost 8% (Shen & Lopes, 2015).

As shown in the review of simulation studies, the majority of such work claims various benefits to the transportation industry from the use of ride-sharing, mode-sharing, and autonomous mobility services, whether it be through reducing VMT, reducing waiting time, increased trip success, or otherwise. Nonetheless, it is important to note that most of these simulation studies focus on optimization and do not project completely rational forecasts for the adoption and usage of such technology. This leads many to use varying adoption rates to account for such fluctuation. However, until such technology has advanced beyond the infancy stage, especially in terms of deployment and adoption, it is unlikely for the precision and accuracy of these simulations to improve beyond optimization-based practices.

2.10 Conclusion

In this review, we focused on car-sharing and ride-sharing services, the associated technology that enables them, and their impacts on various characteristics of the transportation industry, such as impacts on travel behavior, travel demand, as well as recurring and non-recurring congestion. We are in the midst of a paradigm shift in the transportation sector and decision makers will need an understanding of the technology and the associated user choice to inform policy on the links between technology and driving choices in the southeastern region, where the auto-oriented built environment

influences the rate of adoption, cost and, ultimately and the supply and demand for these technologies. AVs and more importantly SAVs appear to be the future of the automotive sector of the transportation industry. Their inception by many is seen as a question of when, rather than if. AVs, SAVs, and ridesharing appear to be the future of the transportation field and arguably offer a more sustainable alternative to the nation's most polluting sector. Further research is needed to resolve this debate, specifically in regards to simulating the most realistic parameters of autonomous vehicle fleets and consumer utility patterns as they relate to technological advances in the field. This review provides a comprehensive synthesis of available literature, as well as available and upcoming technologies for car-sharing and ridesharing applications. Additionally, this review identifies available simulation platforms capable of simulating DRT, car and ridesharing modes, DRS, and peer-to-peer ridesharing. Finally, applications of simulation platforms are identified, including successful applications, barriers to and challenges for simulating DRT, lessons learned, costs, and benefits of the proliferation of car and ridesharing platforms.

3.0 STUDY OF SOUTHEAST MILLENNIALS-NORTH CAROLINA CASE STUDY

3.1 Introduction and Background

This chapter focuses on North Carolina millennials adaption to ridehailing. Recent studies on millennials and their travel behavior have focused on national trends; few have looked at the state level. Studies looking at state or regional level trends have focused on California (Circella et al. 2016), the Midwest (Villwock-Witte & Clouser, 2016), or more traditional urban metropolitan areas (Sakaria & Stehfest, 2013). Since these studies have been published, new issues such as student loan debt are emerging as possibly impacting transportation decisions (Zohdy, Huang, Keegan, & Lukens). North Carolina provides a case study of auto-oriented planning and how millennials navigate within this terrain.

3.1.1 State Characteristics

Demographics

North Carolina is racially and ethnically diverse. The state is one of several southeastern states with a rising Latino population accounting for almost 10% of the population, while its Black population comprises 22% of the population (Stepler & Lopez, 2016). (More information on North Carolina demographics are in Table 3-2).

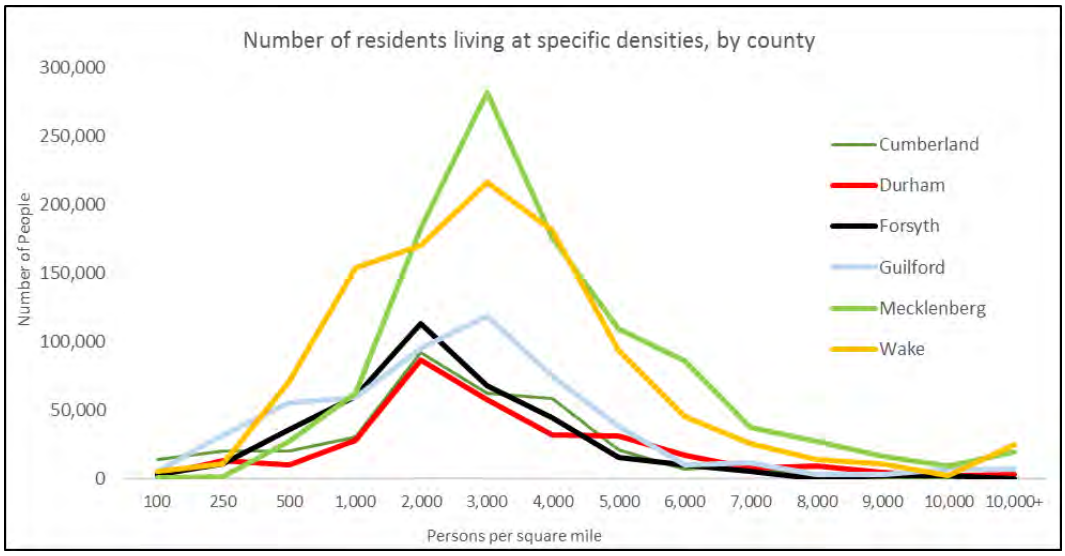


FIGURE 3-1 DENSITY OF COUNTIES CALCULATED BY NUMBER OF PEOPLE LIVING AT CERTAIN DENSITIES.

Source: Social Explorer Tables: ACS 2018 (5-Year Estimates) (SE), ACS 2018 (5-Year Estimates), Social Explorer; U.S. Census Bureau

Land Use

At its highest density, North Carolina's average density is only 1,347 people per square mile. Most respondents (48%) come from counties with population densities between 500-1,000 people per square mile. Thirty-seven percent come from higher density counties of over 1,000 people per square mile. The remaining 15% live in areas of less than 500 people per square mile.

In Figure 3-1, we calculated the densities of the largest counties in North Carolina and where most of our respondents are living. We took into account how many people lived in specific densities. As you can see, most people in these five counties are living in low densities that cannot support more frequent public transit, such as every 10-15 minutes. This would require densities of 5,000 people per square mile. As we later discuss (see 3.1.1.4), despite infrequent transit service, many millennials know how to take transit for their commute but choose not to because it is inconvenient. In line with this, according to the Bureau of Transportation Statistics, North Carolina's public transit ridership is also one of the lowest at 1.1%, while the national average is 5.2%. Therefore, it is no surprise that 81% of North Carolina residents drive to work, exceeding the national average of 76.4%.

Ironically, despite these disconcerting statistics, mobility sharing service companies continue to expand their networks in the state, including ridehailing, bikesharing (docked and dockless), electric scooter sharing, and carsharing services. And millennials have adopted some these services quite quickly (see 3.1.5).

3.2 Methodology

The study team used an online survey to identify awareness, usage, and proliferation of identified technologies with a focus on young commuters aged 22-37 years old in 2018. The online survey was comprised of questions addressing the familiarity with key mobile applications, usage of such applications, frequency of usage, age of respondent, perception of service, value, affordability, etc.

To obtain a better survey sample and after assessing options to cost-effectively target potential survey participants, we focused on millennials in urban areas of North Carolina. North Carolina has 100 counties; most of which are in rural areas with populations under 1000 people. Therefore, we used the Pew Center definition of millennials as those born between 1981-1996 (Dimock, 2018) to identify respondents and focused only on this age group.

We worked with an online research firm, Qualtrics, which sent out invitations to the survey via their panels. Respondents were eliminated from the survey if they were not located in North Carolina using geocoding and if they did not meet the age

requirements. Qualtrics was provided with a list of counties to sample from based on their mix of urban, suburban and rural. These counties included: Forsyth, Wake, Cumberland, Durham, Guilford, Orange, and Mecklenberg. All of the online surveys were collected by Qualtrics, Inc. that used this definition of millennial, race/ethnic group, and gender. We decided to oversample Black (n=189) and Hispanic (n=175) respondents to do stronger analysis amongst groups. We had limited resources and focused on these three groups to have sample size that allowed for cross-comparisons. Responses were collected from May 2018 through June 2018. Table 3-1 shows the racial and ethnic make-up of the respondents.

TABLE 3-1 RACE AND ETHNICITY BY COUNTY

County	City	Population	White	Black	Hispanic	% of respondents from county (n=623)
Cumberland	Fayatteville	332,766	45%	37%	11%	13%
Durham	Durham	300,865	42%	37%	13%	9%
Forsyth	Winston-Salem	364,362	57%	26%	13%	12%
Guilford	Greensboro	517,197	51%	34%	8%	10%
Mecklenberg	Charlotte	1,034,290	48%	31%	13%	24%
Wake	Raleigh	1,023,811	61%	20%	10%	14%

A total of 644 responses were collected and validated. After eliminating incomplete, duplicate, or irregular answers, a total of 623 responses were analyzed.

3.3 Results

In the following section, we describe the main findings from the survey and discuss their implications.

3.3.1 Socio-demographics of respondents

Questions we asked included: Age, Gender, Race, Ethnicity, City, County, Time in North Carolina.

3.3.1.1 Geographical location of respondents

Because ridehailing services seek out urban centers, we targeted the five largest metropolitan areas of North Carolina, a majority of our respondents came from

those areas. Table 3-1is a breakdown of the racial makeup of each of these counties and the percentage of respondents from each of these counties.

The overall racial and ethnic composition of millennials who took part in this research differed significantly from the state's racial composition (see Table 3-2). Because previous research did representative sampling, the numbers were quite small for Blacks and Hispanics. We purposely sought a larger sample size to do comparisons between the different groups that could include age, gender, in addition to race and ethnicity.

TABLE 3-2 RACE AND ETHNICITY OF THOSE SURVEYED COMPARED TO NORTH CAROLINA

Race or Ethnicity	Count	% of all respondents	North Carolina[1]
Non-Hispanic White	205	45%	70.8%
Non-Hispanic Black	189	34%	22.2%
Hispanic	175	28%	9.5%
Other	59	9%	----
Two or more	35	6%	2.2%
American Indian/Alaska Native	13	2%	1.6%
Asian/Native Hawaiian Pacific Islander	24	4%	3.2%

[1] Source: American Community Survey 2017.

3.3.1.2 Gender, Race and Ethnicity

As shown in Table 3-3, overwhelmingly most respondents were women. Only Hispanics had a more even split between men and women.

TABLE 3-3 GENDER RACE AND ETHNICITY

	All n=623	Non-Hispanic White n=205	Non-Hispanic Black n=189	Hispanic n=175
Female	65%	76%	64%	51%
Male	34%	23%	36%	49%
Other	1%	1%	--	--

3.3.2 Tenure in North Carolina

North Carolina has had an annual influx of new residents. We asked how long a respondent had lived in North Carolina. Of the 623 respondents, only 6% were new arrivals having only been in the state less than a year, 23% have been in the state 1-5 years, 37% over 5 years, and 34% stated they were born and raised in North Carolina.

3.3.3 Living Arrangements

Questions we asked included: Type of residence and tenure; living situation; reasons for living with parents or other family members, if selected; household size; presence of children and if they attend school outside of the home; employment status; education level; student loan debt and status.

3.3.3.1 Housing

In terms of housing, White, Hispanic, and multigenerational millennials were more likely to live in single-family homes (see Table 3-4). A higher percentage of Black millennials lived in apartments than any other cohort. Hispanics had larger households with an average of 3.26 and more children under 18 living in the home (55%).

TABLE 3-4 HOUSING TYPE, FAMILY SIZE, AND CHILDREN UNDER 18 IN HOUSEHOLD

Housing type	All n=623	NH White n=205	NH Black n=189	Hispanic n=175	Multi-generational n=120
Single Family Home	55%	58%	48%	59%	74%
Apartment	33%	30%	38%	31%	18%
Condominium	9%	11%	12%	5%	5%
Other	3%	1%	2%	5%	3%
Household size					
Average household size (persons)	2.83	2.81	2.83	3.26	3.46
With children <18 years old	49%	47%	46%	55%	44%

Table 3-5 shows the living situation by race and ethnicity. As shown, 19% of all respondents were living with parents or other family members. The majority of these millennials were Black.

Blacks were more than twice as likely to live with family members as whites, 27% versus 12%. Only 11% of all multigenerational millennials were full-time students, while 20% were unemployed. For Black multigenerational respondents, 23% were unemployed.

Respondents stated three main reasons for living with parents were: "Rents are high in my areas/I need to save money," "I moved back to help my family and/or relatives," and "Student loan payments make it difficult on live on my own or with roommates." Student loan payments were not a major factor for Black respondents; most moved back to help with family and/or because of high rents.

TABLE 3-5 LIVING ARRANGEMENTS BY COHORT

Living Arrangements	All n=623	Non-Hispanic White n=205	Non-Hispanic Black n=189	Hispanic n=175
Married living with spouse	35%	51%	20%	34%
Living with significant other	16%	12%	20%	18%
Living with parents/family	19%	12%	26%	18%
Living with roommate/friends	9%	6%	7%	12%
Living alone	18%	19%	23%	15%
Other	2%	1%	4%	3%

3.3.3.2 Dependents or multigenerational households?

In this survey, we categorize respondents as multigenerational *or Multigen*, as opposed to dependent, if they lived at home with parents or family members. In contrast to other surveys, we asked their reasons for this living arrangement. In our survey, we allowed respondents to choose multiple reasons. Of the 120

respondents that live at home, we found that 18% of these respondents chose "living with family members to assist them" and no other reason. Another 23% chose "rents are high in my area/I need to save money." In future iterations, we will phrase this question to make a clear differentiation between independence but helping family and dependence because of rental costs, unemployment, etc. In addition, even if you live outside of your parents' home, you may still receive some financial assistance from parents/guardians. Some anecdotal evidence points to this.

In addition, multi-generational who worked full-time stated they lived with their parents or family members to help family/relatives 24%. This portrays a more complex picture of what dependency looks like. As stated before, more research needs to be done. Therefore, we are limited in our understanding of financial dependence or independence. We cannot solely label someone as financially independent if they still receive income from family members despite living apart. The place where someone lives or whom they live with needs to be further studied and differentiated.

3.3.4 Employment Status

Table 3-6 summarizes the employment status of responders and Table 3-7 their educational level. Approximately 70% of all respondents were employed, 9% were full time students, 9% managed households, but 10% were unemployed.

TABLE 3-6 EMPLOYMENT STATUS

	All n=623	NH White n=205	NH Black n=189	Hispanic n=175	Multigenerational n=120
Employed, full-time	49%	55%	47%	50%	38%
Employed, part-time	14%	13%	12%	14%	18%
Self-employed	7%	4%	9%	9%	4%
Full-time student	9%	6%	7%	11%	12%
Manage Household	9%	11%	8%	7%	4%
Unemployed	11%	10%	15%	9%	23%

TABLE 3-7 EDUCATIONAL LEVEL OF PARTICIPANTS

Highest degree earned to date	All (n=623)	White (n=20)	Black (n=189)	Hispanic (n=175)	Multigenerational (n= 120)
Less than high school	3%	0%	3%	4%	3%
High school graduate	21%	12%	27%	27%	30%
Some college	26%	21%	28%	29%	32%
2-year degree	10%	11%	9%	9%	10%
4-year degree	28%	41%	24%	17%	20%
Professional degree	10%	12%	8%	11%	4%
Doctorate	2%	2%	1%	2%	1%
Other	1%	0%	1%	2%	0%

The unemployment rate was highest for Black millennials at 15%. This is similar to the state unemployment rate. They also had the lowest full-time employment rate of 47% in comparison to Hispanics with 50% and White millennials with 55%.

3.3.4.1 Student Life and travel behavior

Questions we asked included: type of university, educational objective, travel mode to school, availability of public transit to school, distance between home and school, employment status, main mode of transportation prior to attending university.

Of the 623 respondents, 9% or 57 respondents were full-time students. Most attend public schools, with only 19% attending private schools and 7% attending for-profit schools. In terms of racial and ethnic composition, 37% of students were non-Hispanic Blacks, 37% were non-Hispanic White, and 33% were Hispanic. Travel behavior of students, including commute distance, mode, is discussed in Section 3.1.2 Transportation.

As can be seen from Table 3-8, most university/college students commute to school by car. The 14% of respondents that stated “Other” were doing online courses.

TABLE 3-8 COMMUTE TO SCHOOL VERSUS PREVIOUS MODE OF ALL FULL-TIME STUDENTS

Mode	Previous to attending	Current
Bike	2%	2%
Car	60%	47%
Rides from other people	7%	2%
Other	2%	14%*
Public transit (bus, subway, light rail)	18%	16%
Walk	12%	18%
Ridehailing service (Lyft, Uber)	---	2%

**Online school (5); Scooter (1); None of the above (2)*

Interestingly, when asked if they could take public transit to get to work, almost half or 48% answered that they could take public transit but it was inconvenient. On one hand, it means that they were aware of public transit and nearby routes, but on the other hand, better service including less transfer and more service hours would be needed to get them to use public transit.

We also asked what their main mode of transportation was prior to attending their current school to see if a major life event had shifted their mode. As shown in Table 3-8, driving was the main mode. Since going to school their car use has seen a 13% decrease. That may be further be explained by a similar number of students using online services—14%. A slight decrease in public transit and getting rides from other people could be attributed to greater numbers walking to school and using ridehailing services. And also, many of them live within walking distance to their schools.

When we looked at mode to school and distance, most lived less than five miles away from their school. Yet, many students also worked, a total of 40%, and these students also used a vehicle to get from school to work. Additionally, in North Carolina, they may also not be able to safely walk to a bus stop or public transit hub because of a lack of sidewalks. For example, the City of Durham recently conducted a study identifying areas with non-continuous sidewalks that calculated the cost at over $500 million dollars.

3.3.4.2 Student loan debt and payment status

Forty four percent of all millennials had student loan debt, ranging from less than $1,000 to over $100,000. Fifty two percent of Blacks, 46% of Whites, and 40% of Hispanics have student loans as well.

When considering how debt is distributed amongst all respondents and by race and ethnicity, a larger percentage of Blacks reported of had student loans (52%) than whites (40%) or Hispanics (40%), but a lower level of overall education. Further analysis needs to be done to understand the full impact of student loan debt and its relationship to lack of licensing and/or vehicles. The role of student loan debt and its impact on travel behavior has only recently been studied. Zohdy, et al. studied the impact of student loan debt on transportation choices analyzing several datasets (Zohdy, Huang, Keegan, & Lukens, 2016). They found that: "the student loan impact may be moderated more by fluctuations in income than the mere amount of debt that Millennials owe." Given that Blacks already face significant transportation challenges, increased student loan debt may add further stress, especially if they are unable to complete their degrees.

3.3.5 Transportation

Questions we asked included: typical transportation modes used in a week, the previous day, licensure, number of vehicles available in the household, vehicle ownership status, commute mode of transportation, distance between home and employment, availability of public transit to work, employment transportation benefits provided (e.g. free parking, transit pass, etc.).

3.3.5.1 Modes used in a week

Respondents were asked to select the modes of transportation they used in a typical week. Interestingly, 10% stated they used a ridehailing service, 53% stated they only used cars, and 18% claimed to use public transit.

3.3.5.2 Commuting behavior

Of the millennials who are employed, 79% drive to work; this mirrors North Carolina's average of 80.1% (see Table 3-9).

TABLE 3-9 COMMUTING BY MODE OF EMPLOYED RESPONDENTS

	All	NH-White	NH-Black	Hispanic	Multigen	Student (n=57)

	n=409*	n=141	n=120	n=112	n=69	
Drive	79%	84%	78%	79%	68%	47%
Get a ride	6%	3%	6%	7%	16%	2%
Public Transit	7%	6%	8%	7%	7%	16%
Walk	3%	4%	3%	2%	4%	18%
Bike	1%	---	3%	----	----	2%
Other	3%	3%	3%	5%	4%	14%

*This table only includes respondents who work outside of the home. It does not include the self-employed who work from home, household managers, or the unemployed.

3.3.5.3 Commute Distance

As Table 3-10 shows most millennials commute distance is 10 miles or less. On the surface, this would lend to more public transit. However, taking density into account, as stated previously, most people do not live in dense areas.

TABLE 3-10 COMMUTE DISTANCE TO WORK BY COHORT

	All (n=401)	White (n=140)	Black (n=116)	Hispanic (n=111)
< 5 miles	19%	14%	19%	23%
5-10 miles	30%	29%	33%	29%
10-15 miles	23%	23%	24%	26%
15-20 miles	17%	24%	12%	15%
> 20 miles	10%	10%	12%	7%

3.3.5.4 Public Transit Knowledge

Respondents were asked if they could use public transit for their commute; 52% stated they could use public transit but service hours and transfers kept them from using it. Less than a third stated no public transit went to their work site; others stated they did not know. No significant difference existed between racial/ethnic groups or dependent millennials in terms of most commute modes except getting rides—15% of dependent millennials were more likely to get rides than take public transit (7%).

Respondents who answered that they drove to work were then asked, "Could you use public transit to commute to work?" We asked if the reasons they could not take public transit were because they did not know or if it was because of inconvenience (See Appendix 8.1, survey question 28). If public transit were more convenient or had more hours, 47% of employed millennials (n=323) who drive to work could take public transit to their jobs. Only 12% did not know if they could take public transit.

3.3.5.5 Employee Transportation benefits

Employed respondents were asked about the types of employee benefits they received or were offered. Unsurprising, 63% stated they were given employee parking or discounted parking. Interestingly, 22% stated a carshare membership, carpooling was mentioned by 10%, and discounted transit passes were stated by 14% or respondents. Other types of benefits included gas reimbursement, mileage reimbursement and vehicle discounts.

3.3.6 Licensing

Licensure information of respondents is summarized in Table 3-11.

TABLE 3-11 LICENSURE OF ALL MILLENNIALS AND BY DIFFERENT COHORTS

	All n=623	NH White n=205	NH Black n=189	Hispanic n=175	Multigen n=120
Yes: Has a license	82%	92%	74%	79%	64%
No: Plans to get one	12%	3%	18%	11%	25%
No: Does not want one	5%	4%	6%	4%	5%
Other	1%	1%	2%	5%	6%

Overall, seventeen percent of those surveyed lack a driver's license. Five percent stated they did not want a license; the other 12% planned to get a license within the next year. Most of those without a license were Black, as 26% percent of all Black respondents were without a driver's license. The low license rate amongst Blacks has not changed over time; this echoes the 2000 Census for North Carolina (Census). Only six percent of black respondents stated they did not want a license. More research needs to be done to understand why they do not want a license.

3.3.6.1 License and Access to Household Vehicles

As reported in Table 3-12 seven percent of all millennials both lack a license and live in Zero car households. By cohort, 11% of Black millennials lived in Zero car households and had no driver's license; for Hispanics, it was seven percent, and for white millennials it was less than one percent.

TABLE 3-12 HOUSEHOLD VEHICLES BY COHORT

# of vehicles	All (n=623)	NH White (n=205)	NH Black (n=189)	Hispanic (n=175)
1	43%	39%	47%	41%
2	38%	48%	28%	40%
3	8%	9%	8%	8%
4 and over	2%	1%	2%	3%
Zero	9%	2%	16%	7%

3.3.6.2 Vehicle Ownership: Zero car millennials

In previous studies, access to a vehicle seemed to only consider within the household and not outside the household. Studies have shown, however, that Zero car households may borrow or have access to vehicles outside of the home (Clifton, 2004). For this study, we asked how many cars were in the household and for those who did not have a household car, we asked about access outside of the house, as shown in Table 3-13.

TABLE 3-13 VEHICLE OWNERSHIP BY COHORT

Type of ownership	All (n=623)	NH White (n=205)	NH Black (n=189)	Hispanic (n=175)	Student (n=57)
Own/lease car	76%	87%	69%	77%	68%
Access to car in home	9%	7%	8%	9%	14%
Access to car outside of home	3%	1%	3%	2%	4%
Plan to purchase/lease	7%	1%	12%	8%	9%
Do not plan to own a car	4%	2%	7%	3%	5%

As shown in Table 3-13, of the 189 Black respondents, 16% lack a household vehicle. However, that number was reduced to 7% when taking into consideration access outside of the home.

3.3.7 Technology and sharing mobility use

Questions we asked included: *Cellphone Usage and Types of Apps:* what the cellphone is used for, social media and navigation apps (e.g. snapchat, level of familiarity and frequency of use); ridehailing services (familiarity, frequency, type of trip, last used and reason); bikesharing/carsharing familiarity; food delivery service (UberEats, DoorDash, etc.)

About 98% of respondents stated they had a cellphone. By and large, most respondents (89%) used it for social media or social media in combination with watching videos, communication, navigation, and listening to music. Of significance, no noteworthy differences existed among different racial or ethnic groups in terms of use, type of apps used, or frequency.

3.3.7.1 Ridehailing Services

Most millennials surveyed have used ridehailing services—67% having used Lyft, Uber, or both (see Table 3-14). Over one-third stated they had used Lyft or Uber either daily or weekly and the same amount had used the service within the past 30 days.

TABLE 3-14 RIDEHAILING SERVICE USAGE BY COHORT

	All (n=623)	White (n=205)	Black (n=189)	Hispanic (n=175)
Ridehailing use (% of each cohort)	70%	72%	67%	70%

While previous studies have pointed to concerns over who can actually use or have access to carsharing, ridesharing, and ridehailing services (Shaheen, Cohen, & Zohdy, 2016; Shaheen, Cohen, Zohdy, & Kock, 2016), this suggests that education and race had no significant influence on use or frequency of ridehailing for North Carolina millennials. By race, 68% of Blacks, 72% of whites, and 71% of Hispanics reported having used these services (see Table 3-15).

TABLE 3-15 RIDEHAILING FREQUENCY BY COHORT

	All (n=440)	White (n=147)	Blacks (n=128)	Hispanic (n=123)
Daily	13%	11%	14%	16%
Weekly	20%	14%	15%	26%
Monthly	20%	23%	18%	18%
A few times per year	47%	52%	53%	40%

Differences did exist in terms of reasons for use, as seen in Table 3-16. Over a quarter of Black millennials who used the service used it to get to work as did 23% of Hispanics. All cohorts used it for entertainment purposes. In terms of "Other", respondents stated "going to church," "ordering a ride for someone else," "picking up their car", or "finding a ride when they have been drinking."

TABLE 3-16 REASON FOR MOST RECENT USE OF RIDEHAILING SERVICE

	All (n=440)	White (n=147)	Blacks (n=128)	Hispanic (n=123)
School	6%	1%	10%	8%
Work	23%	18%	27%	23%
Airport	17%	20%	13%	17%
Errands	15%	9%	16%	18%
Entertainment[a]	31%	38%	27%	29%
Other	8%	14%	7%	5%

[a]Entertainment includes going to restaurants, bars, movies, concerts, etc.

3.3.7.2 Bikesharing/Carsharing Service

In addition to the use of ridehailing services, 15% of respondents have used bikeshare, carshare, or both. Given that some of these programs have just arrived in North Carolina, this is encouraging. A total of 9% of all millennials surveyed had used a carsharing service, 2% had used both bikeshare and carshare, and 4% had used a bikeshare service only.

3.3.7.3 Food Delivery

Fifty-one percent of all millennials surveyed stated they had used a food delivery service such as UberEats, InstaCart, GrubHub or the like. In terms of frequency, as shown in Table 3-17, many use it quite frequently, with 63% either monthly or weekly. Many also reported that they had used the service within the last 30 days (69%). With greater use of food delivery and other types of delivery service, increased traffic from deliveries could emerge as an issue in the near future, especially since the road network in many cities is lacking.

TABLE 3-17 USE, FREQUENCY, MOST RECENT USE OF FOOD DELIVERY SERVICE

Have you used a food delivery service?	**All (n=623)**
Yes	51%
No	49%
Frequency	Ridehailing users (n=317)
A few times per year	37%
Monthly	42%
Weekly	21%
Most recent use	Ridehailing users (n=317)
Over 30 days ago	30%
Within the past 30 days	42%
Within the past 7 days	27%

3.4 Conclusions, Recommendations & Future Research

This study helps demonstrate that even in states with small urban areas and lower densities, millennials are aware of and are taking advantage of ridehailing, carsharing, and ridesharing services. Most millennials surveyed have used ridehailing services—with 66% having used Lyft, Uber, or both. Over one-third stated they used Lyft or Uber either daily or weekly and the same amount had used the service within the past 30 days. While previous studies have pointed to concerns over who can actually use or have access to carsharing, ridesharing, and ridehailing services (Shaheen, Cohen, & Zohdy,

2016; Shaheen, Cohen, Zohdy, et al., 2016), initial findings suggest that education and race had no influence on use or frequency.

Initial findings suggest that ridehailing services have become part of the norm. Over 65% of millennials surveyed have used Lyft or Uber or both services; many on a fairly regular basis. We found no significant differences in use or familiarity amongst ethnic or racial groups. Therefore, ridehailing services may be a way to mitigate accessibility issues. Millennials in North Carolina have adapted to these services and more programs are being launched every week. More analysis and research are needed to see how to increase access to these ridehailing and ridesharing services.

In line with other studies, we found that car-less millennials are more likely to be minorities and economically disadvantaged, particularly Blacks. Black respondents were more likely than other ethnic/racial group to live in households without vehicles, lack a driver's license, live with parents/family members, and/or unemployed. Access to a personal vehicle is critical for most residents. As Klein and Smart posit (Klein & Smart, 2017): "Lack of car ownership, inadequate public transit service in many central cities, and metropolitan regions with a high proportion of "captive" transit dependents exacerbate social, economic, and racial isolation, particularly for low-income minorities who have limited transportation options (p.78)".

Taking into consideration that Black millennials comprise 25% of the Black population, it is critical that transportation equity measures are strengthened at the local level, or else "car-less-ness" will continue its pattern of targeting the most vulnerable (Ralph, 2017).

Moreover, there appears to be a latent demand for public transit. A significant percentage of millennials (47%) stated that they could take public transit, but service availability or inconvenience made this impossible. This means that they are familiar with public transit and had considered using it at some time. These millennials may be open to different types of carsharing, carpooling types of commuting programs, in addition to public transit.

Furthermore, 20% of full-time students reported that they took public transit prior to attending university. This is another example of how millennials may be more amenable to using public transit, if not for every commute/school trip, at least for other types of trips.

While some millennials may not be considered financially dependent, given that a number of respondents stated they moved back to help family and were full-time employees, more research needs to be done to understand if the title "dependent" is fairly applied. Millennials may be moving back to financially help their parents or family, instead of the widely-assumed reasons. Also, a recent report has pointed out that independent millennials may still be receiving financial support from their parents

despite living separately from them. Also, the length of the stay is another aspect that should be considered.

In addition, the rise of food delivery service is also concerning. In the next iteration of this survey, we will be looking at the amount of total deliveries millennials and Generation X receive on a weekly basis, from not only food services but retail outlets such as Amazon and the like. New concerns have been raised by many urban areas about the conflicts over curb access as well as increased congestion coming from delivery service. While people may be driving less in some cases, this may be compensated for by using delivery services, thus transferring congestion to another user rather than reducing congestion.

More analysis needs to be undertaken to see how findings from this study compares to the California and the Midwestern studies. Because we were limited in budget, we only focused on millennials. In the next iteration, we will look at two states and two generations. In addition, in the next phase, we will be conducting a study that looks at multiple generations to see how prevalent the issues of transportation inequity are as well as how technology has become ingratiated in people's lives.

Racial differences did emerge in our research. Findings suggest that Black millennials surveyed were more likely than white or Hispanics to live in households without vehicles or to lack a driver's license. This supports findings from Klein and Smart about vehicle ownership and access amongst minorities, particularly Blacks (Klein & Smart, 2017). Blacks also had a higher amount of student loan debt but fewer degrees; student loan debt may further inhibit their transportation choices in the future.

4.0 TRAVEL PREFERENCES AND ATTITUDES OF BIRMINGHAM TRAVELERS

4.1 Introduction

In the recent years, Transportation Network Companies (TNCs) such as Uber/Lyft have led to an expansion of on-demand ride sharing transportation options. Despite the rapid growth of several TNC markets, analysis of potential and actual impacts of TNCs presence on preferences and daily travel patterns of TNC-aware transportation system users are still very limited. Such analyses are hindered by the lack of availability of detailed data due to privacy concerns, as well as technical and financial feasibility issues.

The objective of this part of the study is to understand current travel preferences and practices of transportation users in the Birmingham Metropolitan Area and document their attitudes toward TNC use as a travel mode of choice. To meet this objective, we developed a comprehensive travel diary questionnaire survey and used it to survey a TNC-aware population sample of 451 respondents in the Birmingham Metro Area. The survey requested participants to report detailed trip information for a typical day (i.e., 24-hr travel diary) including origin and destination of each trip, travel time, trip purpose and travel mode used. Demographic data were also obtained and used in the analysis and interpretation of survey findings. The analysis helped to identify indicators that contribute to the use of TNCs and, thus, can create a shift in the travel pattern of TNC-aware populations when TNC services are available in a region. Moreover, the detailed travel diary records provided useful information for the population synthesis performed in Chapter 5.

4.2 Background

In recent years, a service commonly referred to as dynamic ridesharing has emerged. Such service is provided by ride-hailing platforms (such as Lyft and Uber) and arranges one-time rides on an on-demand basis. The use of information and communication technologies through TNCs has the potential to impact travel patterns and modal choices in a region. The literature confirms that technology-enabled services can affect travel behavior in dynamic ways by providing more travel options, reducing travel uncertainty, and potentially replacing other modes (Alemi et al. 2018). Recent research by Sivak (2014) states that the percentage of zero-vehicle households may also increase as a result.

TNC services introduce added convenience to the user and may impact auto ownership and driving licensure trends. However, their impact on transportation network operation is not clear. For example, 40% of TNC users in San Francisco reported that due to the adoption of on-demand mobility sharing services, they use their private vehicle less (Rayle et al. 2014). Others argue that low fare and high availability of TNCs results in

replacement of public transit and taxi trips and increase in vehicle miles traveled (VMT) as TNCs are hovering at certain locations waiting for service calls. Also, little is still known about the reasons that motivate travelers to use TNCs over other modes. Thus, there is a need to identify the determinants that influence the selection and use of TNCs by documenting and analyzing day-to-day travel behavior as well as travelers' perceptions, attitudes towards TNCs.

The study reported herein sheds some light in this direction as it documents current travel preferences and practices of transportation users in the Birmingham Metropolitan Area and their attitudes toward TNC use as a travel mode choice. The data were collected by an online questionnaire survey developed at UAB that sought demographic information, user travel preferences, as well as a 24-hour detailed trip diary of a micro-data sample of Birmingham population. The survey responses were analyzed to document travel patterns and project influential factors in the travel mode choice of Birmingham travelers.

4.3 Methodology

4.3.1 Survey Questionnaire Development

This section of the report discusses the use of a survey developed in this study to identify awareness, usage, and proliferation of identified technologies among transportation system users in Birmingham, AL. To capture such data, an online questionnaire survey was designed in accordance with the ITE Manual on Transportation Engineering Studies (ITE, 2011) and used to obtain information about travel preferences, typical trips, and demographic data. First, an approval was obtained from the Institutional Review Board (IRB) for Human Use to proceed with the survey. The Qualtrics Research Core tool was used to prepare the questionnaire as it provided a user-friendly platform. The questionnaire was modified at various stages and was pretested and fine-tuned prior to use to ensure that it was easy for survey participants to understand the question and provide answers.

The questionnaire asked transportation users about their preferences towards using TNCs, frequency of use and reason for selection, along with demographic information such as gender, age, annual income, education level, and vehicle ownership. The criteria for collecting the demographic data were adopted from the Census criteria. Moreover, the questionnaire solicited detailed 24-hours trip information of the survey participants on a typical day. In the determination of the exact locations of origin and destination of the trips on 24-hr travel diary, we used Google maps API key application. This allowed survey participants to easily insert the location of their origins and destinations.

The survey was administered in the Birmingham, AL region between December 2018 and January 2019. Given a population of 1,141,309 capita in the Birmingham Metro Area as per the 2016 Census data, a sample of 420 responses was deemed sufficient according to the formula shown in the Equation 1 for calculating standard population sample size.

$$n = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + (\frac{z^2 \times p(1-p)}{e^2 N})} \qquad Eq.(1)$$

where n is the sample size, z is the z-score for the corresponding confidence interval, e is the margin of error, N is the population size as per latest Census reports, and p is the standard deviation (assumed to be equal to 0.5).

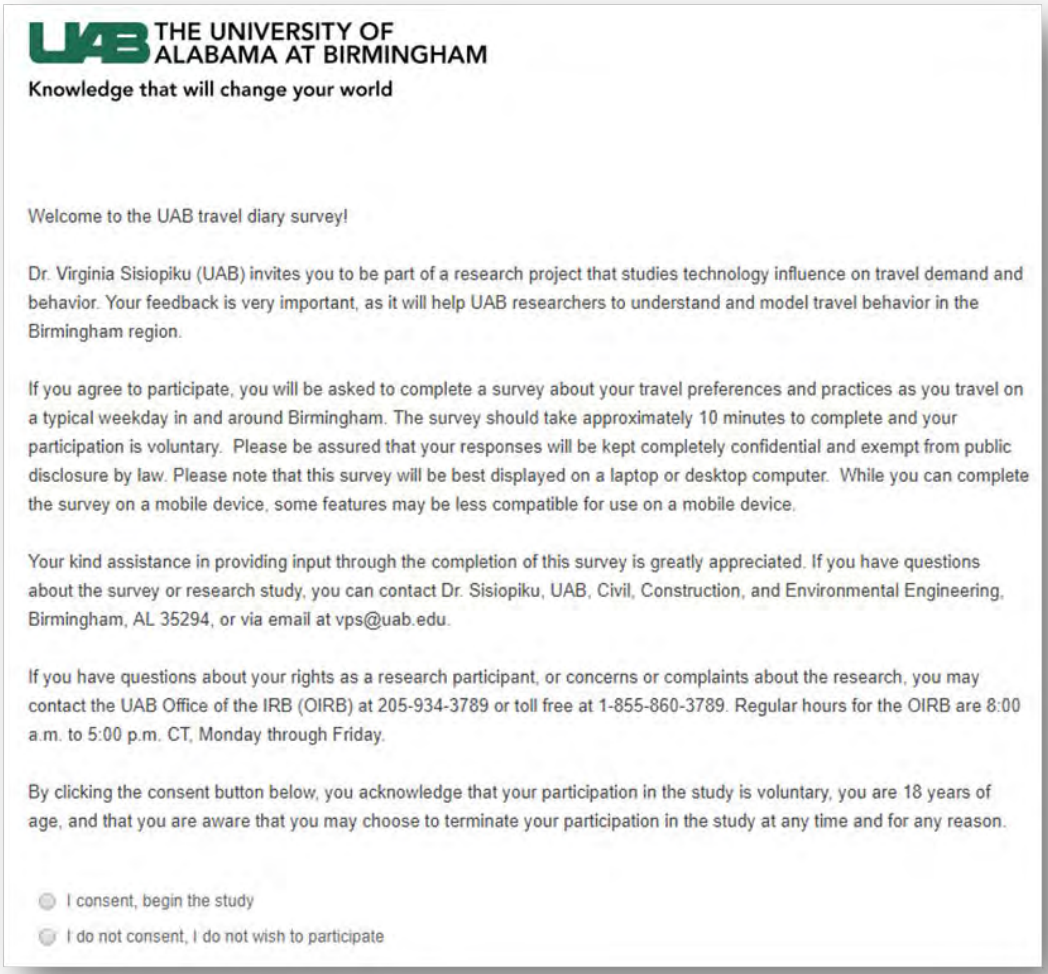
We went through a detailed data verification process to check the responses received using ArcGIS software, built-in tests, and through close manual observation. Several responses were deducted from the database and new responses were collected to replace those that did not pass validation tests or showed mismatch of reported data. A final database of 451 responses from Birmingham Metro Area was used in the analysis of this paper.

4.3.2 Survey Development Tool

The Birmingham Travel Diary Survey was developed by the research team at UAB and conducted using the Qualtrics Research Core. Qualtrics LLC facilitated the identification and recruitment of survey participants and automated the data entry and management process through the use of their software. While developing the survey, the questionnaire was segregated into seven blocks each with the purpose of collecting distinct type of information. The developed seven blocks and the tools used to shape them are as follows:

4.3.2.1 Block 1

This block consisted of a cover page to introduce potential participants to the scope of the survey, encourage them to participate, and request their consent (Figure 4-1). The cover page included a welcome, a brief description of the purpose, format, and expected time commitment, a statement of the participants' rights to privacy, and an invitation to participate in the survey. The participants were provided an option of giving consent and whether to continue to participate in the survey. They were also informed in this page that they can move out from the survey at any certain point of the survey if they wish to.



THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

Knowledge that will change your world

Welcome to the UAB travel diary survey!

Dr. Virginia Sisiopiku (UAB) invites you to be part of a research project that studies technology influence on travel demand and behavior. Your feedback is very important, as it will help UAB researchers to understand and model travel behavior in the Birmingham region.

If you agree to participate, you will be asked to complete a survey about your travel preferences and practices as you travel on a typical weekday in and around Birmingham. The survey should take approximately 10 minutes to complete and your participation is voluntary. Please be assured that your responses will be kept completely confidential and exempt from public disclosure by law. Please note that this survey will be best displayed on a laptop or desktop computer. While you can complete the survey on a mobile device, some features may be less compatible for use on a mobile device.

Your kind assistance in providing input through the completion of this survey is greatly appreciated. If you have questions about the survey or research study, you can contact Dr. Sisiopiku, UAB, Civil, Construction, and Environmental Engineering, Birmingham, AL 35294, or via email at vps@uab.edu.

If you have questions about your rights as a research participant, or concerns or complaints about the research, you may contact the UAB Office of the IRB (OIRB) at 205-934-3789 or toll free at 1-855-860-3789. Regular hours for the OIRB are 8:00 a.m. to 5:00 p.m. CT, Monday through Friday.

By clicking the consent button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

- I consent, begin the study
- I do not consent, I do not wish to participate

FIGURE 4-1. BLOCK 1 OF THE BIRMINGHAM TRAVEL DIARY SURVEY.

4.3.2.2 Block 2

The intent of the survey was to collect data from transportation network users residing in the Jefferson and Shelby counts in the Birmingham region. This block was used for the validation of Zip Codes provided by the survey respondents. This block asked participants to insert their Home Zip Code. If the code was in the list of survey area Zip Codes, participants were directed to next block. If the code was out of the survey area, then the survey participant was redirected to the end of the survey with a thank you message. We used “Matches Regex” function from the “Skip logic” of the Qualtrics tool to complete this check and had to use “inverse” logic of Java in the “Matches Regex” to choose the marked Zip Code for allowing participants to continue the survey.

4.3.2.3 Block 3

This block was divided in two sets of questions. Both sets started with a question about travel modes that survey participants used in the past year. If respondents included TNCs as one of their travel modes, then they were asked three questions about TNCs, i.e., (a) when was their last TNC use, (b) reason for choosing TNCs and (c) Trip purpose for the trips performed by TNCs. On the other hand, if survey participants had not used TNCs in the past year, they received the second set of question which asked them about their indifference towards the TNCs. In both cases, participants were provided with some common reasons as well as an option stated as "Other" where they could write in their answer with a text box.

4.3.2.4 Block 4

Block 4 requested the participants' initial location at 12:00am midnight in a typical day along with the type of location (home, school, work, nightlife/bar etc.). To make it easier to the survey participant, the location was collected using Google Map API (application program interface) key (Figure 4-2). It prompted the participants to locate their address by typing key words and then select their address from the map provided below the box (Figure 4-3). A major advantage of using the Google Map API capability was to obtain the exact latitude and longitude of a location. This was helpful for determining trip origins and destinations and calculating trip lengths.

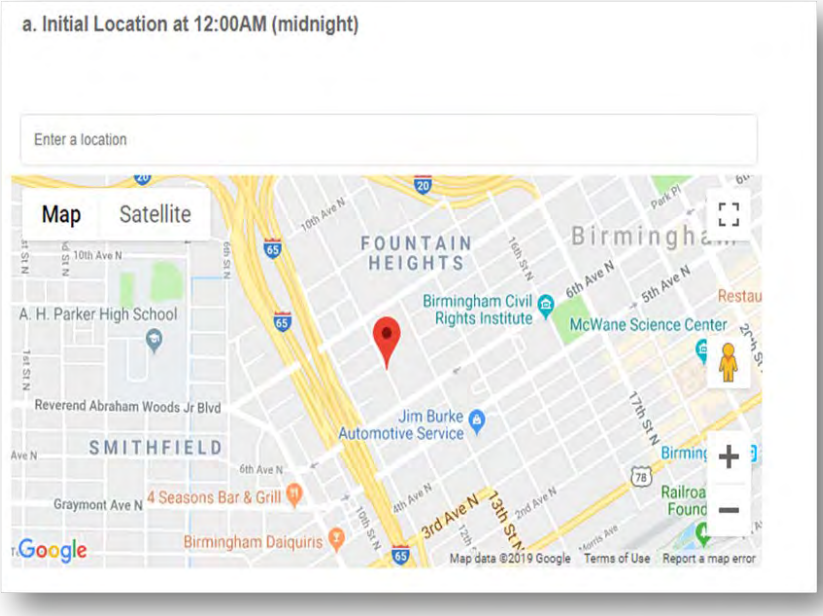


FIGURE 4-2. USE OF GOOGLE MAP API KEY TO COLLECT INITIAL LOCATION INFORMATION.

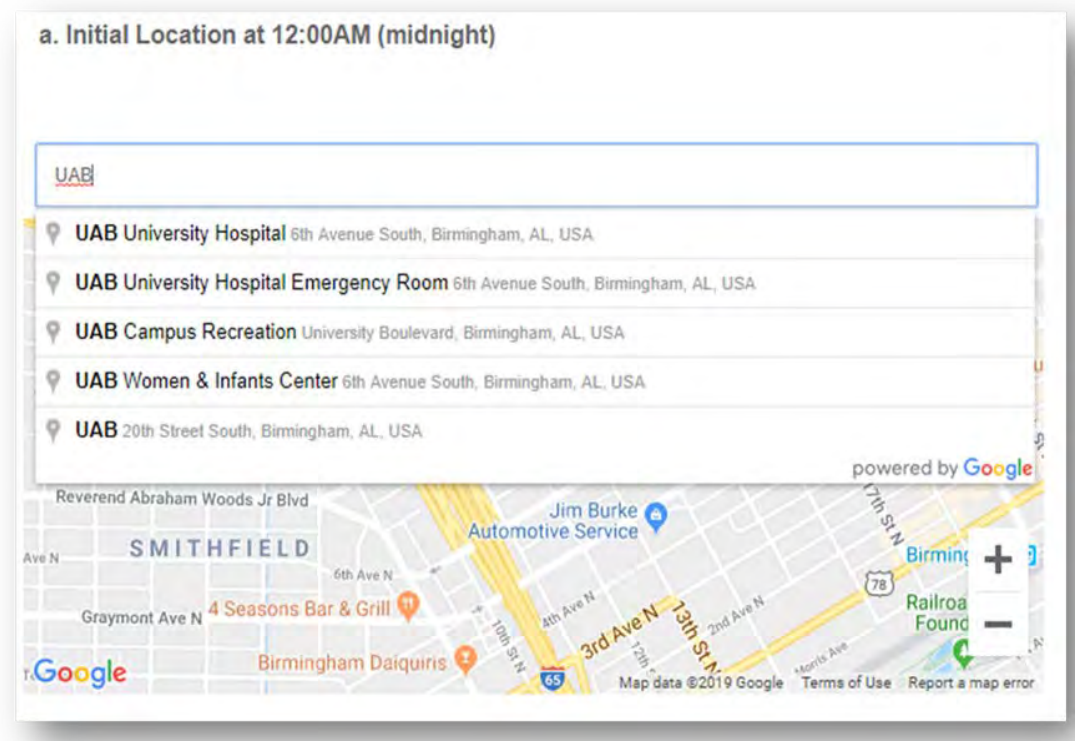


FIGURE 4-3. USING KEY WORDS TO FIND THE EXACT LOCATION.

4.3.2.5 Block 5

This block requested to collect information for each and every trip that the survey participant conducted on a typical day. The block started with asking about details of the first trip of the participant including the destination, departure time, arrival time, trip purpose and the mode used. If the mode was selected as “Uber/Lyft”, the participant was prompted to a new box to provide the information about TNCs such as waiting minutes, preferred company, and vehicle availability.

After the information about the first trip was entered, the survey participant was asked if this was his/her last trip of the day. A “no” response triggered a request to record information about his/her second trip of the day. The process continued until the respondent reported that a trip was his/her last trip for the day.

To automate the process, this block used the “Loop and merge” tool of Qualtrics through the last question as shown in Figure 4-4. If “yes” was selected as the answer, the survey continued to the next block. If “No” was selected as the answer, the survey returned to the start of the Block 5 allowing the participants to report details about their next trip. This way the participants could provide all of their trip information for the 24-hr period.

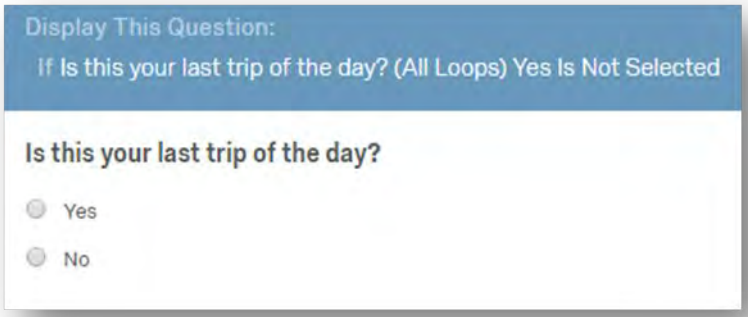


FIGURE 4-4. QUESTION FOR "LOOP AND MERGE".

4.3.2.6 Block 6

This block asked survey participants if they wished to see expansion of services related to Public Transit (bus, light rail), TNCs, Bicycle Lane, Sidewalks, Parking Lots in their area.

4.3.3 Demographic Information:

This block focused on the collection of demographic information of the participants including gender, age, employment status, occupation, industry, annual household income, highest degree, and auto ownership. Most of the questions were created using the "Multiple Choice" and "Drop Down List" tools from Qualtrics. The selection options were modeled according to the US Census categories. This block also asked the home location or the nearest intersection to allow for validation of the Home Zip Code provided at the second block of the survey.

4.3.4 Selection of participants: Location and Standard Sample

Our test bed was the metropolitan area of the greater Birmingham, AL. The area comprises of the cities Birmingham, Homewood, Vestavia Hills, Mountain Brook and Hoover. The area is populated by 1,141,309 capita as per the 2016 Census data. Participants residing in the area that were 18 years or older were eligible to participate in the survey.

Given the population size of the Birmingham Metro Area as per the 2016 Census data, 420 survey responses would be required to provide a representative sample of travel behavior. UAB contracted Qualtrics LLC to recruit and

administer the survey on the Birmingham, AL metropolitan area through Qualtrics Research Core.

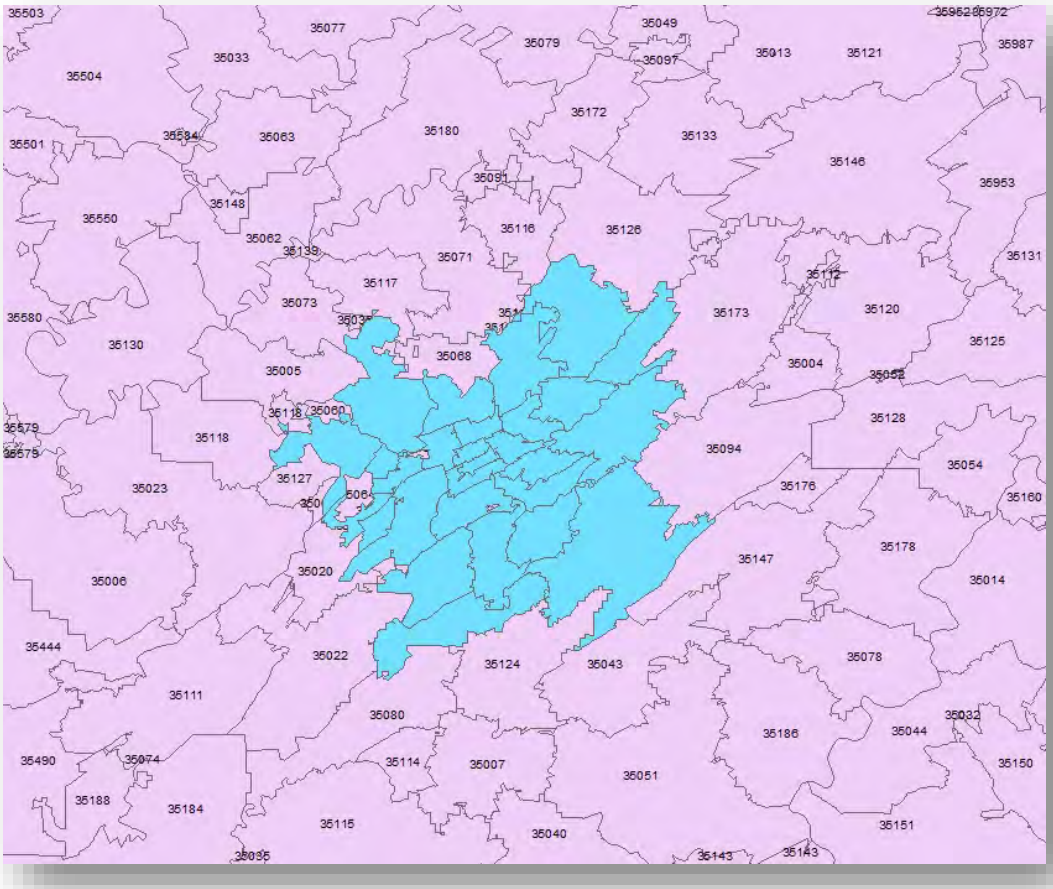


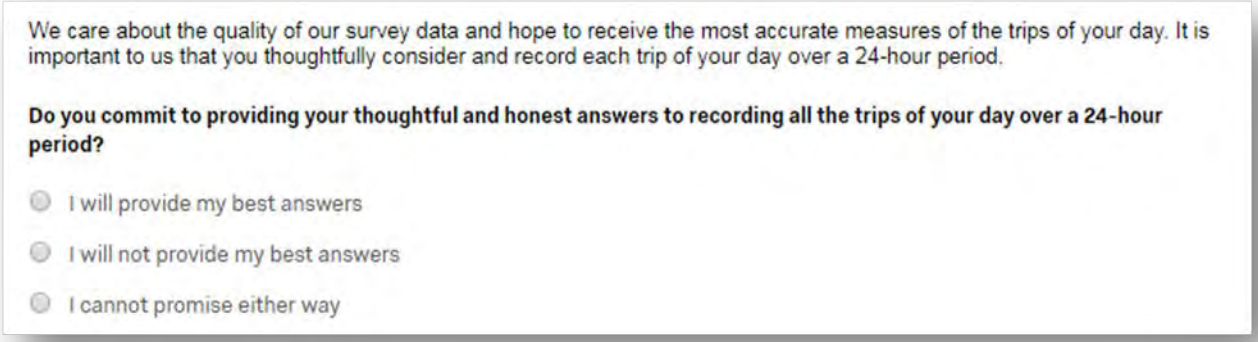
FIGURE 4-5. QUESTIONNAIRE SURVEY SAMPLE GEOGRAPHIC BOUNDARIES.

4.3.5 Survey pretesting and validation process

Before launching the Birmingham questionnaire survey, we went through a thorough pretesting and validation process to ensure that the survey questions were easy to understand, the answers were properly recorded, and the survey duration was reasonable (below 10 minutes).

We first *pre-tested* the draft questionnaire internally and received feedback from 15 UAB respondents which help us update some questions to improve clarity and flow. After the questionnaire survey tool was refined and the study protocol was approved by the IRBs at University of Alabama at Birmingham, we ran a *soft launch* and collected 10% responses to ensure that the instrument is functioning properly. Review of survey responses from the soft launch guided us towards adding one more block to the survey prompting participants to input details

about their every trip of a typical day as shown in Figure 4-6. A copy of the survey instrument used in this study is available in Appendix 8.2.


We care about the quality of our survey data and hope to receive the most accurate measures of the trips of your day. It is important to us that you thoughtfully consider and record each trip of your day over a 24-hour period.

Do you commit to providing your thoughtful and honest answers to recording all the trips of your day over a 24-hour period?

- I will provide my best answers
- I will not provide my best answers
- I cannot promise either way

FIGURE 4-6. NEW BLOCK TO COLLECT THE PROPER TRIP RECORD.

Following this update, Qualtrics proceeded with a full launch of the survey leading to the collection of 473 responses that were made available to the UAB team in January 2019. Among the 473 surveys received, 25 surveys were found to be from participants outside our survey study area and had to be discarded. In fact, these participants provided a residence ZIP Code which was included in Birmingham Metro Area, however, their home address did not match the ZIP Code information. This was verified using ArcGIS and led us to the decision not to keep those responses as we could not trust the accuracy of information provided.

After discarding the 25 surveys our database included 448 surveys. Among the 448 respondents, 104 respondents have given inconsistent home addresses in the trip and home location question. This was an important piece of information for the purposes of our study and a quality check metric thus omitting these surveys was necessary, further reducing our dataset to 344 usable responses. To meet our target sample size, we requested Qualtrics LLC to reopen the survey and recruit additional participants. Identifying additional subjects to part take in the survey proved to be a time consuming and tedious process, but the effort was successful nevertheless. After two iterations, Qualtrics LLC provided an additional 121 responses. Once again, we went through necessary quality control checks to ensure validity of responses and discard those that included inconsistencies and errors.

The process resulted in 451 complete questionnaire responses that we accepted as our final data set. This data set was used for the data analysis as documented in the following sections.

4.4 Data Analysis and Results

4.4.1 Demographic Data:

Among the 451 respondents considered in the analysis, 342 were women and the remaining were men. The overrepresentation of women in the survey was noted but is not alarming as many surveys in the literature reported higher numbers of survey participants as being female. The respondents provided details for 1,085 trips performed over a 24-hr period. Analysis of the data showed that approximately 6.37% of the reported trips were conducted by TNCs with 73% of TNC trips performed by female respondents. Taking exposure into consideration, the finding indicates that TNCs are used almost at the same rate among female and male transportation users in the Birmingham region.

Figure 4-7 displays the distribution of survey participants by age group. The survey participants represented age groups across the lifespan with a peak (25%) at between 25 to 34 years of age. The age distribution of the Birmingham survey sample is relatable to the actual scenario of Birmingham Metro Area, based on analysis of Census records.

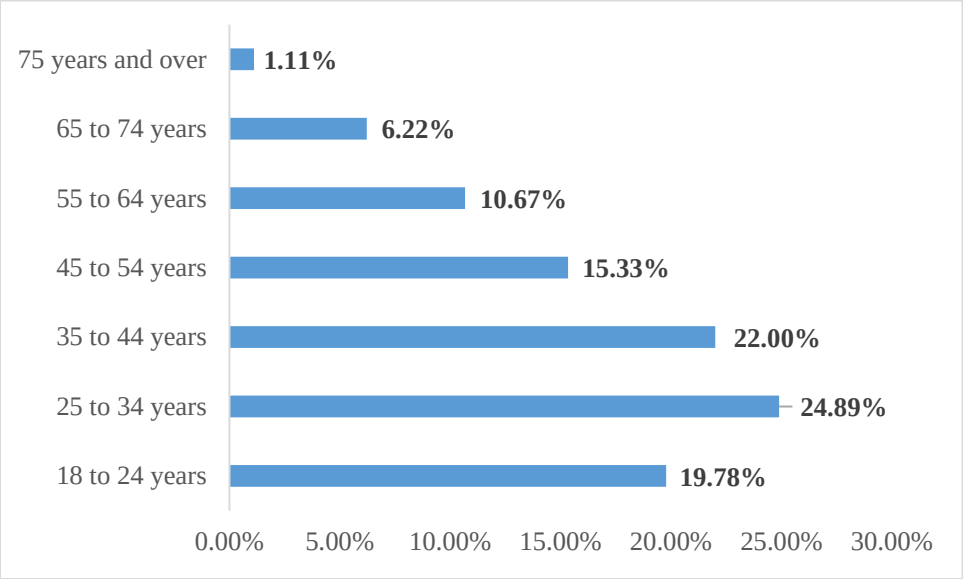


FIGURE 4-7. AGE DISTRIBUTION OF SURVEY PARTICIPANTS.

Figure 4-8 displays information related to survey participants' employment status and occupation. It can be seen that over 55% among the survey participants are full time employee and the remaining 44% contains part-time employees, retired persons, self-employed people, stay-at-home parents, students, unemployed and others.

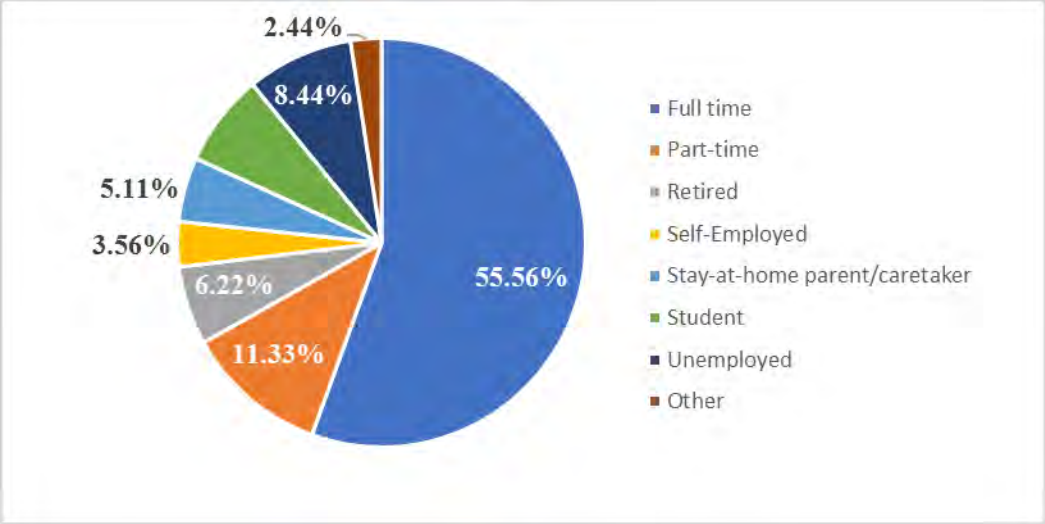


FIGURE 4-8. EMPLOYMENT STATUS OF SURVEY PARTICIPANTS.

Additional details regarding survey participants' occupation and the type of industry they are involved in are displayed in Figure 4-9 and Figure 4-10 respectively.

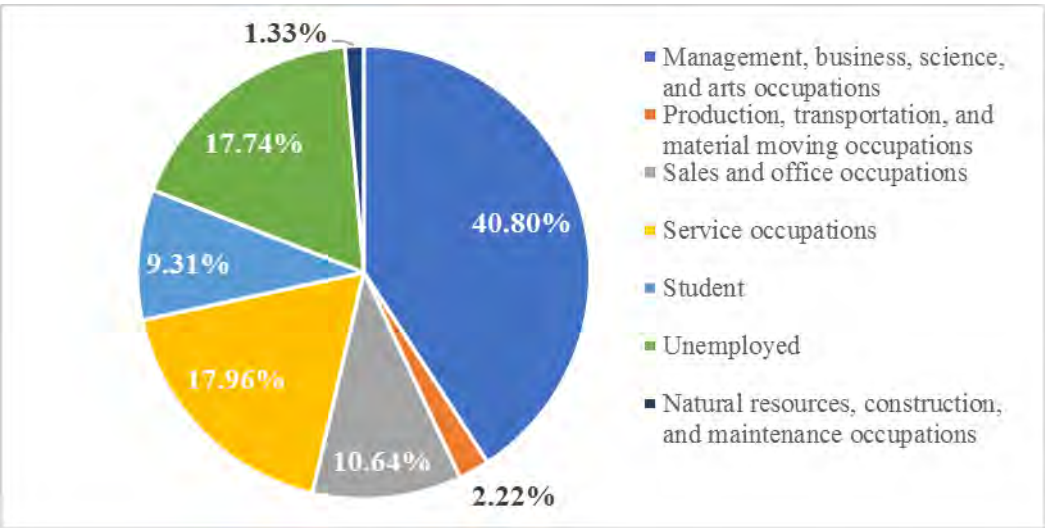


FIGURE 4-9. OCCUPATION OF SURVEY PARTICIPANTS.

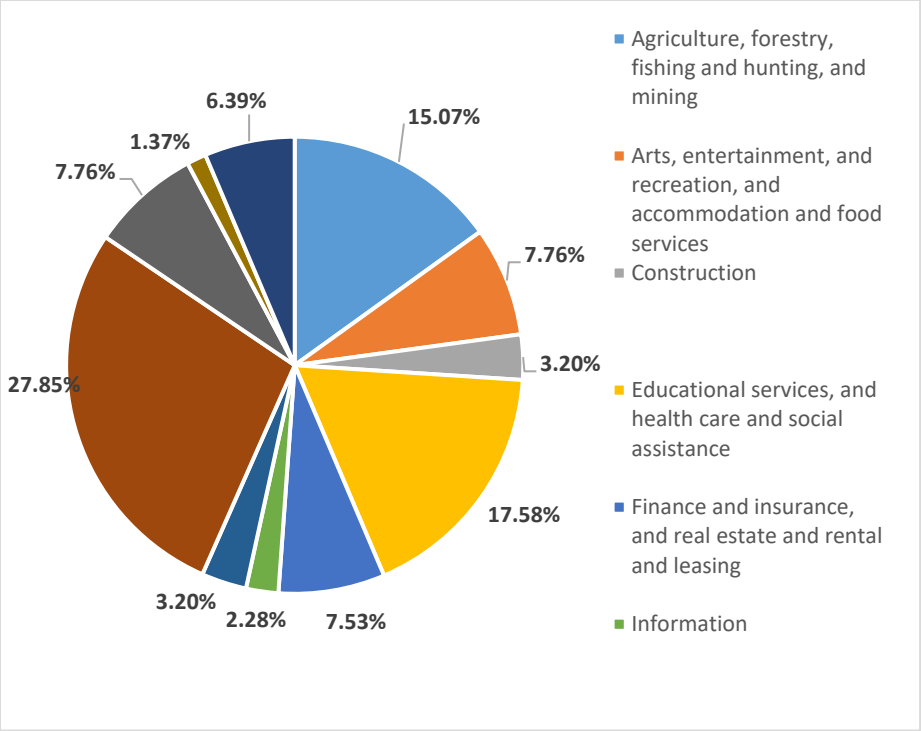


FIGURE 4-10. WORK INDUSTRY OF BIRMINGHAM SURVEY PARTICIPANTS.

Figure 4-11 shows the educational level of the survey participants. Based on the self-reported data of respondents, around 43% of them have a high school degree, 24.6% have bachelor degree, and 15.3% have a master's degree. These are higher than state averages given the proximity of the study area to the University of Alabama at Birmingham which is also the largest employer in the state of Alabama.

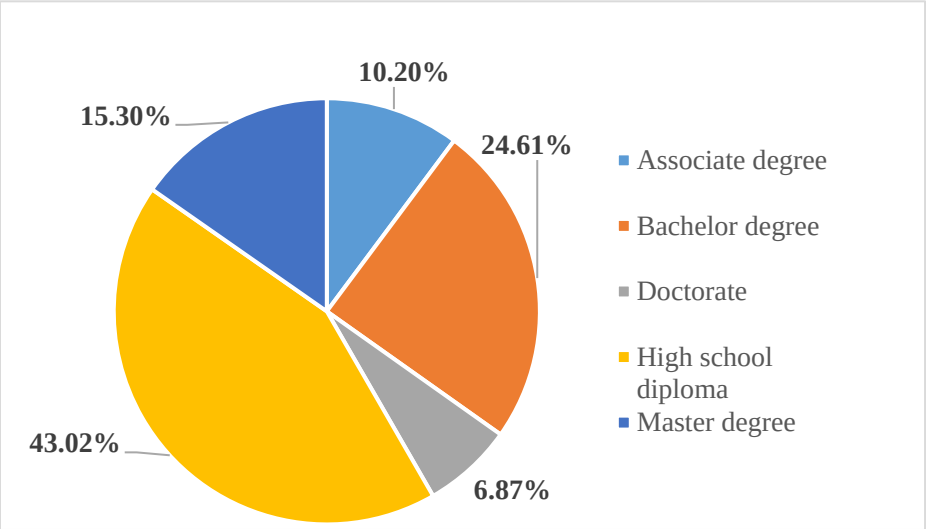


FIGURE 4-11. EDUCATION LEVEL OF BIRMINGHAM SURVEY PARTICIPANTS.

4.4.2 Preferences towards TNCs

To understand the survey participants' mode choices and their exposure to modes of transportation other than automobile we asked them whether or not they have used TNCs, public transit, bicycle, ride sharing program, etc. in the past year. As shown in Figure 4-12, approximately 45% of survey participants indicated that they have used TNCs in the past year. This is an important finding given that only 21% reported use of public transit during the same period and 12.6% of bicycle.

To understand the frequency of TNC use, the respondents were asked when was the last time that they used TNCs in the Birmingham region. Analysis of survey responses revealed that 50% of the TNC users used TNCs within the past month and half of those (about 24.3%) used TNCs at least once within the 7 days preceding the survey.

Additional analysis was performed to determine the potential impact of age on TNCs selection. Table 4-1 provides a cross-tabulation of survey results indicating the frequency of TNC use by age bracket. It can be observed that 25 to 34 year old survey participants use the TNCs the most (about 27.14%) followed by 18 to 25 year old respondents (19.5%). It can be also observed that use of TNC drops steadily as age increases when considering middle aged and elderly users.

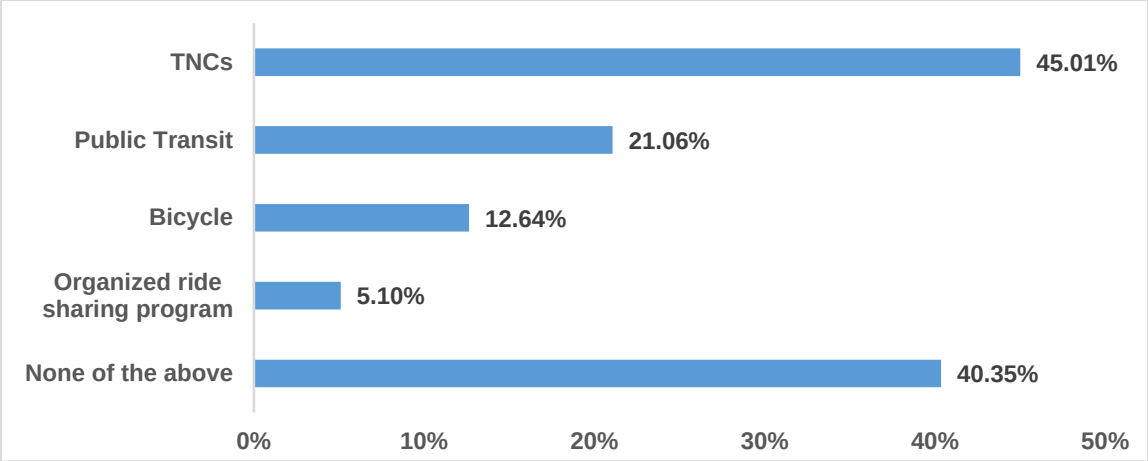


FIGURE 4-12. MODES USED IN THE PAST YEAR BY THE SURVEY PARTICIPANTS.

TABLE 4-1 FREQUENCY OF TNC USE BY AGE LEVEL OF BIRMINGHAM SURVEY PARTICIPANTS.

Age versus TNC Use Frequency	Within the past 7 days	Within the past 30 days	Within the past two months	Within the past 6 months	Within the past year	Total
18 to 24 years	6.67%	3.81%	0.95%	6.19%	1.90%	**19.52%**
25 to 34 years	4.76%	8.57%	5.71%	5.24%	2.86%	**27.14%**
35 to 44 years	3.33%	3.81%	1.43%	5.24%	3.81%	**17.62%**
45 to 54 years	5.24%	4.29%	1.43%	1.43%	3.81%	**16.19%**
55 to 64 years	1.90%	4.29%	0.00%	2.86%	1.43%	**10.48%**
65 to 74 years	2.38%	2.38%	0.48%	1.43%	1.43%	**8.10%**
75 years and over	0.00%	0.00%	0.00%	0.48%	0.00%	**0.48%**

Respondents were also asked to note the reason(s) for using TNCs in the past. In order to identify the most influential reasons for selecting TNC services as a mode of transportation, we factorized each reason to take a value of 1 if selected, and 0 if not selected. Table 4-2 documents the mean and standard deviation according to the survey responses.

The results clearly show that convenience was reported as the main driving force for the selection of TNCs as a mode of transportation by transportation users in the Birmingham area. Safety/avoiding driving when intoxicated and lack of automobile availability were cited as the second and third most important reasons for use of TNCs in the survey of Birmingham users.

TABLE 4-2. MEAN (STANDARD DEVIATION) OF THE FACTORS AFFECTING THE TNC PREFERENCE.

Reasons	Within the past 7 days	Within the past 30 days	Within the past two months	Within the past 6 months	Within the past year	Total
Convenience	0.13 (0.33)	0.15 (0.36)	0.07 (0.25)	0.14 (0.35)	0.07 (0.26)	0.56 (0.5)
Safety; avoid driving under the influence	0.07 (0.26)	0.09 (0.29)	0.03 (0.17)	0.06 (0.23)	0.05 (0.22)	0.30 (0.46)
Car is not available	0.07 (0.26)	0.06 (0.24)	0.05 (0.22)	0.06 (0.24)	0.02 (0.15)	0.27 (0.44)
Destination has limited/no parking availability	0.06 (0.23)	0.08 (0.27)	0.02 (0.14)	0.06 (0.23)	0.03 (0.18)	0.24 (0.43)
Cheaper than alternatives	0.05 (0.22)	0.05 (0.22)	0.01 (0.12)	0.07 (0.25)	0.02 (0.15)	0.21 (0.41)
Parking at destination is expensive	0.06 (0.24)	0.05 (0.21)	0.01 (0.12)	0.05 (0.21)	0.01 (0.12)	0.19 (0.39)
Transit is not accessible	0.02 (0.14)	0.01 (0.12)	0.01 (0.10)	0.01 (0.10)	0.00 (0.07)	0.06 (0.23)
Transit is not reliable	0.02 (0.14)	0.01(0.10)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.03 (0.17)
Other reason	0.01 (0.10)	0.01 (0.10)	0.00 (0.07)	0.00 (0.07)	0.00 (0.07)	0.03 (0.18)
Other modes not available	0.00 (0.00)	0.01 (0.10)	0.00 (0.07)	0.00 (0.07)	0.00 (0.07)	0.02 (0.15)

Survey respondents who had chosen the “other reason” option from the list had stated the following reasons for their preference towards TNCs:

- To the airport and back
- Restaurant with low (number of) parking spaces
- Car was towed to dealership for repair
- Tourists in another city
- Transportation from hospital
- Sightseeing tour
- Was involved in a car accident and needed a ride to home
- Going to work
- To get to (AMTRAK) train station
- Dropped car at auto shop and took Uber for home
- Rental car pickup

The survey also asked respondents who had not used TNCs within the past year to mark the reason for not considering TNCs as a mode of transportation. As shown in Figure 4-13, nearly 30% survey respondents reported that the use of TNCs was not convenient for them while another 20% noted that they do not use TNCs due to associated cost. Other reasons for not using TNCs cited by the respondents include the following:

- I have my own personal vehicle; I have a car on my household
- I drive or ride with friends
- I prefer public transportation because it brings diverse people together. If I desperately needed transportation, I would use a taxicab.
- It's quicker to hop in my own car and drive
- Live where I work so there is not a need for it
- I work too far from home
- Not interested
- It feels weird to ride in another person’s car
- Have not had the opportunity
- I am a wheelchair user

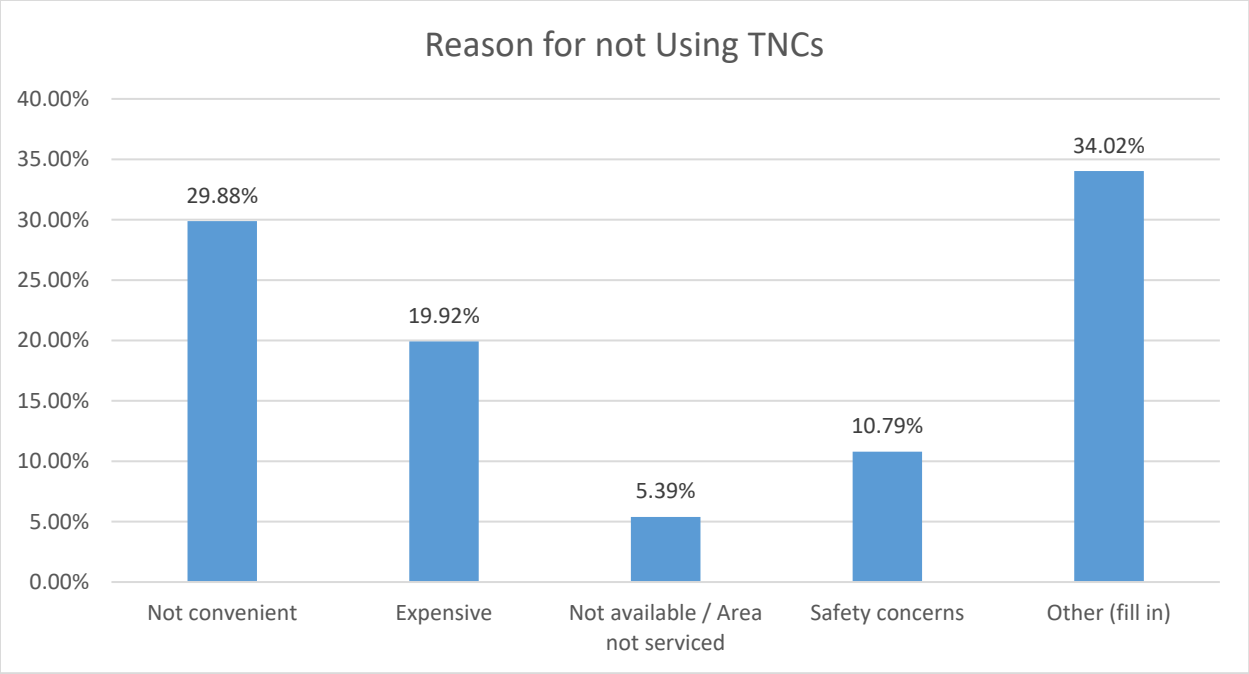


FIGURE 4-13. REASONS FOR NOT USING TNCS.

4.4.3 Trip Mode

The documentation of trips undertaken during a typical day 24-hour travel diary by the 451 Birmingham questionnaire survey respondents provide trip details for 1,130 trips. As shown in Figure 4-14, over 85% of these trips were conducted by private automobile and 6.3% by TNCs (Uber and Lyft). The data are consistent with earlier large-scale surveys in the Birmingham region by Sisiopiku (Sisiopiku, 2018; Sisiopiku and Ramadan, 2017), which reported that over 88% of UAB employees and 82% of UAB students commute to UAB by private automobile.

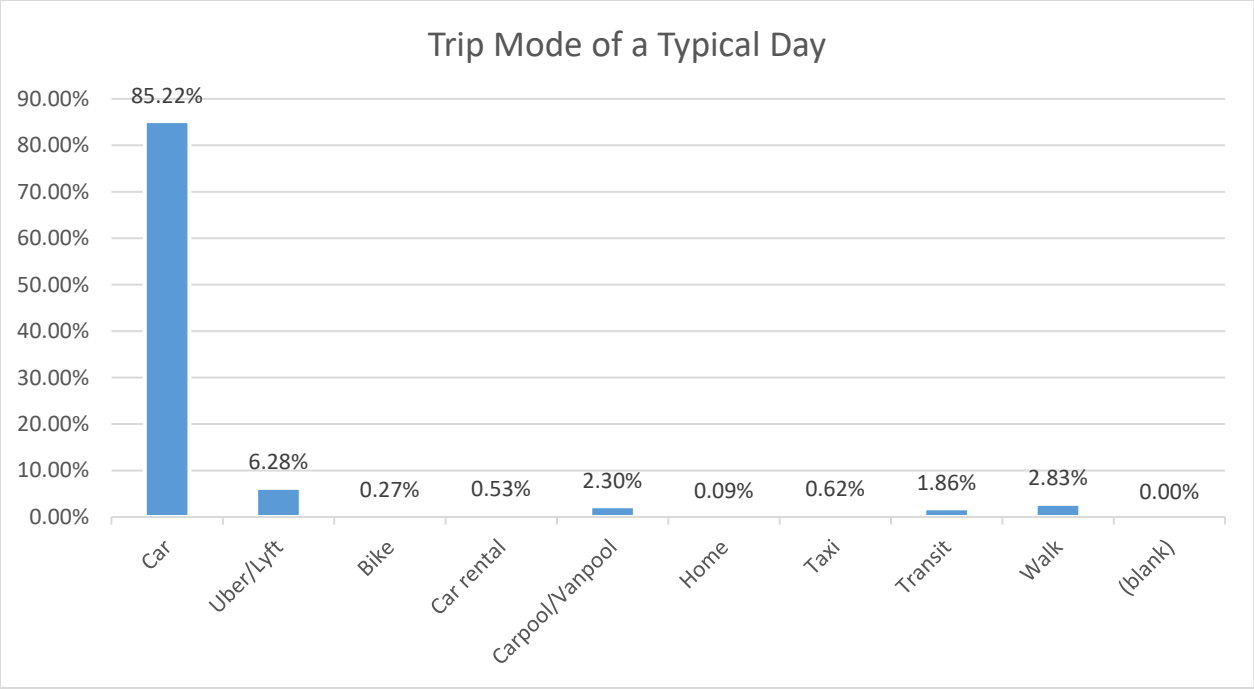


FIGURE 4-14. MODE CHOICES OF BIRMINGHAM SURVEY RESPONDENTS.

Cross tabulation of trip purpose to trip mode in Table 4-3 shows that the majority of the trips conducted by TNCs are trips to work or to home. This is consistent across the other available modes, including the automobile. The results also indicate that respondents use TNC services often for dining out and late-night entertainment. In fact, when accounting for the number of trips performed by each mode, the percentage of trips made for eating out/take out/nightlife with TNCs was found to be 22%, far greater than the same reported for automobile trips (11% of total). This shows a stronger preference for use of TNCs over automobile for dining out and entertainment trips among the Birmingham survey respondents. This is consistent with findings in the literature suggesting that the greatest levels of TNC use are on Friday and Saturday evenings and the busiest time in most cities is between 7 pm and midnight (Feigon and Murphy, 2018).

TABLE 4-3. TRIP PURPOSE VS TRIP MODE OF BIRMINGHAM SURVEY RESPONDENTS.

Trip Purpose to Trip Mode	Car	TNCs	Carpool/ Vanpool	Car Rental	Taxi	Transit	Bike	Walk
Home	27.2%	1.6%	0.5%	0.3%	0.4%	0.5%	0.1%	0.4%
Work	18.4%	2.4%	0.4%	0.0%	0.2%	0.7%	0.1%	0.4%
School	2.8%	0.3%	0.1%	0.0%	0.0%	0.4%	0.1%	0.6%
Eat/Take-out	6.7%	0.5%	0.4%	0.1%	0.0%	0.0%	0.1%	0.4%
Nightlife/ Bar	1.2%	0.4%	0.2%	0.2%	0.0%	0.0%	0.0%	0.0%
Shopping-Grocery	8.8%	0.3%	0.2%	0.0%	0.0%	0.1%	0.0%	0.8%
Shopping-Retail	6.9%	0.3%	0.0%	0.0%	0.0%	0.2%	0.0%	0.1%
Services (e.g. bank, post office)	7.1%	0.2%	0.4%	0.0%	0.1%	0.3%	0.0%	0.2%
Pick-up passenger	3.4%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Drop-off passenger	2.8%	0.4%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Total	85.3%	6.3%	2.3%	0.5%	0.6%	1.9%	0.3%	2.8%

Figure 4-15 represents the distribution of TNC trips by purpose and available TNC option. Two TNC companies operate in Birmingham, namely Uber and Lyft. TNC users in the Birmingham Metro Area reported using Uber for more than 80% of TNC trips. This is expected, given that Uber services have been available for longer time in the Birmingham region than Lyft. The finding is also consistent with national data reporting that Uber has largely dominated the market since its 2009 inception, accounting for over 80% of the market share, though recently this proportion has dropped below 75% (Cortright, 2017).

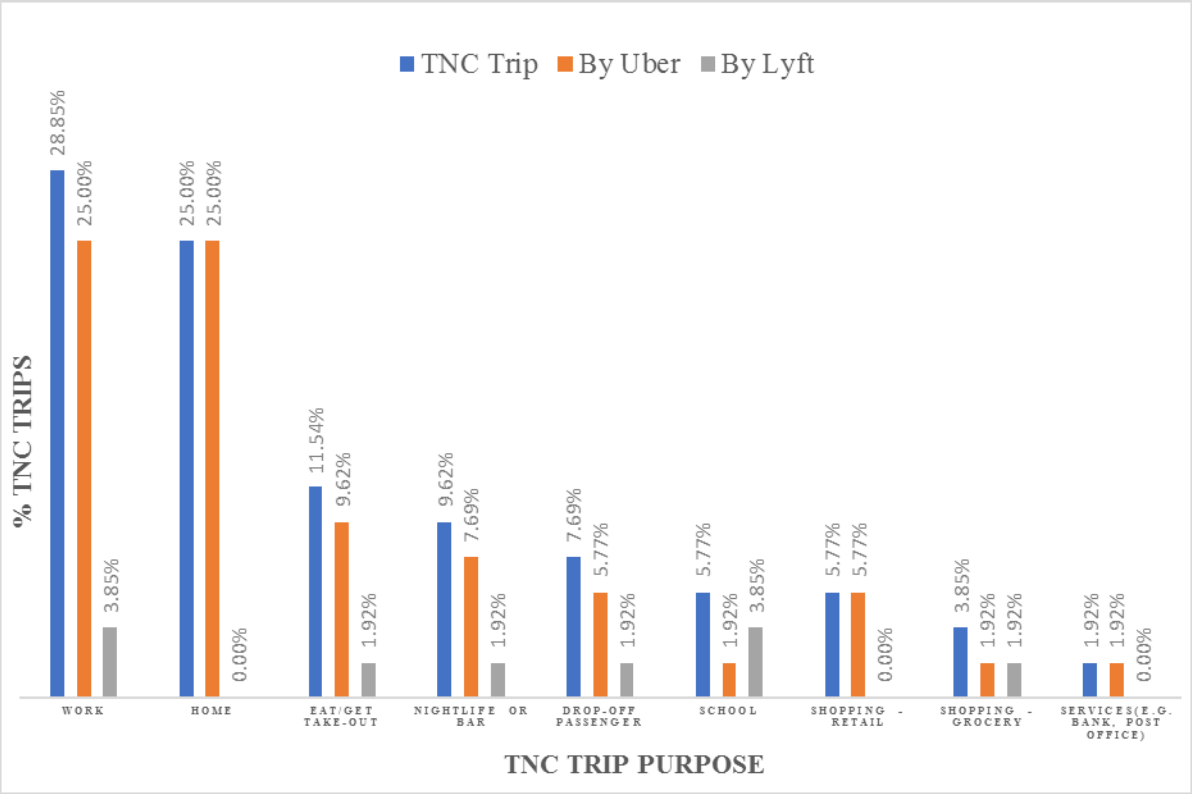


FIGURE 4-15. TRIP PURPOSE OF THE TNC TRIPS AND THE PREFERRED COMPANY.

As mentioned earlier, TNC availability in a transportation market is believed to have a potential impact on the public transportation use. Studies of six individual locations across North American found that between 13.5 percent and 54 percent of carsharing participants take public transit more frequently. However, one study of approximately 9,500 participants across North America found a slight shift away from public transit ridership (Martin & Shaheen, 2010). To understand the connection between auto ownership, transit users and TNC use, we classified the TNC trips reported in the Birmingham survey by vehicle ownership/availability and trip purpose. The results are depicted in Figure 4-16.

It can be observed that 52% of TNC users that completed the Birmingham survey own a vehicle and still use TNCs for select trip purposes. Another 25% reported that they do not own an automobile, but one is available in their household, while the remaining 23% of TNC users reported no vehicle ownership or access. These majorities of this 23% TNC users were public transit users before the introduction of TNCs in Birmingham Metro Area. Thus, distinct types of mode users including the private car users and public transit users are adapting and preferring TNCs as their trip mode.

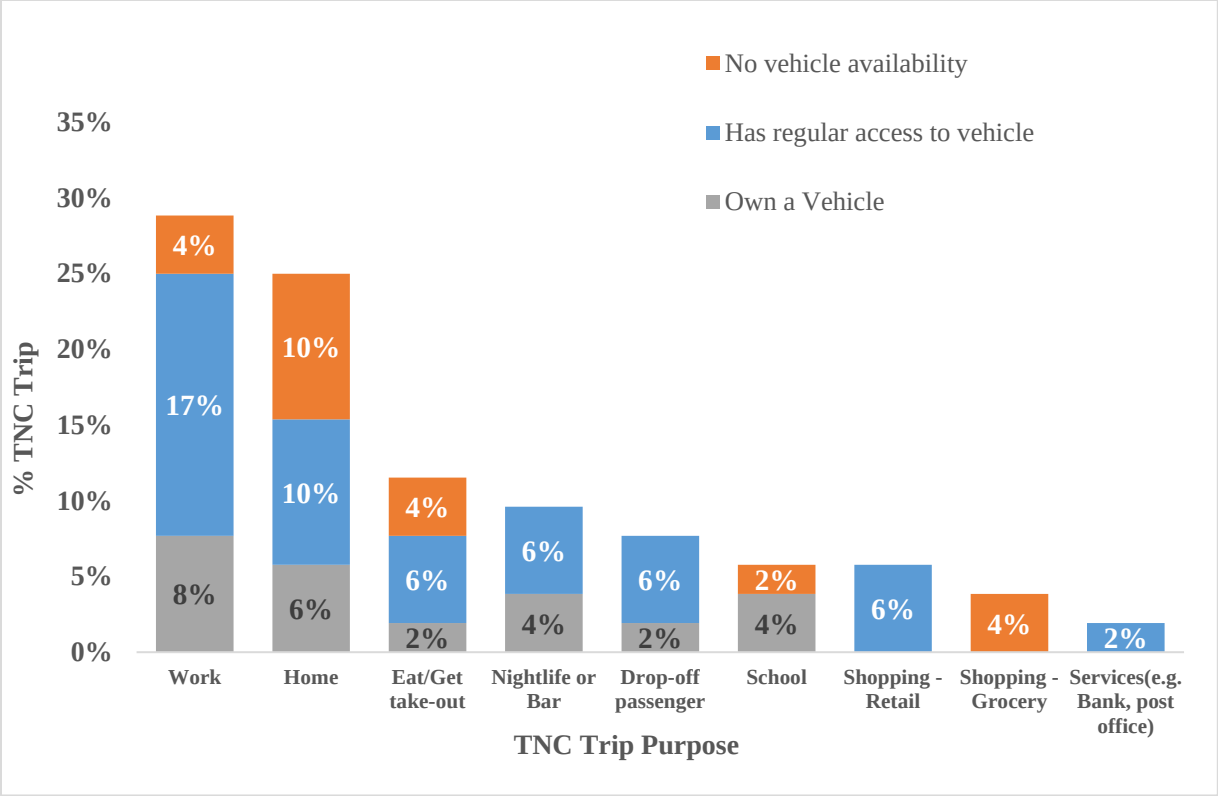


FIGURE 4-16. CAR AVAILABILITY OF TNC USERS.

A factor that was considered as a potential determinant of TNC use was the trip distance. According to the characteristics of the TNC trips reported in our study, TNC users use TNC services for trips under 10 miles. A comparison between TNC and non-TNC trips revealed that the average trip length performed by TNC was 5.19 miles, far lower than the average trip length of automobile trips (9.28 miles) in the region. Further analysis indicated that the longest TNC trips involved drop-off of a passenger or trips to work or home. The average trip length per trip purpose for TNC trips is shown in Figure 4-17.

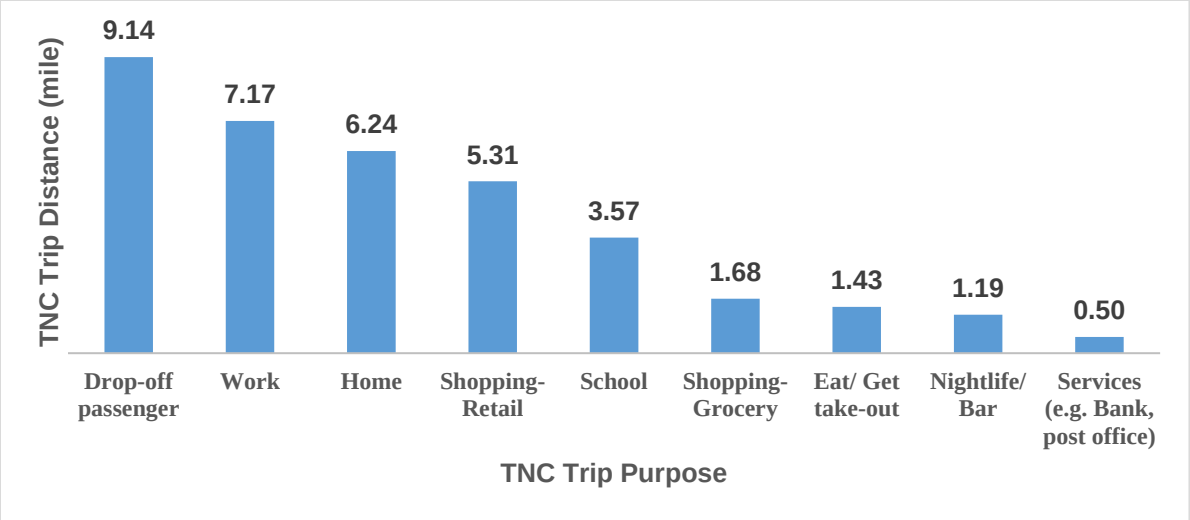


FIGURE 4-17. TNC TRIP DISTANCE (MILES) FOR VARIOUS TRIP PURPOSES.

Similarly, analysis of travel times reported in the survey of the Birmingham transportation system users indicated that the average trip by TNC was 25.64 minutes long, whereas automobile trips averaged 28.38 minutes. This is consistent with findings of the 2013 American Community Survey commuter data for the Birmingham-Hoover Metro Area, which reported average commute in the Birmingham metro of 25.7 minutes.

The findings of the survey also helped us to define the profile of the typical TNC user in the Birmingham region as a 25 to 34 year old that is using the service for commuting trips or for entertainment purposes for short to medium range distances (or average of 5 miles).

4.4.4 Personal Preferences

When Birmingham survey participants were asked about their preferences with respect to future improvements related to transportation infrastructure and services, 26% recommended an expansion of TNC services in the Birmingham region (Figure 4-18).

The current use of TNCs in the Birmingham Metropolitan area, coupled with transportation users' expressed interest in expansion of TNC services, highlight the importance of understanding the potential impacts of such services on traffic operations and traffic congestion in the region.

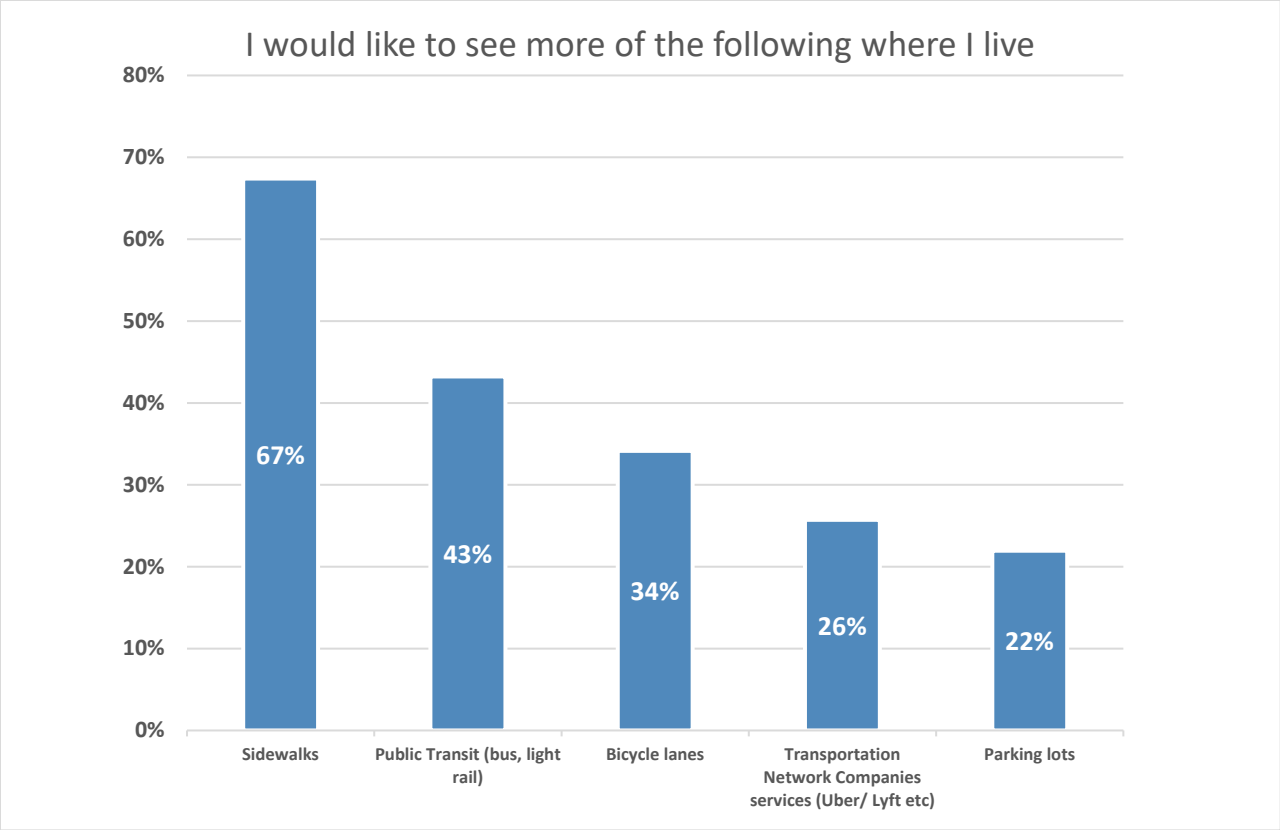


FIGURE 4-18. BIRMINGHAM SURVEY RESPONDENTS' PREFERENCES FOR FUTURE IMPROVEMENT OF TRANSPORTATION INFRASTRUCTURE AND SERVICES.

4.5 Conclusion and Future Work

The analysis of 451 questionnaire surveys of Birmingham transportation system users shed light on transportation system users' awareness and use of TNC services in the Birmingham metropolitan region. Examination of over 1,100 reported trips indicated that approximately 6.3% of those trips were performed using TNCs, with Uber having 80% of the TNC market share. Determinants that made TNCs a preferable mode to travelers include convenience of use, and reduction of concerns for traffic safety (especially for late night trips to bars and eating establishments). Lack of parking availability at destination was also listed as a reason for selecting TNCs as a mode of travel.

Examination of respondents' demographics and cross tabulation analyses provided evidence that TNC users cover a wide range of age groups, with younger users being overrepresented compared to elderly. Lack of vehicle availability was associated with only a quarter of all reported TNCs, thus indicating that the majority of TNC users select TNC services as a mode of choice for certain trips.

The analysis also confirmed that the market share of TNC trips is small (6.3% of trips reported). This is consistent with expectations, given that Uber and Lyft were recently introduced in the region and that transportation users in the Birmingham Metro largely embrace the automobile-dependent commuting culture as confirmed by previous studies. Still, 45% of survey respondents reportedly have used TNC in the past year, an indication of awareness of TNC service availability. This population segment can be targeted with marketing plans and incentives to encourage mode switching to shared modes, including TNCs.

This questionnaire survey is the first of its kind attempt to document the preferences, attitudes, and choices of transportation users in the Birmingham area in the presence of TNC services. The study highlights links between TNC service availability and travel choices among adults in the region, where the auto-dependent built environment likely influences these links. This study is also significant for providing transportation agencies the means to better-plan mobility as a service (MaaS) where car/ridesharing platforms are active. Moreover, study findings can inform TNC- and other shared-mode services about the needs and opportunities of the local market and enable them to better understand how the travel behavior, mode-choice, and travel demand might affect the use of TNCs in the future.

The results reported in this chapter are also expected to help transit agencies, and TNC companies to coordinate their efforts towards achieving integrated system operations that could attract new customers and benefit both types of transportation services in the future.

Last but not least, the Birmingham questionnaire survey gathered detailed trip information for a typical day from a sample of 450+ Birmingham transportation network users. Information gathered includes origin and destination of each trip, travel time, trip purpose and travel mode used. Such information can provide valuable seed trip data for the development a synthetic population, as needed for modeling the Birmingham region using an activity-based simulation platform as described in Chapter 5 of this report.

5.0 SHARED MOBILITY SIMULATION MODELING: A FEASIBILITY STUDY OF THE BIRMINGHAM REGION

5.1 Introduction

Shared mobility and ridesourcing services (such as Uber and Lyft) are emerging transportation strategies that enable transportation users to gain short-term access to transportation modes on an as-needed basis. In the recent years, these transportation options have been adopted at a rapid pace, due to technological advancements and the public's willingness to accept and support sharing economy. However, the full impact of shared mobility and ridesourcing services on local and regional congestion are not yet understood, due to field data acquisition challenges. Thus, a need exists to evaluate and quantify the impacts of shared mobility on the performance of urban transportation facilities and simulation modeling can be utilized for this purpose to the lack of field data availability.

Traditional traffic simulation models lack the ability to simulate shared modes in detail, however, in the recent years, new simulation platforms have emerged to allow shared mobility simulations. Given the limited experience in this area, an extensive literature review was performed, and research case studies were examined to identify available platforms for shared mobility simulation modeling. Four simulation platforms were identified, namely Multi-Agent Transport Simulation (MATSim- Version 0.8.1), Auto Desk Mobility Simulator, the Dynamic Ridesharing (D-Ride- Version 1.0), and ICON Carpooling Demo software. These simulation platforms implement agent/activity-based modeling, data mining and machine learning, and have various advantages and shortcomings for implementation.

Attributes of each simulation platform were reviewed and summarized in Table 5-1. Attention was paid on the types of modes that can be simulated by each tool, system requirements, model development requirements, user friendliness, modeling fidelity, ability to model dynamic events, and cost. Moreover, input requirement and output capabilities were review and contrasted as shown in Table 5-2.

The comparison of simulation model capabilities performed in this study showed that the most promising and well-established platform for simulating ridesharing travel options is MATSim (MATSim). It incorporates time choice, mode choice, and/or destination choice into an iterative loop, leading to a stochastic user equilibrium. Through its computationally efficient-queue based approach, MATSim holds promise toward accurate modelling of technology-based ridesharing modes. Thus, the MATSim model was selected as the best available tool for meeting the needs of the feasibility study.

TABLE 5.0-1 COMPARISON OF ATTRIBUTES OF SHARED MOBILITY MODELING SIMULATION TOOLS

Attributes	MATSIm	Mobility Simulator for InfraWorks 360	D-RIDE-AMS	ICON Carpooling
Simulated modes	Car, bike, train, taxi, truck, car/ride share	Car, taxi, bicycle, walking, bus, train	Car-pooling, ride-sharing, vanpooling	Ridesharing
Simulation architecture	Multi-agent simulation	Multi-agent/ agent-based model	Activity based model	Data Mining/Machine Learning
Pre-requisite skills	Java Prog., XML structures, agent-based	OS, MS Office	OS, MS Office, GIS	PostGRESql, PostGIS, Python
System requirements	4 GB RAM and 200 GB free disk space	8 GB RAM,10 GB free disk space, Core i7	MS Windows 7, Visual Studio Libraries	MS Windows 7 or newer, 2 GB RAM
Model development time	Extensive	Low	Low	Moderate
User-friendliness	Basic GUI without online help	Fully developed GUI, well organized	Good GUI interface, easy to locate tools	Basic GUI, counter intuitive
Modeling fidelity	Mesoscopic: Medium/high fidelity	Microscopic; High fidelity	Macroscopic; Low fidelity	Macroscopic; Low fidelity
Dynamic events modeled	Weather conditions, incidents	No	No	No
Cost	Open-source + €1000 /yr. for Via	$1575 /yr.	Open-source	Open-source

TABLE 5.0-2 COMPARISON OF INPUTS AND OUTPUTS OF SHARED MOBILITY MODELING SIMULATION TOOLS

Attributes	MATSim	Mobility Simulator for InfraWorks 360	D-RIDE-AMS	ICON Carpooling
Input requirements	**Configuration**: Connects other input files, configuration parameters, controller, etc. **Network**: Nodes & links, coordinates, modes using link, link capacity, speed **Demand**: Travel demand and daily plans (tours) for every agent	**Parameters**: Defines agents' behaviors **Network**: Shows roadways and paths **Control**: Traffic signals, pedestrian crossings **Demand**: Trips, origin, destination **Trips**: List of trips, agent, origin, destination, and departure time **Validation**: Validate model performance	**Agent data**: Demand, origin, dest., depart. time, arr. time, capacity **Configuration data**: Iterations, shortest path, vehicle cost/hour **Link data**: Id, start/end node, type, direction, length, lanes, speed limit, capacity **Node data**: Node id, coordinates	GPS trajectories. The study is anchored to a large mobility dataset, consisting of the complete one-month-long GPS trajectories of approx. 10% circulating cars in Tuscany, Italy.
Outputs	• Score Statistics (.png): show the avg. best, worst, executed and overall avg. of all agents' plans for every iteration. • Leg Travel Distance Statistics (.png): plot travel distance • Events (XML): activity start or change, important base for post-analyses • Plans (XML): the current state of the population with their plans • Leg Histogram (.png): agents arriving, departing or en-route, per time unit • Trip Durations (.txt): listing number of trips and their durations • Link and Network Stats (.txt): count values, travel times, emissions Accessibility measures	• Summaries for People/ Cyclists/ Public Transport/ Private Vehicles/ Freight: • Distance (m), time (sec.), stops for each mode (number of stops) • Modes includes walking, passenger, driving, waiting • Lane changes • Loop activations • Emissions (CO2 (kg/ton), NO (g/ton), PM10 (g/ton)) • Detailed Public Transport Information • Economic Evaluation (detailed costs for each trip) Level of Service Reports	• AgentPlus: suggests each vehicle's pickup, delivery sequence, and corresponding paths to satisfy all passengers' needs while minimizing the overall cost. • DTALite: determines the best dynamic pricing strategy for vehicles, to have a sustainable development of D-RIDE applications. • Agent routing • Agent scheduling: a path containing a sequence of time stamps • Assignment of vehicles to passengers • Updated agent serving value ($) Upper bound, Lower bound, and the gap percentage between these two	

5.2 The MATSIM Platform

5.2.1 Introduction

MATSim is a non-traditional, open source simulation platform, implemented as a Java application, that provides a framework to implement large-scale agent-based transportation simulations of various transportation modes, including shared modes. The framework consists of several modules which can be combined or used stand-alone. Currently, MATSim offers a framework for demand-modeling, agent-based mobility-simulation (traffic flow simulation), re-planning, a controller to iteratively run simulations as well as methods to analyze the output generated by the modules.

The platform adopts the activity-based approach to generate agents' activities. Within the context of MATSim, agents are the individual travelers, and agent behavior refers to an individual's daily activity travel plan and route choice.

MATSim designs two layers: a) the physical layer, which simulates the physical world where the agent (or traveler) moves, and b) the mental layer, in which the agents generate strategies, including routes, mode choice, and daily activity plans. MATSim runs its activity plan, microsimulation, activity re-plan, microsimulation, and so on, iteratively until it reaches a stationary state of the system, where an agent cannot improve its score by revising the plan. The MATSim simulation steps are listed below:

- A set of initial plans is generated.
- The plan selection mechanism of the agent database chooses one plan per agent for execution.
- The model runs the simulation to execute the plans, produce a new travel time for each trip, and re-score the plans.
- A subset of the agents is chosen to undergo plan adjustment or new plan generation by external strategy modules.
- The model runs external strategy modules, and each agent is updated with a new or revised plan.
- The model runs the mode and route choice module to produce a route for each agent.
- If the stop criterion is satisfied, then the simulation stops; otherwise, the process continues for additional iterations, as needed.

5.2.2 MATSim Capabilities and Requirements

The MATSim tool is designed to simulate large-scale scenarios by adopting a computationally efficient queue-based approach (Horni, Nagel, & Axhausen, 2016). It incorporates mode/time choice, and/or destination choice into an iterative process loop by removing the lowest score plan until the average plans become steady.

A MATSim run contains a number of replications starting with an initial demand that emerged from the travel diaries for travelers in the study area. Activity chains are derived from empirical data through sampling or discrete choice modeling to establish the initial demand. The initial demand is optimized individually for each traveler during iterations. Each traveler selects a plan prior to simulation in each iteration, the selection is dependent on plan scores, which are calculated after each mobility simulation (mobsim) run based on plan performances. MATSim replanning module is performed to modify travel plans by considering four dimensions: departure time, route, mode, and destination (Horni et al. 2016). The MATSim loop is demonstrated in Figure 5-1 below.

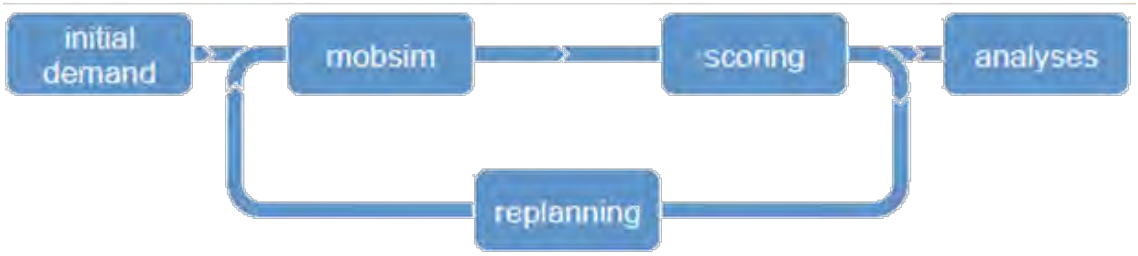


FIGURE 5.0-1 MATSIM LOOP (SOURCE: MATSIM BOOK 2016, P. 5).

In addition to external mobility modulus that can be plugged into MATSim, the traffic flow model in MATSim provides two internal Mobility Simulations (mobsims), namely Queue Simulation (QSim), and Java Discrete Event Queue Simulation (JDEQSim). Figure 5-2 shows the traffic flow model developed by MATSim.

The MATSim's default mobsim parameter in the configuration file is QSim; the queue and time-step based. The MATSim traffic flow model works based on two link attributes, namely the storage and flow capacity. Storage refers to the number of vehicles that can fit onto a network link so that vehicles can only enter a link when a link is not full. Flow capacity refers to outflow capacity of the section, e.g. number of vehicles that can leave the section per time (Horni et al. 2016).

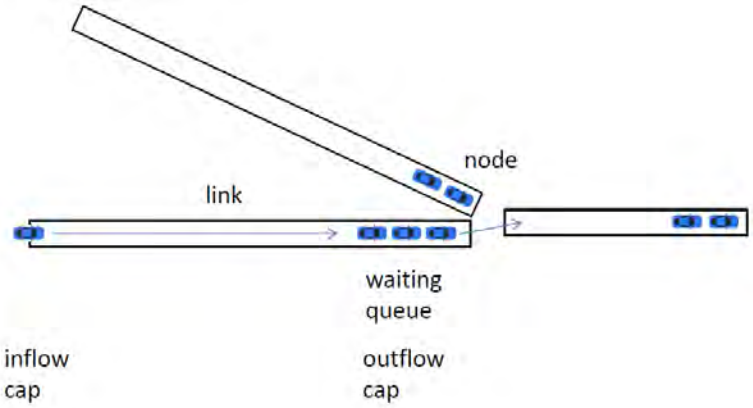


FIGURE 5.0-2. MATSIM TRAFFIC FLOW MODEL (SOURCE: MATSIM BOOK, 2016).

MATSim optimization is performed in terms of agents' plans scoring based on a co-evolutionary algorithm (which leads to a stochastic user equilibrium) until reaching an equilibrium (Horni, Nagel, & Axhausen, 2016).

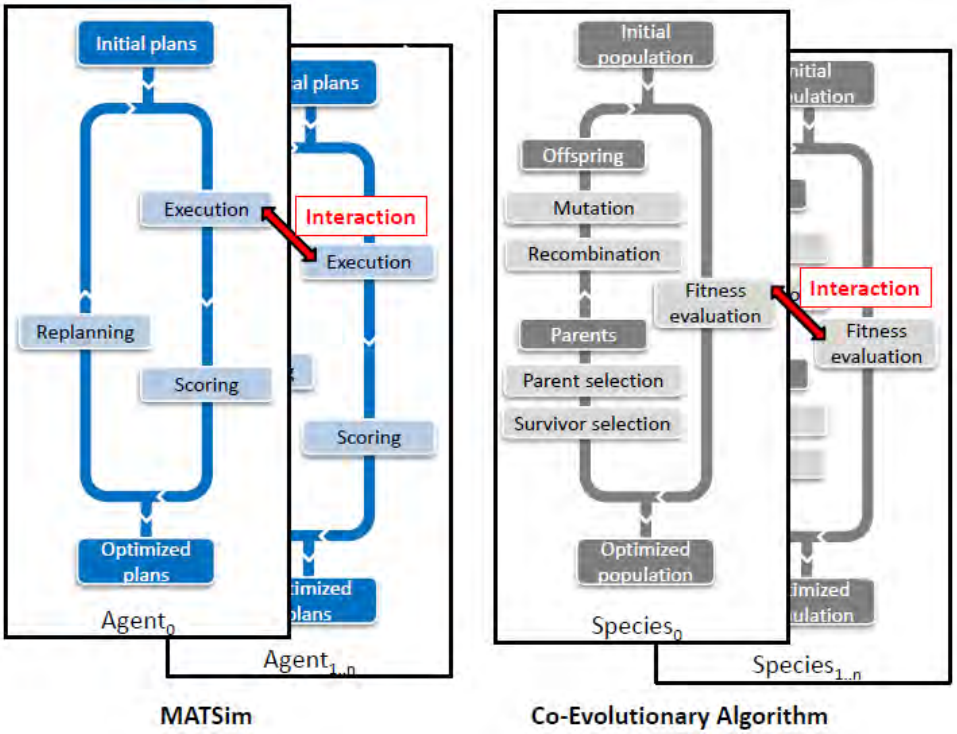


FIGURE 5.0-3. THE CO-EVOLUTIONARY ALGORITHM IN MATSIM (SOURCE: MATSIM BOOK, 2016).

5.2.3 Typical Model Input

MATSim is open-source software that requires its input files to be XML files. Minimum input files required to run the software are:

- Configuration file
- Network file
- Population/plans file

The subsections below explain the purpose and the required inputs for each file.

Configuration file (config.xml)

The configuration file contains a list of settings that influence how the simulation behaves (e.g. the number of iterations, and the end time of the mobsim, etc.). It builds the connection between MATSim tool and all other input files (e.g. network, population, etc.), Figure 5-4 shows a typical file structure for the MATSim configuration file.

```
<module name="network">
   <param name="inputNetworkFile" value="<path-to-network-file>" />
</module>

<module name="plans">
   <param name="inputPlansFile" value="<path-to-plans-file>" />
</module>

<module name="controler">
   <param name="firstIteration" value="0" />
   <param name="lastIteration" value="0" />
</module>

<module name="planCalcScore" >
   <parameterset type="activityParams" >
      <param name="activityType" value="h" />
      <param name="typicalDuration" value="12:00:00" />
   </parameterset>
   <parameterset type="activityParams" >
      <param name="activityType" value="w" />
      <param name="typicalDuration" value="08:00:00" />
   </parameterset>
</module>
```

FIGURE 5.0-4 TYPICAL FILE STRUCTURE FOR MATSIM CONFIGURATION FILE (SOURCE: MATSIM BOOK, 2016).

In the MATSim replanning strategy (i.e., module ReRoute), x% of the agents reroute one of their plans in each iteration. The remaining percentage of the agents select their highest score plan for re-execution in the same iteration (module BestScore). Plans will be deleted if the memory is full, defined by maxAgentPlanMemorySize. The plan with the lowest score is removed. Figure 5-5 below shows the reroute strategy settings of MATSim.

```
<module name="strategy">
   <param name="maxAgentPlanMemorySize" value="5" /> <!-- 0 means unlimited -->

   <parameterset type="strategysettings" >
      <param name="strategyName" value="ReRoute" />
      <param name="weight" value="0.1" />
   </parameterset>

   <parameterset type="strategysettings" >
      <param name="strategyName" value="BestScore" />
      <param name="weight" value="0.9" />
```

FIGURE 5.0-5. "STRATEGY" SECTION OF THE CONFIG FILE (SOURCE: MATSIM BOOK, 2016).

The routing algorithm is shown in Figure 5-6 below:

```
<param name="routingAlgorithmType" value="{Dijkstra | FastDijkstra |
   AStarLandmarks | FastAStarLandmarks}" />
```

FIGURE 5.0-6. THE ROUTING ALGORITHM (SOURCE: MATSIM BOOK, 2016).

The performance of the plan in the synthetic reality is taken to compute each executed plan's score as in Figure 5-7 below.

```
<module name="planCalcScore" >
   ...
</module>
```

FIGURE 5.0-7. SCORING. (SOURCE: MATSIM BOOK, 2016).

Network file (network.xml)

MATSim's network file consists of nodes and links representing the infrastructure on which agents move around. Nodes are defined by coordinates while links require the definition of several attributes including the length of the link, capacity, speed, and the number of lanes that modes used. Figure 5-8 below shows an example of the structure of a network file in MATSim. The network file creation is an important step in the model development process that requires the flowing steps:

- Download of Java Open Street Map "JOSM". It is a tool that can be employed to create the network file.
- Activation of the MATSim plugin in Preferences mode which allows to edit and extract network information.
- Use of JOSM to download map data by selecting the map boundary and then convert it to the MATSim layer.

- Validation of the map in JOSM, after which the map can be saved as a MATSim network file to be easily used as an input.

```
<network name="example network">
   <nodes>
      <node id="1" x="0.0" y="0.0"/>
      <node id="2" x="1000.0" y="0.0"/>
      <node id="3" x="1000.0" y="1000.0"/>
   </nodes>
   <links>
      <link id="1" from="1" to="2" length="3000.00" capacity="3600"
            freespeed="27.78" permlanes="2" modes="car" />
      <link id="2" from="2" to="3" length="4000.00" capacity="1800"
            freespeed="27.78" permlanes="1" modes="car" />
      <link id="3" from="3" to="2" length="4000.00" capacity="1800"
            freespeed="27.78" permlanes="1" modes="car" />
      <link id="4" from="3" to="1" length="6000.00" capacity="3600"
            freespeed="27.78" permlanes="2" modes="car" />
   </links>
</network>
```

FIGURE 5.0-8. NETWORK (SOURCE: MATSIM BOOK, 2016).

Population file (population.xml)

It provides information about travel demand, e.g. a list of agents and their travel diaries. The travel demand is described by the daily plans of each agent. The population file contains a list of transportation users and their daily plans, activities, and legs. Figure 5-9 below shows the typical population/plans file in MATSim.

```
<population>
   <person id="1">
      <plan selected="yes" score="93.2987721">
         <act type="home" link="1" end_time="07:16:23" />
         <leg mode="car">
            <route type="links">1 2 3</route>
         </leg>
         <act type="work" link="3" end_time="17:38:34" />
         <leg mode="car">
            <route type="links">3 1</route>
         </leg>
         <act type="home" link="1" />
      </plan>
   </person>
   <person id="2">
      <plan selected="yes" score="144.39002">
         ...
      </plan>
   </person>
</population>
```

FIGURE 5.0-9 TYPICAL POPULATION FILE STRUCTURE (SOURCE: MATSIM BOOK, 2016, P. 16).

It should be noted that because it is practically impossible to get detailed activity-based data for the whole population in any study area, a population synthesis is needed to create the population data based on a sample of data (e.g. travel dairy survey data) using modeling techniques that mirror the true population. Hence, modelers opt for population syntheses based on travel diary surveys, land use data, and census data. The most prominent techniques are iterative proportional fitting (IPF), iterative proportional updating (IPU), combinatorial optimization, Markov-based, fitness-based synthesis, and other emerging approaches. Yet, in the literature, there is no clear guideline on using any of the available techniques. To address this gap, as part of this study, a comprehensive review of population synthesis (Ramadan & Sisiopiku, 2019) options for activity- and agent-based models was performed by Ramadan and Sisiopiku (2019) and is available in Appendix 8.3.

The critical review of the literature on population synthesis by Ramadan and Sisiopiku (2019) concluded that, despite its identified limitations and drawbacks, the IPF approach is the most feasible and widely used population synthesizer, is being used in state-of-the-art simulation platforms like MATSim and is still being preferred by modelers and practitioners. Thus, brief summary of population synthesis efforts related to IPF is provided next.

5.2.3.1 Population Synthesis

Activity-based travel demand modeling demands comprehensive socio-economic and travel data for the population of the study area. As it is prohibitively expensive to harvest such data for a whole population, and in most cases practically impossible, population synthesis has been offered as an alternative that can synthesize the data on the basis of a sample (Choupani & Mamdoohi, 2016).

As disaggregate models and microsimulation techniques become increasingly popular, researchers undertaking modeling are learning to create synthetic populations, that are disaggregate representations of an area's population that mirrors the true population (present or predicted) in terms of a number of elements (either factual or predicted), for example income level or size of household (Abraham, Stefan, & Hunt, 2012).

Along these lines, a far more economical means of predicting the characteristics of the population is rooted in iterative proportional fitting (IPF), which has attracted the interest of many researchers as it has a number of clear benefits. In 2016, Choupani & Mamdoohi, undertook a critical review of the current progress with IPF through a comprehensive review of literature that summarized the benefits and difficulties related to the IPF technique. They also identified areas that would be fruitful for researchers to study in the future to make IPF more valuable in terms of synthesis. This study has demonstrated that two of the most central difficulties which require empirical investigation are: a) the

integer conversion, and b) the zero-cell. Thus, there is a need for the development of unbiased tabular (controlled) rounding methodologies in order to integerize fractional numbers regarding the frequency of household types from the IPF estimation.

To address the limitation of the traditional IPF technique, research conducted by Fournier, Christofa, Akkinepally, and Azevedo (2018) offers a method that combines a number of techniques when access to both aggregated and disaggregated data is possible. To demonstrate the concept, a synthetic population was generated for the Greater Boston Metropolitan Area, comprising of around 4.6 million individuals in 1.7 million households. This research demonstrates that the limitations of IPF can be resolved by combining five methods into an integral framework for population synthesis, namely Seeding algorithm; Iterative Proportional Fitting; Integerization; Iterative Proportional Updating; and Monte Carlo Sampling.

5.2.3.2 MATSim Typical Outputs

Typical output files from MATSim include the following:

- Log File which contains information needed later for analyses or debugging
- Warnings and Errors Log File which identifies problems in the simulation
- Score Statistics which shows the average best, worst, executed and overall average of all agents' plans for every iteration
- Leg Travel Distance Statistics
- Stopwatch which contains the computer time
- Events which record every action
- Plans which contain the final iteration plans
- Leg Histogram which describes the number of agents arriving, departing or enroute, per time unit
- Trip Durations, and
- Link Stats which contains hourly count values and travel times on every network link.

5.3 Birmingham Pilot Study

5.3.1 Introduction

The purpose of the Birmingham pilot study was to demonstrate the feasibility of building a model for the Birmingham area that can be later used as a testbed for

the evaluation of the impact of shared mobility options on traffic operations and congestion. The study site covered the metropolitan area of greater Birmingham, AL. This area comprises of the cities of Birmingham, Homewood, Vestavia Hills, Mountain Brook, and Hoover. The majority of the area is in Jefferson County and a portion of it within Shelby County. The area is populated by 1,141,309 capita as per the 2016 Census data.

Based on the findings from the literature review and comparison of available simulation platforms, the platform selected to be used for the model development in this study was the agent-based transportation simulation platform MATSim. The model required transportation network data, population travel activity data, carsharing stations data (i.e., most likely locations of vehicles hovering the network), and facility data (i.e., location of buildings and points of interest). As part of the model development processes, the Open Street Map of the study area was acquired and coded it into MATSim format. Details about the MATSim Birmingham model development process are discussed next.

5.3.2 Birmingham MATSim Model Development Requirements

As discussed earlier, a MATSim model requires a configuration file, a network file; and Population/plans file. The configuration file describes the settings that influence how the simulation behaves and builds the connection between the various input files (e.g. network, population, etc.). The network file consists of the transportation network nodes and links representing the infrastructure that agents use to move around. The population file contains information about the detailed daily travel plans of each agent that utilizes the network.

Specifically, the Birmingham area MATSim model was developed following sub-tasks described below.

1) Determine and Code Network. The study team determined the study network boundaries using GIS. As a result, a GIS map showing the study network was developed. Then, the study team generated the network code as per the requirements of the simulation platform. To build the Birmingham region network and due to the map size limitation in JOSM, the Mapzen website was used to download the full map for Birmingham, AL area. The Mapzen file then imported to JOSM as a MATSim scenario. A map projection in JOSM used the following:

    - Projection Method: Mercator
    - Display coordinates as: Decimal Degrees
    - System of measurement: metric

Figure 5-10 shows an image of the Birmingham, AL network using JOSM and after being converted as a MATSim layer. Grey color represents links while the light blue color represents the nodes.

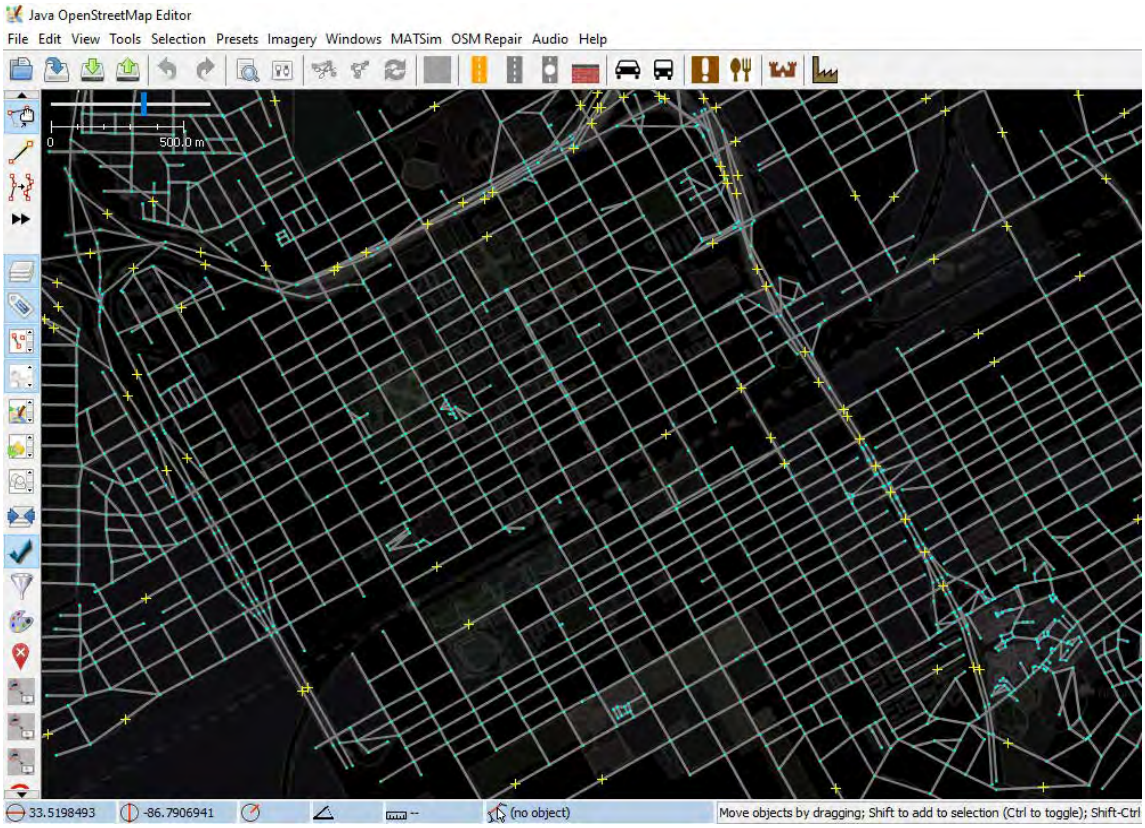


FIGURE 5.0-10 THE NETWORK IMAGE FOR BIRMINGHAM, AL USING JOSM AND MATSIM PLUGIN.

2) Collect Travel Diaries and Behavior Data. In an effort to generate the data needed in the population file, the study team designed and tested a travel diary survey to document actual travel diaries from a representative sample of the population in the study area (approximately 420 responders). The needed number of respondents were determined using Equation 1. The details of this effort are summarized in Chapter 4 of this report.

   In summary, the Qualtrics Research Core tool was used to prepare the questionnaire survey as it provided a user-friendly platform. The questionnaire was modified at various stages and was pretested and fine-tuned prior to use in order to ensure that it was easy for responders to understand the questions and provide answers. The questionnaire asked transportation users about their preferences towards using TNCs, frequency of use, reason for selection as well as asked for demographic information such as gender, age, annual income, education level, and vehicle

ownership. The criteria for collecting the demographic data were adopted from the Census criteria. Moreover, the questionnaire solicited detailed trip information of the respondents on a typical day over a 24-hour period. In the determination of the exact locations of origin and destination of the trips reported, we used Google maps API key application. This allowed respondents to easily insert the location of their origins and destinations.

Prior to launching the survey, the study team obtained an approval from the UAB office of the Institutional Review Board (IRB) for Human Use to proceed with the survey. The survey was administered in the Birmingham, AL region between December 2018 and January 2019. For quality control purposes, we went through a detailed data verification process and checked the responses received using ArcGIS software, built-in tests, and through close manual observation. Several responses were deducted from the database and new responses were collected to replace those that did not pass validation tests or showed mismatch of reported data. The final database consisted of a total of 451 responses from transportation users in the Jefferson and Shelby's counties. The stated preferences of the survey responders were used to understand the leading reasons and conditions driving the use of TNCs services in the Birmingham Metro Area. The detailed trip records provided seed data for the population synthesis task.

3) Population Synthesis. Using the travel diary data collected from the survey of Birmingham users, along with Census data, and OpenAddresses data, the study team attempted the development of a Java program to synthesize population agents (travelers) using the Iterative Proportional Fitting and the Iterative Proportional Updating (IPF/IPU) techniques. The synthesized population has to be coded as per the requirements of the simulation platform. The synthetic population can then be fed into MATSim for large-scale simulation of the Birmingham region that can produce as output a sequence of events which can be visualized and queried. While the questionnaire survey provided some useful insights regarding travel patterns in the region, the small sample size compared to the population raised questions related to data sparsity and its undesirable effects on simulation model quality. To address such concerns and enable a more realistic simulation, the study team adopted a data-driven approach to model the different aspects of travel in the Birmingham region including (1) time, (2) location, (3) activities, and (4) mode. Accordingly, the questionnaire survey data sample was enriched with public data in order to generate a more realistic population for Birmingham with realistic day plans.

More specifically, to tackle the data sparsity, we used kernel density estimation to obtain the probability density distribution (PDF) of leg start

time, leg lasting time, leg mode and leg length, which are used to sample an agent plan. We also enriched the locations captured by the survey with locations from OpenAddresses.io, but when sampling the start and end locations of legs from them, we favored locations that respect leg length distribution and closeness to a survey point. The number of legs generated among different regions were determined using IPF with the survey data as the seed, and US Census Bureau regional aggregates on user commuting. The resulting agent plans for the population of Birmingham was then fed to MATSim for simulation. The details of this effort are described next.

5.3.2.1 Population Synthesis for the Birmingham MATSim Model

Sample travel data challenges

A crucial step in the development of agent-based models is the definition of agents, e.g. household and persons. While model developers wish to capture typical workday travel patterns of the entire study population of travelers, such detailed data are unavailable due to privacy concerns, and technical and financial feasibility issues. Hence, modelers opt for population syntheses based on travel diary surveys, land use data, and census data. A conventional approach is to generate a population from a seeding survey using Iterative Proportional Fitting (IPF) (Choupani & Mamdoohi, 2016), which attempts to align the demographic attributes of survey participants with those of the entire population, whose marginal distributions are obtained from U.S. Census Bureau (U.S. Census Bureau). However, this approach has a few weaknesses:

- For user-designated geographic regions, IPF only generates the number of agents for each combination of demographic attribute values. There is no information on where these agents are, what activities they conduct and when they conduct these activities.
- Given the number of people generated above, denoted by n, a conventional population generation algorithm simply samples the pool of survey participates with the designated combination of demographic attribute values for n day plans (Choupani & Mamdoohi, 2016). While this approach compensates for the lack of spatial-temporal information on user activities, it requires a large survey sample. The detailed survey of 451 transportation system users in the Birmingham region provides a small sample compared to the population and thus each person's day plan should be sampled many times. This leads to a synthetic population where many travelers (agents) commute to work from the same origin (e.g., home location) to the same destination (e.g., work location) at the same time, which gives rise to unrealistic representation of traffic conditions.

The rest of this section briefly describes the survey data and explains the data sparsity problem that hampered the use of the conventional demography-based IPF approach for population generation in this study.

Each participant of the questionnaire survey first reported their location at 12:00AM midnight (e.g., home). Then the participant reported his/her 24-hour travel plan that provided information on the number of trip legs completed, where each leg indicates the travel from the previous location to the next location (aka. destination), along with the activity (e.g., work, home, shopping) and arrival time at the destination, the mode of travel (e.g., Car, Bike, Uber/Lyft), and if Uber/Lyft is the mode, the waiting time.

As an illustration, Figure 5-11 and Figure 5-12 show samples of data from survey responders (with identity anonymized) whose day plans have 2 legs and 4 legs, respectively. The figures report details for each trip leg, including the UTM coordinates, time interval of the travel, destination activity, and mode. For example, in Figure 5-11, Participant 009 left his/her origin at 8:10 AM for work, traveled by car and arrived at 9:00 AM (1st trip leg). The same person left work at 5:00 PM and arrived at home by car at 5:40 PM (2nd trip leg).

```
In [18]: print_trajs(2)

========== 001 ==========
(529350.66, 3727717.76),    [16:30:00, 16:37:00],    Services (e.g. Bank, post office),    Car
(529195.93, 3725994.35),    [17:10:00, 17:15:00],    Home,    Car

========== 002 ==========
(532350.68, 3716840.13),    [05:00:00, 05:20:00],    Work,    Car
(536274.28, 3721074.83),    [21:00:00, 21:20:00],    Home,    Car

========== 003 ==========
(528162.16, 3716958.22),    [16:05:00, 16:20:00],    Shopping- Retail,    Car
(524383.15, 3713958.45),    [18:10:00, 18:30:00],    Home,    Car

========== 009 ==========
(532046.64, 3688166.77),    [08:10:00, 09:00:00],    Work,    Car
(514502.46, 3712184.43),    [17:00:00, 17:40:00],    Home,    Car

========== 010 ==========
(514692.56, 3704450.57),    [15:20:00, 15:31:00],    Shopping- Grocery,    Car
(517791.45, 3706693.85),    [16:31:00, 17:00:00],    Home,    Car
```

FIGURE 5.0-11 EXAMPLES OF SURVEY RESPONDERS WITH DAY PLANS OF 2 TRIP LEGS.

```
In [20]: print_trajs(4)
(526403.17, 3697107.27),    [15:00:00, 15:30:00],    Home,    Car

========== 014 ==========
(517539.26, 3696245.51),    [03:20:00, 03:40:00],    Work,    Car
(514561.19, 3684138.06),    [12:30:00, 12:55:00],    Pick-up passenger,    Car
(528108.24, 3716180.15),    [12:58:00, 01:35:00],    Shopping- Retail,    Car
(529339.67, 3722913.29),    [15:00:00, 15:10:00],    Home,    Car

========== 016 ==========
(513811.97, 3710229.8),    [12:30:00, 12:50:00],    Services (e.g. Bank, post office),    Car
(516433.42, 3698821.97),    [14:00:00, 14:20:00],    Home,    Car
(519219.47, 3700595.9),    [18:10:00, 18:15:00],    Shopping- Grocery,    Car
(516433.42, 3698821.97),    [19:00:00, 19:10:00],    Home,    Car

========== 046 ==========
(517056.27, 3708248.89),    [17:10:00, 17:20:00],    Shopping- Retail,    Uber/Lyft
(517478.44, 3708718.93),    [17:38:00, 17:44:00],    Home,    Uber/Lyft
(521722.09, 3711567.16),    [16:42:00, 17:00:00],    Nightlife/ Bar,    Car rental
(517478.44, 3708718.93),    [19:00:00, 07:18:00],    Home,    Car rental
```

FIGURE 5.0-12 EXAMPLES OF SURVEY RESPONDERS WITH DAY PLANS OF 4 TRIP LEGS.

Using the data from the survey of the 451 Birmingham travelers described in Chapter 4, Figure 5-13 shows a histogram where the x-axis denotes the number of trip legs, and the y-axis denotes the number of persons whose day plan contained that number of trip legs. It can be seen that day plans with 2 legs are the most common, such as Home-Work-Home and Home-Shopping-Home. A small number of responders (about 20%) report 4 or more legs, including one (an Uber driver) who reported 13 trip legs in one day.

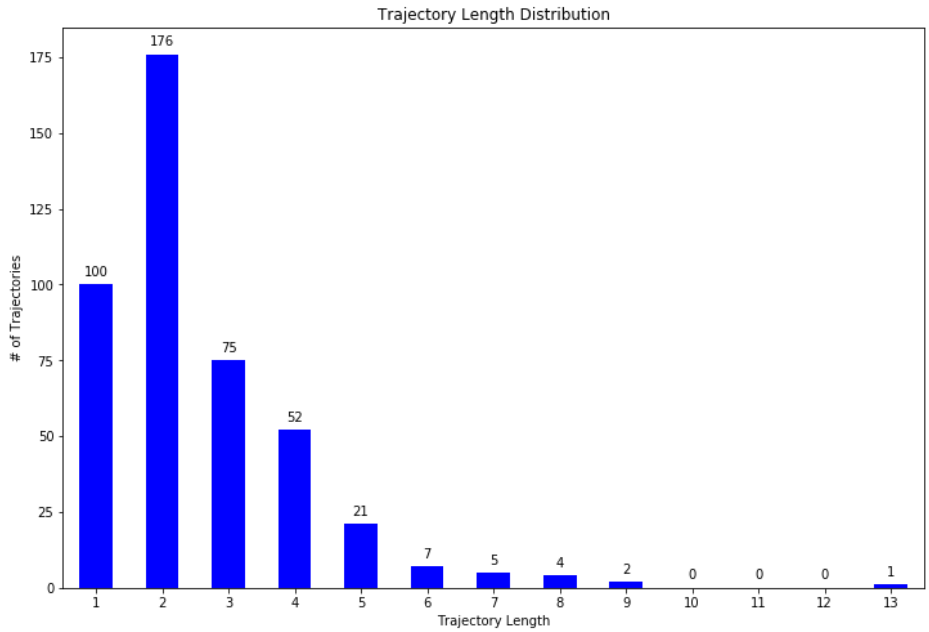


FIGURE 5.0-13 DISTRIBUTION OF NUMBER OF TRIP LEGS OF SURVEY RESPONDERS.

Additionally, each person shared his/her demographic data. A sample of related demographic attributes (such as gender at birth, age, annual household income along with their values) are shown in Figure 5-14.

```
================== Gender at birth ==================

Female     334
Male       109

================== Age ==================

25 to 34 years        120
35 to 44 years        102
18 to 24 years         89
45 to 54 years         60
55 to 64 years         45
65 to 74 years         23
75 years and over       4
```

```
================== Annual Household Income ===

$50,000 to $74,999       66
$100,000 to $149,000     59
$35,000 to $49,999       59
$75,000 to $99,999       55
Less than $10,000        49
$25,000 to $34,999       44
$15,000 to $24,999       40
$10,000 to $14,999       25
$200,000 or more         24
$150,000 to $199,999     22
```

FIGURE 5.0-14 EXAMPLES OF DEMOGRAPHIC ATTRIBUTES OF SURVEY RESPONDERS.

After consideration of the available data from the survey of Birmingham travelers, one can understand the data sparsity problem that we faced in this project. Assuming that just the 3 demographic attributes shown in Figure 5-13 are considered for IPF, then there are 2 x 7 x 10 = 140 possible combinations of attribute values, were 2 refers to the 2 options for gender, 7 refers to the 7 listed options for age group and 10 refers to the 10 annual income range options. Given that the total number of survey participants in the Birmingham study is 451, on average each combination would have data from around 3 participants. This is a too small of a sampling pool to allow diversity when scaled up to the population scale. Thus the use of the limited sample data available would result to an incorrect and unrealistic representation of activity-based trips on the MATSim network with too many transportation users sharing the same origins and destinations as well as departure times.

To address the data sparsity issue discussed above there are two options: (a) significantly increase the sample size of survey responses or (b) search for alternative ways to enrich the survey sample using publicly available data sources. Option (a) was deemed not viable due to two reasons, namely the difficulty to identify thousands of additional subjects in the Birmingham being willing to participate in the survey and the cost associated with the collection of additional survey responses that was prohibitive. Thus the research team proceeded with option (b) and adopted a data-driven approach to model the different aspects of travel including (1) time, (2) location, (3) activities, and (4) mode and generate improved and more realistic day travel plans for the Birmingham population. Details of the approach followed and related findings are discussed next.

Addressing the time modeling issue

The time aspects of a trip leg are modeled as a pair (start_time, time_span) and can be used to derive other information like: end_time = start_time +

time_span. We consider both start_time and time_span as random variables that follow a distribution conditional on the specific activity adopted. Once the activity of a trip leg is determined, we can sample (*start_time*, *time_span*) from the distribution for that activity to timestamp that trip leg.

To illustrate why conditioning time on activity makes good sense, we plotted the probability density functions (PDFs) of start_time and end_time for 4 activities, namely "Work", "Home", "Shopping-Grocery" and "Eat/Get take-out" in Figure 5-15. These PDFs are estimated from the Birmingham survey data using kernel density estimation (KDE). For example, to get the PDF of start_time for the activity "Work", we collect the start time of all legs in the survey whose destination activity is "Work", which are then input to the KDE model to fit a PDF.

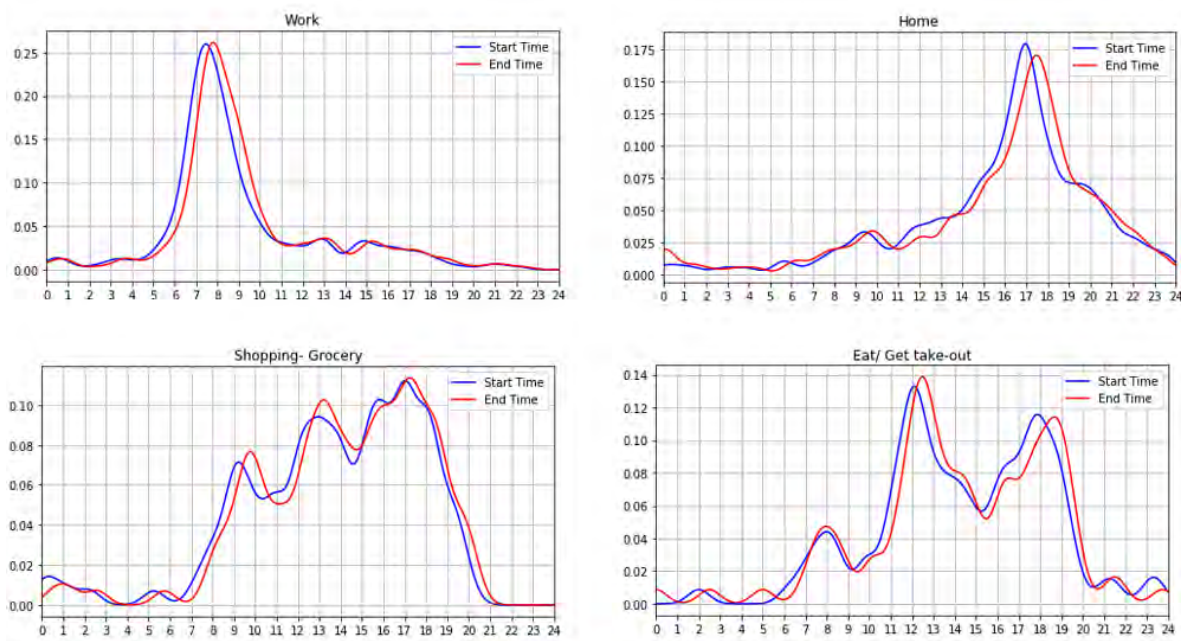


FIGURE 5.0-15 PDFS OF *START_TIME* AND *END_TIME* OBTAINED BY KERNEL DENSITY ESTIMATION (KDE).

The plots in Figure 5-15 align well with our intuition. For example, the time to "Work" peaks at 8 am and the time back "Home" peaks at 5 pm. Also, "Shopping" happens mostly between 9 am and 8 pm whereas the "Eat/Get take-out" time peaks at noon and right after work.

Figure 5-16 shows the PDF of time_span for various activities, where the x-axis is in the unit of hours. It can be seen that most trip legs finish within 1 hour, which is within expectation for travel times within the Birmingham area.

Addressing the location modeling issue

Publicly available data such as census data usually report aggregate statistics for individual geographic regions. Thus, a need exists to align the surveyed locations with these geographic regions in order to scale up the population and their locations in each region. Since we are only interested in two counties (namely Jefferson and Shelby) in our case study, a county-level granularity is too coarse for an accurate modeling. We thus choose the granularity of ZIP Code Tabulation Areas (ZCTAs). Figure 5-17 shows the counties and ZCTAs in Alabama, which are obtained from TIGERweb (TIGERweb).

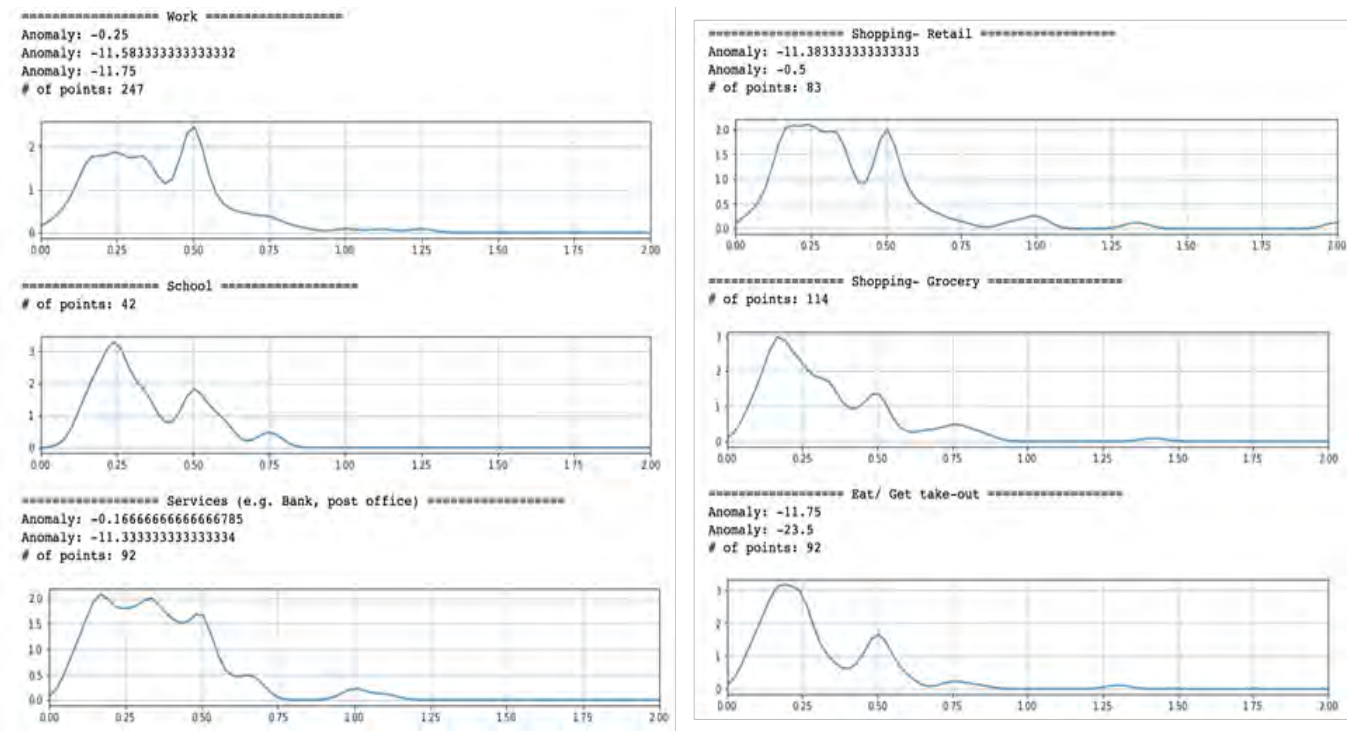


FIGURE 5.0-16. PDFS OF TIME_SPAN OBTAINED BY KERNEL DENSITY ESTIMATION (KDE).

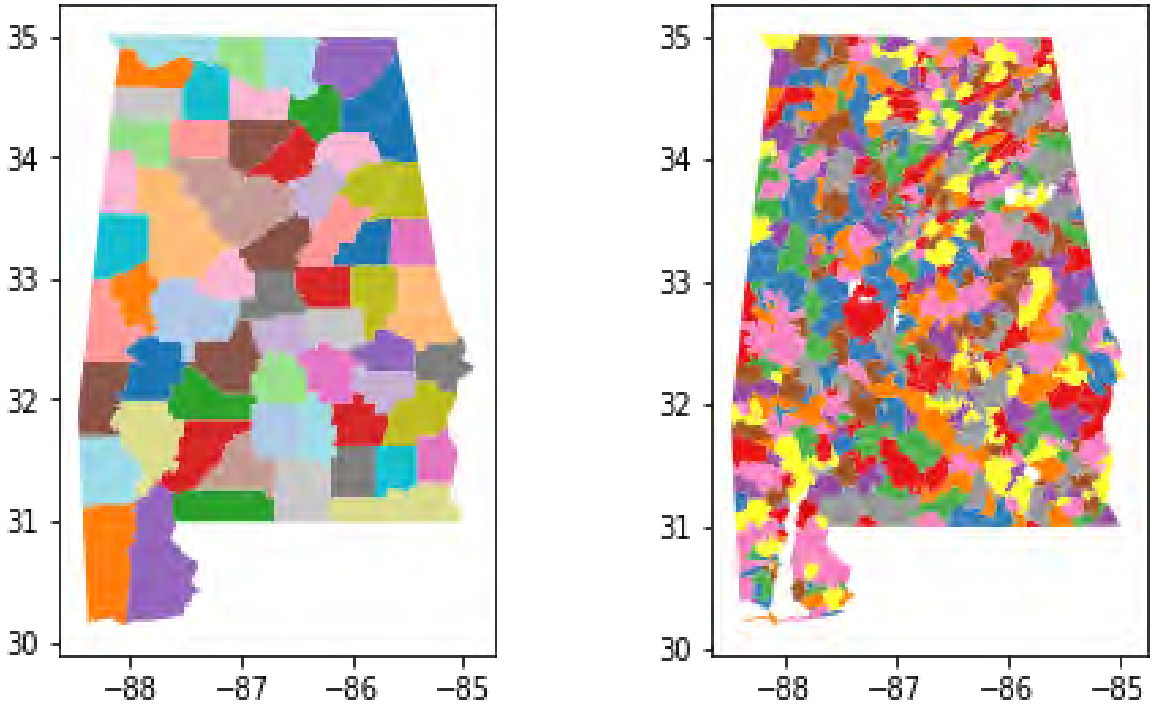
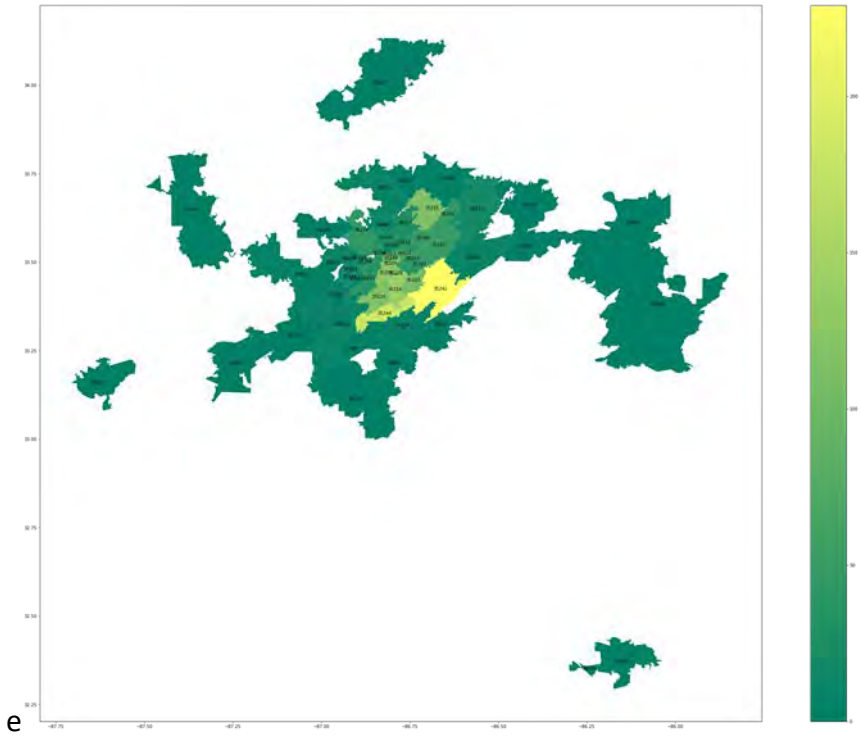


FIGURE 5.0-17. ALABAMA COUNTIES (LEFT) AND ZIP CODE TABULATION AREAS (ZCTAS) (RIGHT).

Using the questionnaire survey data, we first partitioned the surveyed locations by ZCTAs as shown in Figure 5-18, where the brighter a ZCTA is, the more surveyed locations it contains as shown in the color scale on the right-hand site. Figure 5-19 further shows how many surveyed locations each ZCTA located within the study area contains.



e

FIGURE 5.0-18. ZCTAS WITH SURVEYED LOCATIONS (MARKED WITH ZIP CODES).

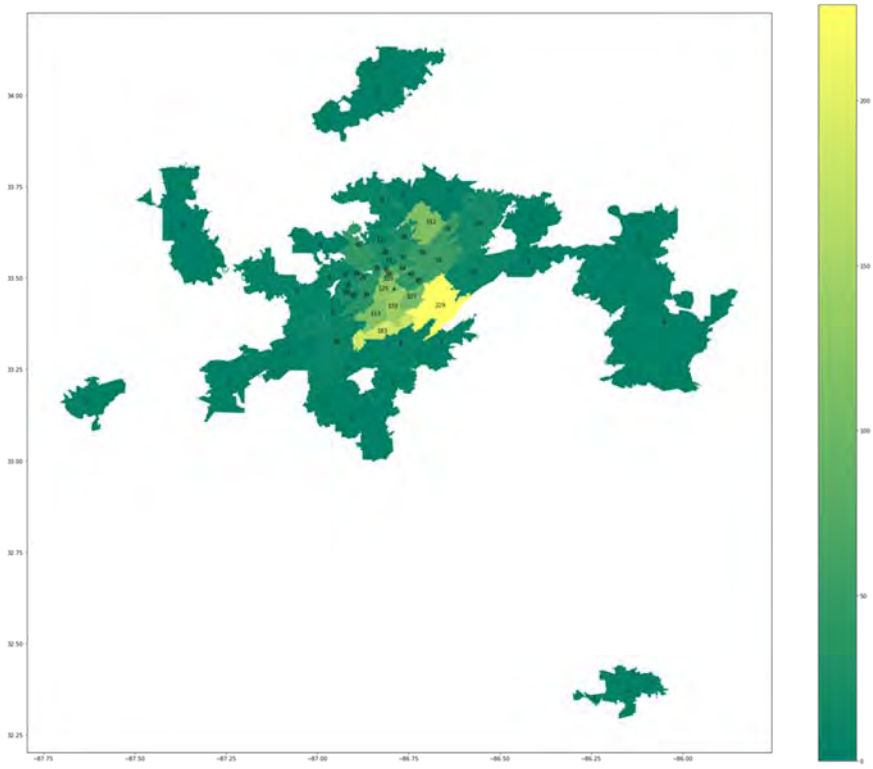


FIGURE 5.0-19 NUMBER OF SURVEYED LOCATIONS WITHIN EACH STUDY ZCTA.

In order to scale up the population from the seed survey, we can create a seed matrix *A* where each element *A*[*i*][*j*] indicates how many legs have source in ZCTA *i* and destination in ZCTA *j* and scale up the travel plans of the population

for different activities individually. For example, for "Work" trips which is often a major factor for traffic congestion, we can obtain marginals such as how many people commute to work from each ZCTA (i.e., they live there), and how many people commute to work at each ZCTA (i.e., they work there) from public data. Assuming that the seed matrix A is also constructed out of only those legs whose activity is "Work", then we can adjust A to align with the marginals using IPF as illustrated in Figure 5-20.

	35215	35173	35206	35209	35226	35205	35242	35223	35207	35233	...
35215	6979	997	2493	499	0	499	0	0	0	997	...
35173	2458	4916	1229	1229	0	1229	0	0	0	0	...
35206	2357	295	589	295	295	0	295	0	589	589	...
35209	522	0	0	4441	2090	1567	2090	784	261	1045	...
35226	446	0	0	2674	4457	446	1783	0	446	891	...
35205	210	210	0	1468	0	3145	419	0	0	1887	...
35242	0	281	281	1684	1403	561	16557	281	0	1123	...
35223	0	0	0	227	0	453	227	1812	0	227	...
35207	0	0	277	277	0	0	0	0	555	0	...
35233	45	0	23	136	23	181	0	23	0	113	...
35211	1381	0	691	691	1381	0	0	0	0	691	...
35235	3337	667	667	0	0	0	667	0	0	0	...
35222	181	0	181	363	181	363	181	181	0	363	...
35243	0	0	167	1004	167	1004	2175	335	0	836	...

FIGURE 5.0-20 A FRAGMENT OF THE ADJUSTED SEED MATRIX BY IPF.

To get the marginals of how many people commute to work from home in each ZCTA, we can use American Community Survey (ACS), more specifically, variable "P03_0018E: COMMUTING TO WORK" which indicates how many people commute to work at each region. This can be obtained using the Census API of ACS; for example, in our scenario we used the following URL request:

https://api.census.gov/data/2017/acs/acs5/profile?get=NAME,DP03_0018E&for=zip%20code%20tabulation%20area:35215,35173,35206,35209,35226,35205,35242,35223,35207,35233,35211,35235,35222,35243,35203,35216,35217,35244,35208,35210,35214,35213,35218,35221,35204,35228,35212,35094,35234,35224,35071,36117,35023,35022,35020,35126,35043,35401,35096,35080,35490,35068,35124,35116,35128,35229,35064,35077,35115,35160,35007,35579,35127,36106,35120,35005,35111

The returned data are illustrated in Figure 5-21. For example, ZCTA 35401 has 11,636 people commuting to work whereas ZCTA 35218 has 2,583 commuters.

	NAME	DP03_0018E	zip code tabulation area
0	ZCTA5 35401	11636	35401
1	ZCTA5 35071	7141	35071
2	ZCTA5 35005	2777	35005
3	ZCTA5 35213	6409	35213
4	ZCTA5 35215	20017	35215
5	ZCTA5 35111	7809	35111
6	ZCTA5 35124	12974	35124
7	ZCTA5 35203	947	35203
8	ZCTA5 35206	6862	35206
9	ZCTA5 35209	16154	35209
10	ZCTA5 35218	2583	35218

FIGURE 5.0-21 VALUES OF VARIABLE DP03_0018E FOR SELECTED ZCTAS.

Unfortunately, even though ACS has another variable "B08604_001E: WORKER POPULATION FOR WORKPLACE GEOGRAPHY", it only delves into the county level and the ZCTA level data is missing (nulls are returned). Thus we had to obtain the related data from other sources such as the major employer list from Birmingham Business Alliance (BBA), as shown in Figure 5-22. In fact, our estimate is that the list covers more than 35% of the employees in Birmingham, and we can generate workplaces for the other 65% employees following the same distribution. The marginals of how many people commute to work at each ZCTA can then be obtained for IPF.

Addressing the location scarcity issue

Now that we have the value for each A[i][j] indicating how many legs have origin in ZCTA i and destination in ZCTA j, we can generate that many trip legs about "Work" whose time information can be sampled from the PDFs for the "Work" activity as described earlier. However, it is still unclear how to generate the origin (e.g., home) and destination (e.g., office) of each leg.

2017 - 18 METROPOLITAN BIRMINGHAM MAJOR EMPLOYERS:

EMPLOYER RANK	EMPLOYER	ESTIMATED NUMBER OF EMPLOYEES	PRODUCTS OR SERVICES	COUNTY	MUNICIPALITY
1	University of Alabama at Birmingham	23,000	Education and health care services	Jefferson	Birmingham
2	Regions Financial Corporation	9,000	Financial services, banking, corporate headquarters	Jefferson	Birmingham
3	St. Vincent's Health System	5,100	Health care services, hospital network serving metro Birmingham	Jefferson	Birmingham
4	Children's of Alabama	5,000	Health care services, regional specialized health care	Jefferson	Birmingham
5	AT&T	4,517	Telecommunications, regional operations	Jefferson	Birmingham
6	Honda Manufacturing of Alabama	4,500	Manufacturing, vehicle assembly plant	Talladega	Lincoln
7	Brookwood Baptist Health	4,459	Health care services, management	Jefferson	Birmingham
8	Jefferson County Board of Education	4,400	Government, public education	Jefferson	Birmingham
9	City of Birmingham	4,200	Government, city administration	Jefferson	Birmingham
10	Mercedes-Benz U.S. International, Inc.	3,600	Manufacturing, vehicle assembly plant	Tuscaloosa	Vance
11	Blue Cross-Blue Shield of Alabama	3,100	Financial services, insurance, corporate headquarters	Jefferson	Hoover
12	Alabama Power Company	3,092	Utilities services, electrical, corporate headquarters	Jefferson	Birmingham
13	Birmingham Board of Education	2,721	Government, public education	Jefferson	Birmingham
14	Jefferson County Commission	2,500	Government, county administration	Jefferson	Birmingham
⋮					
142	Satellites Unlimited, Inc.	300	Telecommunications, corporate headquarters, wired telecommunications carrier	Jefferson	Birmingham
138	STERIS Corporation	300	Manufacturing (advanced), surgical appliance and supplies	Jefferson	Birmingham
144	Tractor & Equipment Company	300	Wholesale distribution, parts and equipment	Jefferson	Birmingham
137	U. S. Security Associates Inc.	300	Maintenance, security services	Jefferson	Birmingham
143	Wood-Fruitticher Grocery Company	300	Wholesale distribution, food	Jefferson	Birmingham
140	Yorozu Automotive Alabama	300	Manufacturing, vehicle assembly plant supplier	Walker	Jasper

Source: Birmingham Business Alliance (BBA)

FIGURE 5.0-22 BIRMINGHAM MAJOR EMPLOYER LIST (ONLY SHOWING THE HEAD AND TAIL).

One approach is to sample a random location in the corresponding ZCTA, but this is unlikely to be realistic as many places are sparsely populated or even far from the road network. Another approach is to sample only from the surveyed locations which are real. However, given that we only have 451 participants in our survey each reporting a few trip legs, there are only slightly more than 2,000 reported locations in total. This is quite a small pool to sample from if we want to generate the source and destination locations for the entire population in Birmingham, that would lead to unrealistic scenarios (e.g., many travelers commuting to work from the same house.

Fortunately, OpenAddresses (OpenAddresses) collects address data which can be used to enrich our location pool for sampling purposes. Figure 5-23 shows Jefferson county (in green) and Shelby County (in blue) with ZCTA boundaries marked. ZCTAs marked are the ones that are contained within or intersect with any of the two counties. Moreover, Figure 5-23 shows the addresses from OpenAddresses inside the two counties in red, and the surveyed locations in black. It is easy to observe that the red points significantly enrich the black ones.

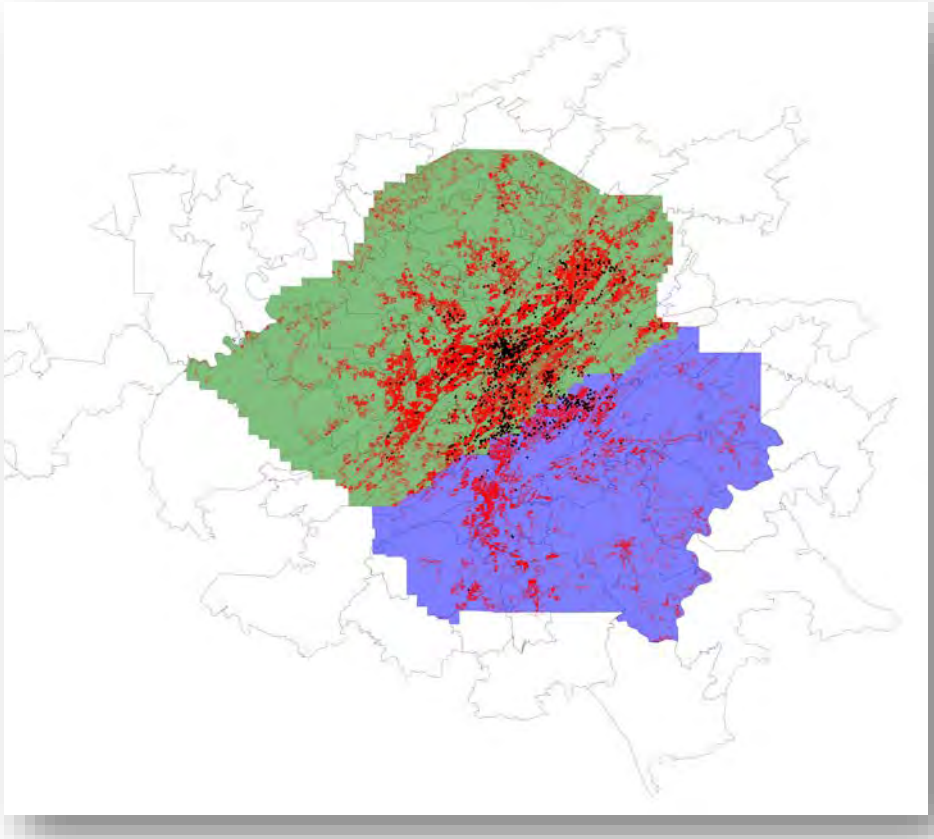


FIGURE 5.0-23 OPENADDRESSES LOCATIONS IN JEFFERSON AND SHELBY COUNTIES.

In fact, the OpenAddresses locations capture well the road network and building block structures, as shown in Figure 5-24 which is a zoomed-in version of Figure 5-23. Moreover, close inspection of Figure 5-24 shows that our surveyed locations align well with the OpenAddresses locations as the black points (that represent data from the survey of Birmingham users) appear more often at those regions where the red points are dense.

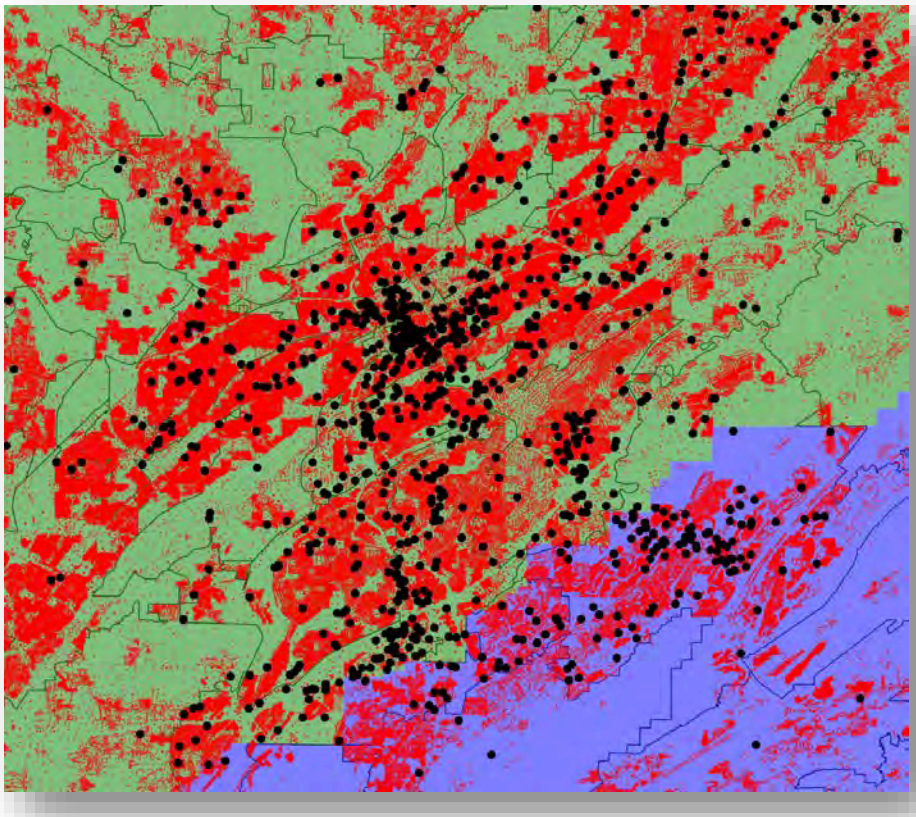


FIGURE 5.0-24 SAMPLE OF STUDY OPENADDRESSES LOCATIONS; ZOOMED-IN VERSION.

Therefore, in order to address the location sampling issue, we propose to sample "Home" locations from OpenAddresses locations which better reflect the spatial distribution of the entire population and better align with the road network. In contrast, "Work" locations can be sampled from the locations of the major employers shown in Figure 5-22 with probability proportional to the number of employees. This addresses around 35% of the possible "Work" locations. The other 65% are sampled from OpenAddresses locations, but with probability decided by fitting a Gaussian mixture model to the major employers' data. Another benefit is that this approach can accurately model trips during peak hours, for example, when the majority of the 23,000 employees of UAB (the largest employer in Birmingham) commute from work to home at 5 pm.

So far, we focused on the activities related to "Home" and "Work". However, the sample pool should consider different activity. Fortunately, OpenStreetMap maintains different types of points-of-interest (POIs) (OpenStreetMap, 2018) that correspond to different activities, as listed in Table 5-3.

TABLE 5.0-3 SURVEYED ACTIVITIES AND THE CORRESPONDING OPENSTREETMAP POI TYPES

Surveyed Activities	OpenStreetMap POI Types
Shopping-Grocery	supermarket, convenience
Services (e.g. Bank, Post Office)	bank, atm, post_office
School	school
Shopping-Retail	department_store, mall
Eat / Get Take-out	restaurant, fast_food, café, deli, bar, pub
Nightlife / Bar	nightclub, bar, pub
Drop-off Passenger	all POIs (50%) and OpenAddresses locations (50%)
Pick-up Passenger	all POIs (50%) and OpenAddresses locations (50%)

Figure 5-25 and Figure 5-26 show some examples of POIs in types "supermarket" and "school", respectively. However, it should be noted that the list of POIs in each category may not be complete since the purpose of OpenStreetMap is to share data by collaborative efforts but there is no guarantee of data completeness.

```
In [11]: key = "shop"
         value = "supermarket"
         supermarkets = extract(key, value)

         print_pois(supermarkets)

         <71> supermarket (Walmart Neighborhood Market): 33.212574, -86.828967
         <72> supermarket (Walmart Neighborhood Market): 33.428559, -86.791833
         <73> supermarket (Walmart Supercenter): 33.460574, -86.96832
         <74> supermarket (Walmart Supercenter): 33.971218, -86.446599
         <75> supermarket (Walmart Supercenter): 33.745493, -87.041799
         <76> supermarket (Walmart Supercenter): 33.421996, -86.67595
         <77> supermarket (Walmart Supercenter): 33.4471, -86.821
         <78> supermarket (Walmart Supercenter): 33.579201, -86.9243089
         <79> supermarket (Piggly Wiggly): 33.4142908, -86.8448869
         <80> supermarket (Piggly Wiggly): 33.8072708, -86.8104039
         <81> supermarket (Aldi): 33.4261712, -86.7029532
         <82> supermarket (Piggly Wiggly): 33.4814219, -86.7119523
         <83> supermarket (Publix): 33.3858798, -86.7405938
         <84> supermarket (Fresh Market): 33.4215003, -86.7000679
         <85> supermarket (Restaurant Depot): 33.4488057, -86.8268891
         <86> supermarket (Super Ofertas): 33.4709642, -86.8269303
         <87> supermarket (Brito's Supermarket): 33.4684959, -86.8257972
         <88> supermarket (Whole Foods Market): 33.3771436, -86.8007887
         <89> supermarket (Aldi): 33.6421365, -86.6226526
```

FIGURE 5.0-25. EXAMPLE OF POIS FROM OPENSTREETMAP RELATED TO "SUPERMARKETS".

```
<403> school (Calera High School): 33.0926965, -86.7672835
<404> school (Fultondale Elementary School): 33.5969094, -86.7976472
<405> school (Shades Mountain Elementary School): 33.4265383, -86.8246601
<406> school (Rock Quarry Middle School): 33.2659563, -87.513503
<407> school (Rock Quarry Elementary School): 33.2655619, -87.512763
<408> school (Cordova High School): 33.7662019, -87.1833124
<409> school (Bankhead middle School): 33.7689489, -87.1836772
<410> school (Cordova Elementary School): 33.7673257, -87.181274
<411> school (Chelsea Intermediate School): 33.3343651, -86.6365877
<412> school (Hewitt-Trussville Middle School): 33.6557363, -86.5935731
<413> school (The Horizons School): 33.4978367, -86.7928196
<414> school (Paine Primary School): 33.6490781, -86.5682636
<415> school (Paine Intermediate School): 33.6497727, -86.5670892
<416> school (Hewitt-Trussville High School): 33.6657798, -86.590175
<417> school (Hueytown High School): 33.4306238, -87.0342663
<418> school (Highlands School): 33.5128455, -86.7034581
<419> school (Nicholls-Lawson Middle School): 33.1876716, -86.2400171
<420> school (Riverchase Elementary School): 33.3596026, -86.8123587
<421> school (Glen Iris Baptist School): 33.493258, -86.8118965
```

FIGURE 5.0-26. EXAMPLE OF POIS FROM OPENSTREETMAP RELATED TO "SCHOOLS".

In order to estimate the coverage of POIs for each activity, we use *k* nearest neighbor (*k*-NN) queries to find the nearest POIs of the destination of each leg of that activity. Figure 5-27 shows the 3-NN query results for activities "Shopping-Grocery" (left) and "Eat / Get Take-out" (right) and their walking distances. We can see that many of them are an obvious hit with walking distance < 3 minutes, but a few of them are missed with walking distance > 10 min.

```
============ (33.46527270638653, -86.81967834785836) ============
>> NN 1 : dist =  0 min
supermarket (Publix): 33.465192, -86.819874
>> NN 2 : dist =  2 min
supermarket (aldi): 33.4639637, -86.8213989
>> NN 3 : dist =  3 min
supermarket (Walmart Neighborhood Market): 33.4678637, -86.8215147

============ (33.40365486174883, -86.80835731664584) ============
>> NN 1 : dist =  0 min
supermarket (Publix): 33.403505, -86.8082925
>> NN 2 : dist =  27 min
supermarket (Whole Foods Market): 33.3771436, -86.8007887
>> NN 3 : dist =  29 min
supermarket (Walmart Neighborhood Market): 33.428559, -86.791833

============ (33.42631593134768, -86.70288355495192) ============
>> NN 1 : dist =  0 min
supermarket (Aldi): 33.4261712, -86.7029532
>> NN 2 : dist =  5 min
supermarket (Fresh Market): 33.4215003, -86.7000679
>> NN 3 : dist =  12 min
supermarket (Winn-Dixie): 33.4173675, -86.6945961

============ (33.479436640015166, -86.84185798281771) ============
>> NN 1 : dist =  17 min
supermarket (Super Ofertas): 33.4709642, -86.8269303
>> NN 2 : dist =  19 min
supermarket (Brito's Supermarket): 33.4684959, -86.8257972
>> NN 3 : dist =  21 min
supermarket (Mi Pueblo): 33.4666125, -86.8243888

============ (33.42862167991548, -86.79182192815595) ============
>> NN 1 : dist =  0 min
supermarket (Walmart Neighborhood Market): 33.428559, -86.791833
>> NN 2 : dist =  17 min
supermarket (Western Supermarket): 33.4264824, -86.7742867
>> NN 3 : dist =  30 min
supermarket (Publix): 33.403505, -86.8082925
```

```
============ (33.45572623258383, -86.73523675081049) ============
>> NN 1 : dist =  0 min
cafe (Starbucks): 33.4562302, -86.7356415
>> NN 2 : dist =  1 min
fast_food (Domino's Pizza): 33.4552876, -86.7362266
>> NN 3 : dist =  2 min
restaurant (Martin's Bar-B-Que Joint): 33.4564046, -86.7374946

============ (33.47161173663921, -86.82211216492074) ============
>> NN 1 : dist =  0 min
restaurant (Full Moon Barbecue): 33.471604, -86.8220535
>> NN 2 : dist =  0 min
fast_food (Wendy's): 33.4717103, -86.8216855
>> NN 3 : dist =  0 min
restaurant (El Sol): 33.47206, -86.8229897

============ (33.645658313884795, -86.68360004711136) ============
>> NN 1 : dist =  30 min
fast_food: 33.6149233, -86.6853804
>> NN 2 : dist =  36 min
fast_food (Hardee's): 33.6538403, -86.6481455
>> NN 3 : dist =  38 min
fast_food (Subway): 33.6561182, -86.6463107

============ (33.36390800523452, -86.99999602280803) ============
>> NN 1 : dist =  3 min
restaurant (Applebee's Neighborhood Grill & Bar): 33.3645948, -87.0035811
>> NN 2 : dist =  3 min
restaurant (Cracker Barrel): 33.3654801, -87.0035613
>> NN 3 : dist =  60 min
restaurant (The Bright Star Restaurant): 33.402533, -86.9537598

============ (33.64034668569537, -86.63060998561306) ============
>> NN 1 : dist =  0 min
restaurant (Buffalo Wild Wings): 33.6405638, -86.6306627
>> NN 2 : dist =  3 min
restaurant (Costa's Mediterranean Cafe): 33.6409647, -86.6271773
>> NN 3 : dist =  3 min
fast_food (Milo's): 33.641917, -86.6271865
```

FIGURE 5.0-27. THREE NEAREST POIS AND THEIR WALKING DISTANCES.

The hit rate can be used to estimate the fraction of POIs covered, denoted by *p*. During location sampling for an activity, we sample from the OpenStreetMap POIs with probability p, and sample a random location from OpenAddresses locations with probability (1 - p). For activities related to picking up and dropping off passenger, we expect that one end of a trip leg is home and the other end is another activity, and thus currently we sample locations from OpenAddresses with 50% probability (as home locations) and from POIs of all types with 50% probability. Alternatively, the probability may also be estimated using the actual ratio observed in our survey rather than 50%.

OpenStreetMap allows us to download map data in a region with a bounding box as illustrated in Figure 5-28. Besides POIs, we also used OpenStreetMap to get the road network as required by MATSim.

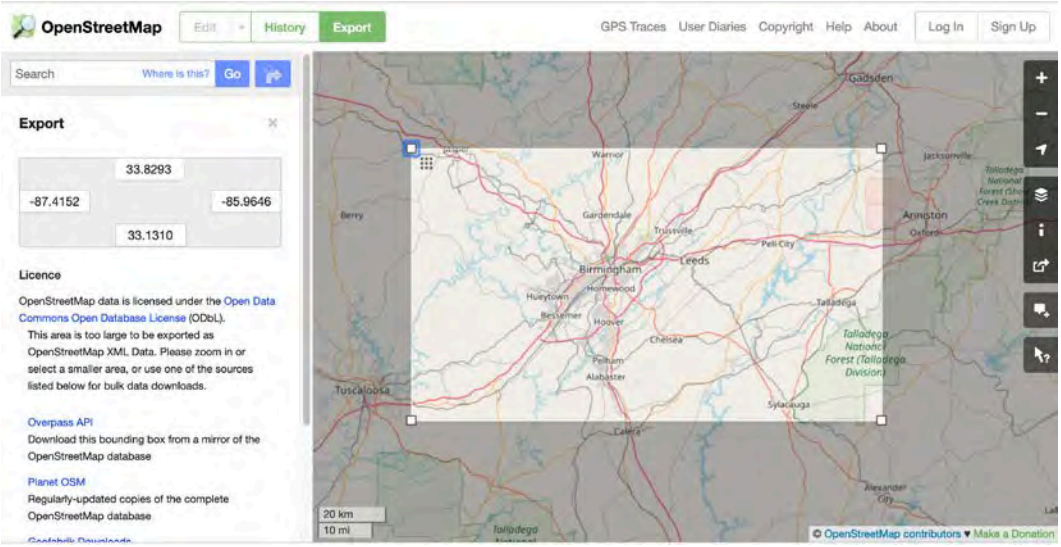


FIGURE 5.0-28. GETTING MAP DATA FROM OPENSTREETMAP.

Modeling Activity Sequence

So far, we discussed our approach for conditioning time and location generation on activities, and travel mode can be similarly generated. There is, however, one key problem remaining: how to generate a day plan as a sequence of activities?

Our solution is to generate the day plans according to the Birmingham survey data. The intuition is that the day plan of people usually does not change much with their locations; for example, most people go to work at around 8 am and leave at around 5 pm, regardless of where they are. This allows us to use the entire 451 survey participants dataset to generate day plans to counteract the data scarcity issue.

A straightforward approach is to sample from the 451 day plans in our survey. However, given a sequence of activities A1, A2, …, we need to ensure that the A1 ends before A2 starts, which means that we need to fix our previous model of time generation to one that is conditioned on an activity sequence rather than an individual activity.

Suppose that we already sampled activities A_1, A_2, …, A_{i-1}, and the next activity sampled is A_i (or we may decide that the sequence ends). To maximize the pool of samples that we can use to fit a PDF for the travel time of A_i (e.g., using KDE), we need to find all day plans among our 451 participants that have A_1, A_2, …, A_{i-1} as a subsequence. While we can find the pool in a brute-force manner for each activity sequence when we need it, here we propose a more efficient method that indexes the 451-day plans as a preprocessing step to enable much more efficient branch-and-bound search given any activity sequence as a query.

One way is to index all 451 activities using a trie (or prefix tree), in other words, an ordered tree data structure. However, in that case, (1) Home-Work-Home and (2) Home-Eat-Work-Home will diverge into two different branches. So when we have a sequence Home-Work-Home sampled and would like to decide the time for the last activity (Home) and its time (e.g., 5 pm), we will only have access to Sequence (1) following the trie, while Sequence (2) is lost, even though it also well captures the time off work. For example, the two sequences may be contributed by one person who had breakfast at home while the other had breakfast at Chick-fil-A before heading to work. Similarly, (3) Home-Work-Eat-Work-Home will not contribute to the estimation of off-work time, even though it makes perfect sense.

To tackle this problem, we first mined all frequent sequential activity patterns from the 451 day plans in our survey, where we considered an activity sequence as a frequent pattern if it is a subsequence of at least 4 persons' entire activity sequence in our survey. We used 4 as the frequency threshold since if a pattern only appears in 3 or less day plans among a total of 451, it is not statistically significant and is likely to be an outlier that should not generalize to the population.

Figure 5-29 shows the frequent patterns in non-increasing order of their frequency of appearances in the 451 day plans, based on the Birmingham survey data. It can be observed that the most frequent pattern is Home-Work-Home that appears in 154 of all the 451 day plans in our survey followed by Home-Shopping Grocery-Home with 65 appearances.

```
In [28]: ps = PrefixSpan(activities_table)

         lst = filterSeqs(ps, 2, 2)
         def getKey(item): return item[0]
         lst.sort(key = getKey, reverse = True)
         for item in lst:
             print(item)

         (154, ['Work', 'Home'])
         (65, ['Shopping- Grocery', 'Home'])
         (63, ['Eat/ Get take-out', 'Home'])
         (47, ['Shopping- Retail', 'Home'])
         (46, ['Services (e.g. Bank, post office)', 'Home'])
         (37, ['Home', 'Home'])
         (23, ['Work', 'Work'])
         (23, ['Work', 'Shopping- Grocery'])
         (23, ['Work', 'Eat/ Get take-out'])
         (23, ['School', 'Home'])
         (23, ['Pick-up passenger', 'Home'])
         (19, ['Work', 'Eat/ Get take-out', 'Home'])
         (18, ['Work', 'Work', 'Home'])
```

FIGURE 5.0-29. FREQUENT SEQUENTIAL ACTIVITY PATTERNS.

Then, instead of building a trie over all day plans, we build a trie over these frequent patterns, and for each tree-path $A_1 \rightarrow \ldots \rightarrow A_{i-1} \rightarrow A_i$ that stops at node labeled with activity A_i, we maintain the set of day plans that contain $A_1 \rightarrow \ldots \rightarrow A_{i-1} \rightarrow A_i$ as a subsequence, denoted by $D(A_i)$. Note that to construct $D(A_i)$ we only need to filter $D(A_{i-1})$ rather than going through all the 451 day plans (Figure 5-30).

```
In [222]: root.print_maxDepth(3)

          +- ROOT, 443
             +- Home, 422
             |  +- Home, 23
             |  |  +- Home, 23
             |  +- Work, 170
             |  |  +- Home, 80
             |  |  +- Shopping- Grocery, 16
             |  |  +- Eat/ Get take-out, 10
             |  |  +- Work, 10
             |  |  +- Pick-up passenger, 9
             |  |  +- Shopping- Retail, 8
             |  |  +- Services (e.g. Bank, post office), 9
             |  |  +- Drop-off passenger, 4
             |  +- Shopping- Grocery, 49
             |  |  +- Home, 23
             |  |  +- Shopping- Grocery, 8
             |  +- Services (e.g. Bank, post office), 43
             |  |  +- Home, 12
             |  |  +- Shopping- Grocery, 6
```

FIGURE 5.0-30. A FRAGMENT OF ACTIVITY TRIE

Figure 5-30 shows such a tree where each activity node A_i is also marked with how many day plans contain pattern $A_1 – … – A_{i–1} – A_i$. These frequency numbers are used to estimate the probability of sampling the next activity. For example, if we already sampled Home-Work stopping at the node with frequency 170, then we sample the next event as 'Home' with probability 80/170, while we sample 'Eat/Get take-out' with probability only 10/170.

Once the next event is sampled, we can estimate its time with the maximum possible day plans. For example, consider again the previous 3 day plans (1) Home-Work-Home, (2) Home-Eat-Work-Home and (3) Home-Work-Eat-Work-Home. Given that subsequence Home-Work-Home is already sampled and corresponding tree-path reaches a node A_i marked "Home", $D(A_i)$ contains all 3 day plans, namely (1), (2) and (3) which are used collectively to estimate the time off work. Of course, those day plans whose time off work is before the end time of $A_{i–1}$ should be filtered during the KDE fitting of the time PDF. If the number of day plans for KDE fitting is less than 4, we consider the sequence as ending at $A_{i–1}$ since the subsequent activities sequence may not generalize to the population.

The procedure described above was used to generate day plans for all agents for the Birmingham population for our case study. Our approach used publicly available datasets to supplement data obtained from the survey of Birmingham transportation network users in an effort to overcome the data sparsity encountered as we prepared the initial demand file for the MATSim simulation.

5.3.3 Birmingham MATSim Simulation

As discussed in Section 5.2.2.1, the MATSim model requires a configuration file, a network file, and the population/plans file as an input in order to run the software. After obtaining the road network from OpenStreetMap for the Birmingham study area and generating the day plans of all agents of the entire population, we fed these into MATSim as input files for simulation as shown in the job configuration file shown in Figure 5-31.

FIGURE 5.0-31. MATSIM JOB CONFIGURATION FILE FOR BIRMINGHAM CASE STUDY.

As Figure 5-32 shows and as discussed earlier, MATSim runs its activity plan iteratively until it reaches a stationary state of the system, where an agent cannot improve its score by revising the plan. Thus the agent plans that we generated are just the initial demand that is input into the simulation. MATSim executes these plans on the road network in an iterative manner that optimizes activity planning. Thus some plans may not be totally followed (e.g., due to delays related to traffic congestion, and these plans will gain a low score. Replanning will be executed on plans so that the score may go up. Each agent keeps a few plans and keeps dropping poor ones to attempt to boost the overall score of simulation. This follows the idea of evolutionary algorithm. For more detailed introduction of MATSim, please refer to the MATSim Guide (Horni et al., 2016).

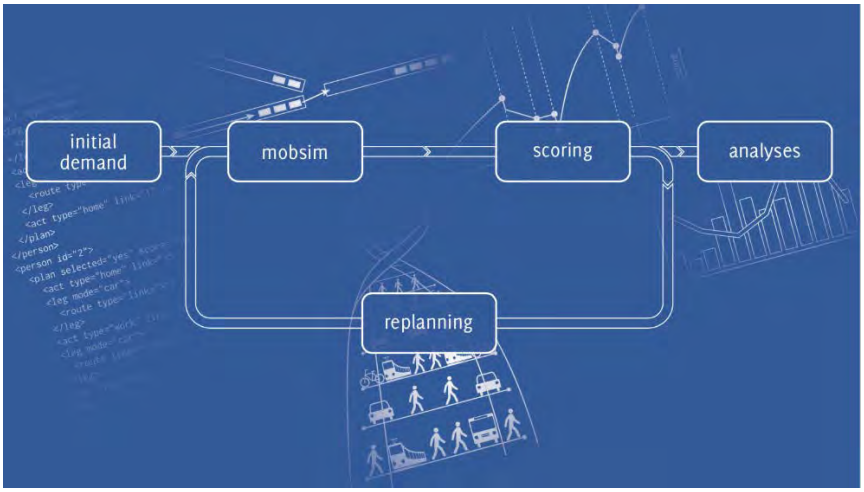


FIGURE 5.0-32. THE WORKFLOW OF MATSIM SIMULATION (HORNI ET AL., 2016).

We scaled the seed questionnaire survey into a population of the scale of Birmingham using census data (ACS) which is close to half a million. The Birmingham MATSim simulation takes around 6 hours to finish, and generates a sequence of events in chronological order. We will explain how to visualize and evaluate the simulation output in the next two sections.

5.3.3.1 Visualization of the Birmingham MATSim Simulation Output

There are two tools available for visualizing MATSim Output, namely Via and OTFViz.

Via is a powerful tool that supports advanced querying, such as automatic plotting of a traffic volume histogram on a road segment along the time dimension, by simply clicking the road segment in the visualization panel. The free version of Via supports up to 500 agents, far less than the half a million agents that we had in the Birmingham case study. Given budget restrictions, the acquisition of the Via tool for visualization of the Birmingham MATSim simulation model output was not feasible. Thus we searched for alternative options.

Another tool considered was the "On the Fly Visualizer", better known as OTFViz. This tool is MATSim's open-source visualizer that is available for free and has no restrictions on the number of agents. Given these advantages, we decided to utilize OTFViz for the Birmingham feasibility study. Figure 5-33 displays a snapshot of the Birmingham traffic in OTFViz. It should be noted that OTFViz only supports very primitive querying and while it may be appropriate to use in this feasibility study, a future extension of the project involving intensive traffic evaluation and analysis would require the purchase of Via.

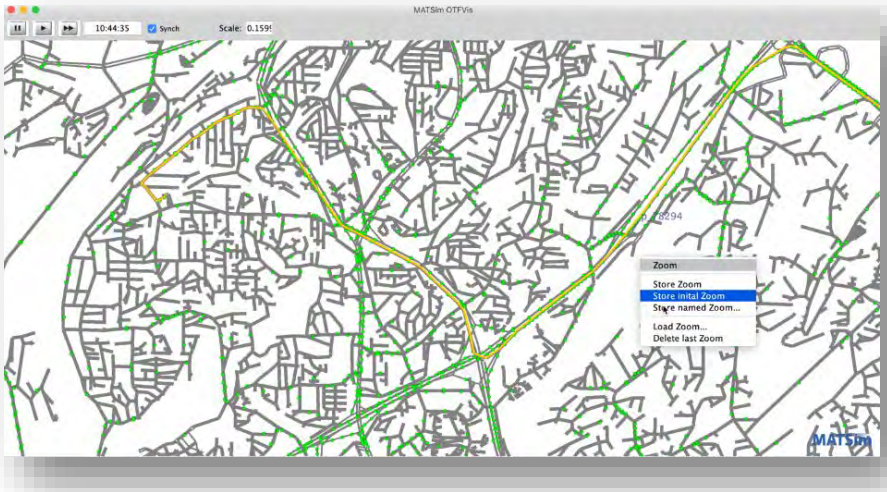


FIGURE 5.0-33. A SNAPSHOT OF THE BIRMINGHAM TRAFFIC IN OTFVIZ.

OTFViz supports the creation of a video from the file of events output by MATSim simulation, by taking a user-defined parameter indicating the time gap between two consecutive frames. Figure 5-34 shows a snapshot of such a video for the Birmingham MATSim simulation where the peripheral regions are relatively empty due to lack of data about agents in those locations. This is an artifact caused by getting of a road network that covers a wider region than our surveyed area. Each green and red point indicates the current location of an agent with the red color indicating that the corresponding agent is experiencing congested condition.

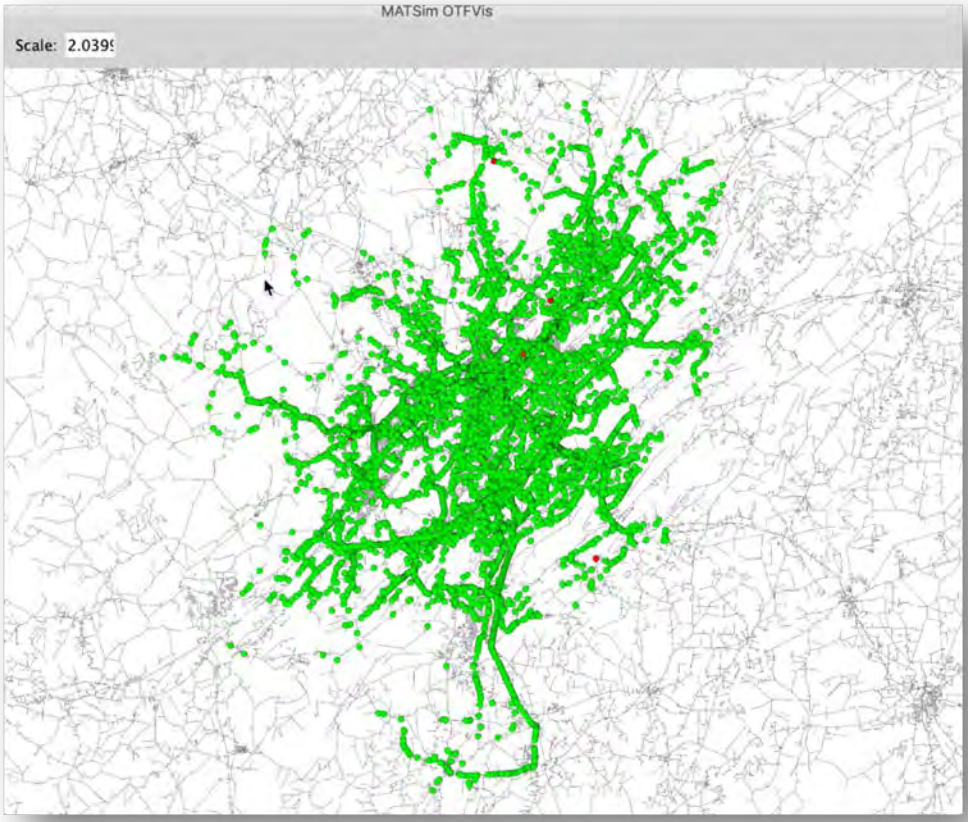


FIGURE 5.0-34. A VIDEO FRAME IN OTFVIZ FOR THE BIRMINGHAM AREA.

The user can also zoom in to see more details by dragging a bounding box using a mouse. For example, Figure 5-35 shows a zoom-in operation right before OTFViz zooms in to the scenario in Figure 5-33.

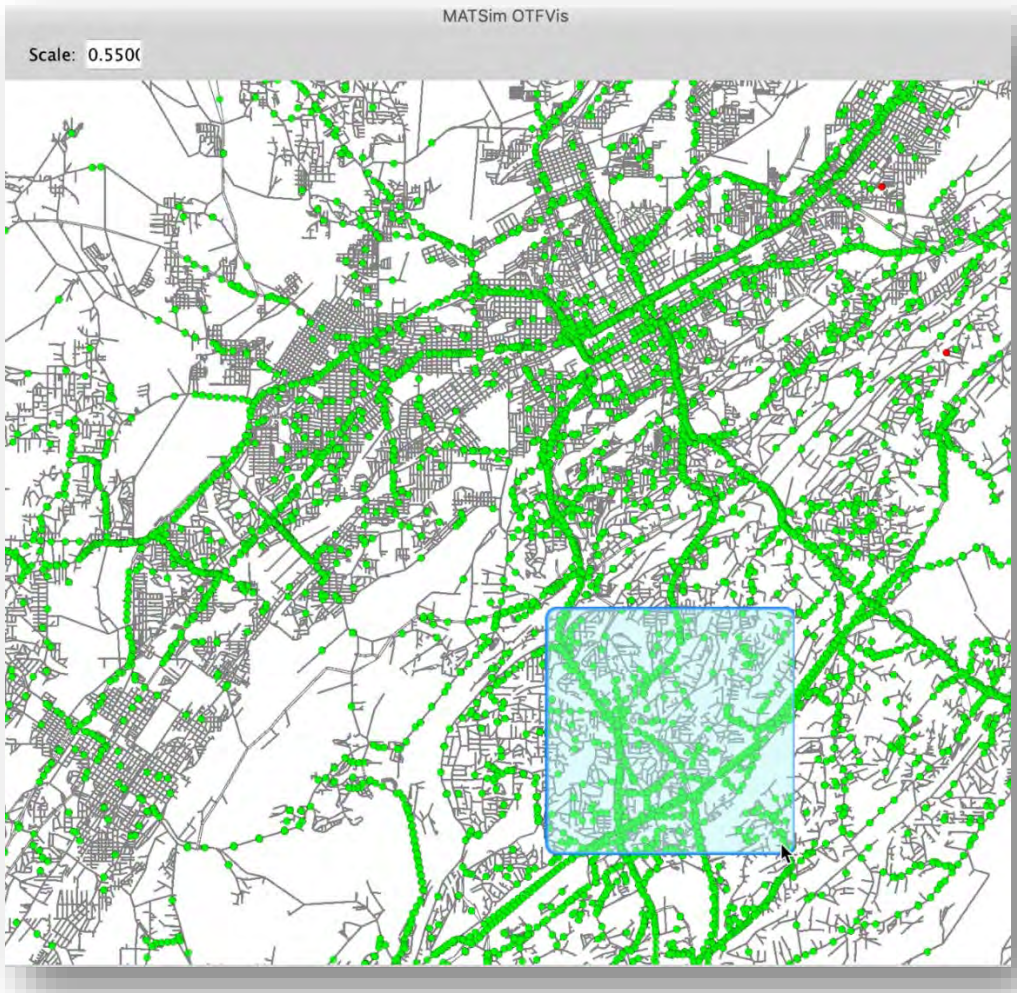


FIGURE 5.0-35 AN EXAMPLE OF THE ZOOM-IN OPERATION OF OTFVIZ.

Finally, OTFViz supports some querying operations. However, users should be cautioned that OTFViz does not support the plotting of traffic flow characteristics in a day on road segment as Via does. Although users can extend OTFViz to support more advanced visualization and querying, this requires a lot of development efforts delving into the MATSim code.

5.3.3.2 Evaluating the Birmingham MATSim Simulation Output and Future Work

The MATSim outputs consists of a sequence of events in chronological order as illustrated in Figure 5-36 using the Birmingham MATSim Simulation model output.

Load Data

```
In [2]: import xml.etree.ElementTree as ET

        tree = ET.parse('output_events.xml')
        root = tree.getroot() # <events>

In [3]: for child in root:
            print(child.attrib)

{'time': '21510.0', 'type': 'actend', 'person': '3', 'link': '1', 'actType': 'h'}
{'time': '21510.0', 'type': 'departure', 'person': '3', 'link': '1', 'legMode': 'car'}
{'time': '21510.0', 'type': 'PersonEntersVehicle', 'person': '3', 'vehicle': '3'}
{'time': '21510.0', 'type': 'vehicle enters traffic', 'person': '3', 'link': '1', 'vehicl
e': '3', 'networkMode': 'car', 'relativePosition': '1.0'}

{'time': '26758.0', 'type': 'actstart', 'person': '100', 'link': '1', 'actType': 'h'}
{'time': '27810.0', 'type': 'actend', 'person': '3', 'link': '20', 'actType': 'w'}
{'time': '27810.0', 'type': 'departure', 'person': '3', 'link': '20', 'legMode': 'car'}
{'time': '27810.0', 'type': 'PersonEntersVehicle', 'person': '3', 'vehicle': '3'}
{'time': '27810.0', 'type': 'vehicle enters traffic', 'person': '3', 'link': '20', 'vehicl
e': '3', 'networkMode': 'car', 'relativePosition': '1.0'}
{'time': '27811.0', 'type': 'left link', 'vehicle': '3', 'link': '20'}
{'time': '27811.0', 'type': 'entered link', 'vehicle': '3', 'link': '21'}
{'time': '28171.0', 'type': 'left link', 'vehicle': '3', 'link': '21'}
{'time': '28171.0', 'type': 'entered link', 'vehicle': '3', 'link': '22'}
{'time': '29431.0', 'type': 'left link', 'vehicle': '3', 'link': '22'}
{'time': '29431.0', 'type': 'entered link', 'vehicle': '3', 'link': '23'}
{'time': '29791.0', 'type': 'left link', 'vehicle': '3', 'link': '23'}
{'time': '29791.0', 'type': 'entered link', 'vehicle': '3', 'link': '1'}
{'time': '30150.0', 'type': 'vehicle leaves traffic', 'person': '3', 'link': '1', 'vehicl
e': '3', 'networkMode': 'car', 'relativePosition': '1.0'}
{'time': '30150.0', 'type': 'PersonLeavesVehicle', 'person': '3', 'vehicle': '3'}
{'time': '30150.0', 'type': 'arrival', 'person': '3', 'link': '1', 'legMode': 'car'}
{'time': '30150.0', 'type': 'actstart', 'person': '3', 'link': '1', 'actType': 'h'}
```

Figure 5.0-36. Sample of the event file output by MATSim for Birmingham.

One can see in the top part of Figure 5-36 that Person #3 ends its stay at home (Event Type: actend) and goes to work, and he/she then enters Vehicle #3 on Link #1. At the bottom part, the same person finishes work and drives home, where one can see events like “entered link 20”, “left link 20”, “entered link 21”, “left link 21”, “entered link 22”, “left link 22”, “entered link 23”, “left link 23”, “entered link 1” (where home locates), “vehicle leaves traffic” and “PersonLeavesVehicle”.

Using a one-pass streaming algorithm over the events, we can calculate performance measures to help determine the traffic conditions on every road segment for the entire day. Such data can then be compared with field data to validate the model. Model validation is an important process for determining to what extend the MATSim model properly represents actual conditions.

A detailed validation exercise was not carried out in this part of the project, which is a limitation of this part of the work. However, the research team plans to perform model validation in a future study where the Birmingham MATSim model will be further refined, expanded, and used to study congestion impacts from multi-modal integration (including Uber/Lyft and public transit) in urban settings. As part of this effort, model validation will be done by using speed data from the National Performance Management Research Data Set (NPMRDS). NPMRDS data will be obtained from Federal Highway Administration (FHWA) with the help of the Regional Planning Commission of Greater Birmingham

(RPCGB). Moreover, traffic count data will be obtained along selected study corridors through the Alabama Department of Transportation (ALDOT) and used in model validation. A map overlay showing the relevant data collection points will be generated and overlaid on the GIS map showing the study network.

In the follow up study, the research team plans to use two techniques to validate and calibrate the MATSim model. The first technique is graphical and will rely on control limits of ±15% (Dowling, Skabardonis, & Alexiadis, 2004). The second technique is analytical and will implement ANOVA to check for statistically significant differences between the simulation outputs and the validation data.

5.3.4 Conclusions and Planned Future Work

While the current project is still under development, we have demonstrated reasonable match of our simulation with the real traffic data on various study road segments. This shows that our data-driven approach using open data to address the small data problem of our user survey is effective.

In future work that has currently under way the prototype Birmingham MATSim model will be further expanded to incorporate public transit into the existing model. Test simulation scenarios will be developed and used to quantify the impacts of shared mobility options on traffic operations and regional congestion patterns. Link-based and corridor-based performance measures will be evaluated, including vehicle miles traveled (VMTs), link average speed, and link average delay. These will be used to quantify potential benefits of travel demand shift from private automobile to shared modes, in terms of improved network performance, and reduced frequency and reduced severity of congestion. The ultimate goal is to provide practical guidelines to transportation agencies that can help them better plan and operate the transportation system as a truly mode-integrated environment in the era where technology availability can facilitate such integration.

6.0 Summary Findings and Study CONTRIBUTIONS

This report summarizes results from a multidisciplinary study that used a mixed methods approach to examine and document technology influence on travelers' attitudes, preferences, and choices and their potential impact on transportation services in the Southeast. More specifically, the study investigated the influence of Transportation Network Companies (TNCs) such as Uber and Lyft, on travelers' behavior in two medium size cities in the Southeast based on three distinct but interrelated case studies, in addition to a comprehensive literature review and synthesis.

The survey of 600 millennials in North Carolina showed that most millennials surveyed had used ridehailing services—with 66% having used Lyft, Uber, or both; many on a fairly regular basis. Initial findings suggest that ridehailing services have become part of the norm. Over 65% of millennials surveyed have used Lyft or Uber or both services; many on a fairly regular basis. We found no significant differences in use or familiarity amongst ethnic or racial groups. Therefore, ridehailing services may be a way to mitigate accessibility issues. The case study findings demonstrate that even in states with small urban areas and lower densities, millennials are aware of and are taking advantage of ridehailing, carsharing, and ridesharing services.

The survey of 450 transportation users in the Birmingham Metro area documented travel practices as well as attitudes toward TNC use as a travel mode of choice. The data analysis showed that Birmingham travelers are aware of TNC services and 45% of those surveyed have used TNC services. The determinants that make TNCs a preferable mode to Birmingham travelers included convenience of use, and reduction of concerns for traffic safety (especially for late night trips to bars and eating establishments). Lack of parking availability at destination was also listed as a reason for selecting TNCs as a mode of travel along with lack of vehicle availability. Using the Birmingham metro as a case study, the analysis of survey responses provided valuable insights on the leading reasons and conditions that drive people towards the use of TNCs services in medium size cities.

The third case study evaluated the feasibility of building an agent-based simulation model of the Birmingham Metro Area in order to study the impact of shifts in travel demand due to applications of shared-use economy on local and regional congestion. Due to the fact that commonly-used traffic simulation models lack the ability to simulate shared modes in detail, the Birmingham prototype model was developed using the Multi-Agent Transport Simulation (MATSim) modeling platform and was a major undertaking in itself. The scope of this case study was limited to identifying data needs and requirements for model development and demonstrating the feasibility of a data-driven approach for addressing data sparsity issues encountered.

Future research by the research team in STRIDE Project I2 will extend this work by expanding the prototype Birmingham MATSim model to incorporate public transit and

quantifying the impacts from the integration of TNCs and transit on travel demand and congestion for various market penetration rates.

Future work also includes survey of users in Florida. A comparative analysis will be performed to identify similarities and differences in responses received in Alabama, Florida and North Carolina. While the results presented in this report establish links between technology and driving choices among transportation users in the presence of TNC service they are limited to the Southeast region. Thus, robustness of the results for other regions would need to be established by expanding the scope of the data collection to include users from other regions. The findings from the survey of users in NC and AL show that users adapt to TNC services when available. The literature confirms that more programs are being launched in the southeast and beyond. More analysis and research are needed to see how to increase access to these ridehailing and ridesharing services for all users.

The report is generating new knowledge: (a) documenting travel behavior of millennials in US Southeast and their attitudes toward transportation network companies and public transit, (b) documenting travel behaviors and transportation network companies use in Birmingham metro, and (c) demonstrating how data from such surveys can be used to support agent-based simulation modeling that can assist researchers and cities to assess TNCs impacts on transportation network's performance. The findings of the study serve as a means to understand the influence of transportation network services and better-plan mobility in mid-size cities. They also assist to address needs/opportunities of the local market and help to better understand the behaviors/choices of users and the possible influence on travel demand patterns.

7.0 REFERENCES

1. AAA. (2016). Three-Quarters of Americans "Afraid" to Ride in a Self-Driving Vehicle. Retrieved from Three quarters of Americans afraid to ride in a self-driving vehicle.

2. Abraham, J. E., Stefan, K. J., and Hunt, J. D. (2012). "Population Synthesis Using Combinatorial Optimization at Multiple Levels." Proc., Transportation Research Board 91st Annual Meeting, Transportation Research Board, Washington, DC.

3. Agatz, N., Erera, A. L., Savelsbergh, M. W., & Wang, X. (2011). Dynamic ride-sharing: A simulation study in metro Atlanta. Procedia-Social and Behavioral Sciences, 17, 532-550.

4. Alemi, F., Circella, G., Handy, S., and Mokhtarian, P. (2018). "What influences travelers to use Uber? Exploring the factors affecting the adoption of on-demand ride services in California." Travel Behaviour and Society, 13, 88-104.

5. Alemi, F., & Rodier, C. (2018). Simulation of Ridesourcing Using Agent-Based Demand and Supply Models Regional: Potential Market Demand for First Mile Transit Travel and Reduction in Vehicle Miles Traveled in the San Francisco Bay Area (National Center for Sustainable Transportation No. 18-05283). Retrieved from: The Repository and Open Science Access Portal

6. Alexander, L. P., & González, M. C. (2015). Assessing the impact of real-time ridesharing on urban traffic using mobile phone data. *Proc. UrbComp*, 1-9.

7. Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., & Rus, D. (2017). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, 201611675.

8. Althoff, M. (2010). Reachability analysis and its application to the safety assessment of autonomous cars. (Doctoral dissertation, Technische Universität München).

9. American Community Survey. Retrieved from U.S. goverment census program surveys

10. Arentze, T., Timmermans, H., and Hofman, F. (2007). "Creating Synthetic Household Populations: Problems and Approach." *Transp. Res. Rec.*, 2014, 85-91.

11. Auld, J., and Mohammadian, A. (2010). "Efficient Methodology for Generating Synthetic Populations with Multiple Control Levels." *Transp. Res. Rec.*, 2175, 138-147.

12. Axhausen, K. W., and Gärling, T. (1992). "Activity-Based Approaches to Travel Analysis: Conceptual Frameworks, Models, and Research Problems." *Transport Reviews*, 12(4), 323-341.

13. Ayed, H., Khadraoui, D., & Aggoune, R. (2015). Using MATSim to simulate carpooling and car-sharing trips. Paper presented at the 2015 World Congress on Information Technology and Computer Applications (WCITCA).

14. Balac, M., Ciari, F., & Axhausen, K. W. (2015). Carsharing Demand Estimation: Zurich, Switzerland, Area Case Study. Transportation Research Record: Journal of the Transportation Research Board, 2536, 10-18. doi:10.3141/2536-02.

15. Balac, M., Ciari, F., & Axhausen, K. W. (2017). Modeling the impact of parking price policy on free-floating carsharing: Case study for Zurich, Switzerland. *Transportation Research Part C: Emerging Technologies*, *77*, 207-225.

16. Balac, M., Ciari, F., & Waraich, R. A. (2016). Modeling the Impact of Parking Price Policy on Free-Floating Carsharing: Case Study for Zurich, Switzerland. Transportation Research Procedia.

17. BBA. Retrieved from Birmingham Business Alliance - major employers

18. BBC. (2018a). Uber halts self-driving car tests after death. Retrieved from An article from BBC Business News.

19. BBC. (2018b). Toyota suspends US driverless car tests after fatal Uber accident. Retrieved from An article from BBC Business News.

20. Beckman, R. J., Baggerly, K. A., and McKay, M. D. (1996). "Creating synthetic baseline populations." *Transportation Research Part A: Policy and Practice*, 30(6), 415-429.

21. Beer, R., Brakewood, C., Rahman, S., & Viscardi, J. (2017). Qualitative Analysis of Ridehailing Regulations in Major American Cities. *Transportation Research Board 96th Annual Meeting Compendium of Papers*, Paper No 17-00755.

22. Bischoff, J., & Maciejewski, M. (2016). Simulation of City-wide Replacement of Private Cars with Autonomous Taxis in Berlin. *Procedia Computer Science*, *83*(The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Affiliated Workshops), 237-244. doi:10.1016/j.procs.2016.04.121

23. Bliss, L. (2017). In Cities, Uber and Lyft Mean More Cars, Trips, and Miles. Retrieved from CityLab article on the ride hailing effect, more cars more trips more miles.

24. Bowman, J. L., and Rousseau, G. (2008). "Validation of Atlanta, Georgia, Regional Commission Population Synthesizer." *Proc., Innovations in Travel Demand Modeling Conference*, Transportation Research Board, Washington, DC, 54-62.

25. Brasuell, J. (2018). *Uber's CEO Wants to Take Over for Buses*. Planetizen - Urban Planning News, Jobs, and Education. Retrieved from www.planetizen.com

26. Casati, D., Müller, K., Fourie, P. J., Erath, A., and Axhausen, K. W. (2015). "Synthetic Population Generation by Combining a Hierarchical, Simulation-Based Approach with Reweighting by Generalized Raking." *Transportation Research Record: Journal of the Transportation Research Board*, 2493, 107–116.

27. Chan, N. D., & Shaheen, S. A. (2012). Ridesharing in North America: Past, present, and future. *Transport Reviews*, *32*(1), 93-112.

28. Chavis, C., & Gayah, V. V. (2017). Development of a Mode Choice Model for General Purpose Flexible Route Transit Systems. Transportation Research Record, 2650 (1), 133-141.

29. Choupani, A.-A., and Mamdoohi, A. R. (2015). "Population Synthesis in Activity-Based Models." *Transportation Research Record: Journal of the Transportation Research Board*, 2493, 1-10.

30. Choupani, A.-A., & Mamdoohi, A. R. (2016). Population Synthesis Using Iterative Proportional Fitting (IPF): A Review and Future Research. Transportation Research Procedia, 17, 223-233. doi:10.1016/j.trpro.2016.11.078.

31. Ciari, F., Balac, M., & Axhausen, K. W. (2016). Modeling Carsharing with the Agent-Based Simulation MATSim. Transportation Research Record: Journal of the Transportation Research Board, 2564, 14-20. doi: 10.3141/2564-02.

32. Chowdhury, S., & Ceder, A. (2016). "Users' willingness to ride an integrated public-transport service: A literature review." *Transport Policy* 48 (2016) 183–195.

33. Ciari, F., Balac, M., & Axhausen, K. W. (2016). Modeling Carsharing with the Agent-Based Simulation MATSim: State of the Art, Applications, and Future Developments. *Transportation Research Record: Journal of the Transportation Research Board*, (2564), 14-20.

34. Circella, G., Alemi, F., Berliner, R., Tiedeman, K., Lee, Y., Fulton, L., Handy, S. & Mokhtarian, P. L. (2017). *The Multimodal Behavior of Millennials: Exploring Differences in Travel Choices between 1 Young Adults and Gen Xers in California 2* (University of California, Davis Institute of Transportation Studies Research Report UCD_ITS_RR_17-54) Retrieved from itspubs.ucdavis.edu.

35. Circella, G., Lew Fulton, Farzad Alemi, Rosaria M. Berliner, Kate Tiedeman, Patricia L. Mohktarian, & Susan Handy. (2016). What affects Millennials' Mobility? Part I: Investigating the Environmental Concerns, Lifestyles, Mobility-Related Attitudes and Adoption of Technology of Young Adults in California.

36. Clewlow, R. (2017). *New Research on How Ride-Hailing Impacts Travel Behavior*. Planetizen - Urban Planning News, Jobs, and Education. Retrieved from www.planetizen.com features new research how ride hailing impacts travel behavior.

37. Clewlow, R., & Laberteaux, K. L. (2016). Shared-use mobility in the United States: current adoption and potential impacts on travel behavior. In *Annual Meetings of the Transportation Research Board, Washington DC*.

38. Clewlow, R. R., & Mishra, G. S. (2017). *Disruptive transportation: the adoption, utilization, and impacts of ride-hailing in the United States* (Vol. 7). Research Report–UCD-ITS-RR-17.

39. Clifton, K. J. (2004). Mobility Strategies and Food Shopping for Low-Income Families: A Case Study. Journal of Planning Education and Research, 23(4), 402-413. doi:10.1177/0739456X04264919

40. Cortright, J. (2017). *City Observatory - Uber and Lyft: A dynamic duo(poly)?* City Observatory. Available at: cityobservatory.org.

41. Deming, W. E., and Stephan, F. F. (1940). "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known." *The Annals of Mathematical Statistics*, 11(4), 427-444.

42. Dimock, M. (2018). "Defining generations: Where Millennials end and post-Millennials begin". Retrieved from www.pewresearch.org.

43. Dowling, R., Skabardonis, A., & Alexiadis, V. (2004). Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Software (FHWA-HRT-04-040).

44. Dubernet, T., Rieser-Schüssler, N., & Axhausen, K. W. (2013). Using a Multi-agent Simulation Tool to Estimate the Carpooling Potential. Paper presented at the Transportation Research Board 92nd Annual

45. Fagnant, D. J., & Kockelman, K. M. (2018). Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas. *Transportation*, *45*(1), 143-158.

46. Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167-181.

47. Fagnant, D. J., Kockelman, K. M., & Bansal, P. (2015). Operations of Shared Autonomous Vehicle Fleet for Austin, Texas, Market. *Transportation Research Record: Journal of the Transportation Research Board,* 2536, 98-106. DOI: 10.3141/2536-12.

48. Farooq, B., Bierlaire, M., Hurtubia, R., and Flötteröd, G. (2013). "Simulation based population synthesis." *Transportation Research Part B: Methodological*, 58, 243-263.

49. Feeney, M. (2015). *Policy Analysis: Is Ridesharing Safe?* Cato Institute. Retrieved from mygalaw.com, Cato Institute on Ridesharing.

50. Feigon, S., & Murphy, C. (2016). *Shared mobility and the transformation of public transit* (Transit Cooperative Research Program (TCRP), Project J-11, Task 21) doi:10.17226/23578.

51. Feigon, S., and Murphy, C. (2018). "Broadening Understanding of the Interplay Among Public Transit, Shared Mobility, and Personal Automobiles." TCRP Research Report 195, the national academic press.

52. Fournier, N., Christofa, E., Akkinepally, A. P., and Azevedo, C. L. "An Integration of Population Synthesis Methods for Agent-Based Microsimulation." Proc., Transportation Research Board 97th Annual Meeting.

53. Gilbert, G. N. (2008). *Agent-based models*. Los Angeles: Sage Publications, c2008.

54. Guo, J. Y., and Bhat, C. R. (2007). "Population Synthesis for Microsimulating Travel Behavior." *Transportation Research Record: Journal of the Transportation Research Board*, 2014, 92-101.

55. Greenwood, B. N., & Wattal, S. (2017). Show Me the Way to Go Home: An Empirical Investigation of Ride-Sharing and Alcohol Related Motor Vehicle Fatalities. MIS quarterly, 41(1), 163-187.

56. Hafezi, M. H., and Habib, M. A. (2015). "Synthesizing Population for Agent-Based Microsimulation Modeling in Atlantic Canada." *Proc., Transportation Research Board 94th Annual Meeting*, Transportation Research Board, Washington, DC.

57. Henao, A. (2017). *Impacts of ridesourcing – Lyft and Uber – on transportation including VMT, mode replacement, parking, and travel behavior* (doctoral dissertation, University of Colorado Denver).

58. Horni, A., Nagel, K., & Axhausen, K. W. (2016). The Multi-Agent Transport Simulation MATSim: ubiquity press London.

59. Huang, Y., Bastani, F., Jin, R., & Wang, X. S. (2014). Large scale real-time ridesharing with service guarantee on road networks. *Proceedings of the VLDB Endowment*, *7*(14), 2017-2028.

60. Institute of Transportation Engineers (ITE) (2011). Manual of Transportation Engineering Studies, 2nd edition.

61. Kao, S.-C., Kim, H. K., Liu, C., Cui, X., and Bhaduri, B. L. (2018). "Dependence-Preserving Approach to Synthesizing Household Characteristics." *Transportation Research Record: Journal of the Transportation Research Board*, 2302, 192-200.

62. Kim, J., and Lee, S. (2015). "A Simulated Annealing Algorithm for the Creation of Synthetic Population in Activity-based Travel Demand Model." *KSCE Journal of Civil Engineering*, 20(6), 2513-2523.

63. Klein, N. J., & Smart, M. J. (2017). Car today, gone tomorrow: The ephemeral car in low-income, immigrant and minority families. Transportation(44), 495-510.

64. Lee, D.-H., and Fu, Y. (2011). "Cross-Entropy Optimization Model for Population Synthesis in Activity-Based Microsimulation Models." *Transportation Research Record: Journal of the Transportation Research Board*, 2255, 20-27.

65. Lewyn, M. (2018). *Why Is Transit Ridership Declining?* Planetizen - Urban Planning News, Jobs, and Education. Retrieved from www.planetizen.com why transit ridership is declining.

66. Li, Z., Hong, Y., & Zhang, Z. (2016a). An empirical analysis of on-demand ride sharing and traffic congestion. Thirty Seventh International Conference on Information Systems, Dublin.

67. Li, Z., Hong, Y., & Zhang, Z. (2016b). Do Ride-Sharing Services Affect Traffic Congestion? Evidence from Uber Entry. Available at SSRN: ssrn.com abstract or dx.doi.org

68. Ma, L., and Srinivasan, S. (2015). "Synthetic Population Generation with Multilevel Controls: A Fitness-Based Synthesis Approach and Validations." *Computer-Aided Civil and Infrastructure Engineering*, 30(2), 135-150.

69. Malodia, S. & Singla, H. (2016) A study of carpooling behavior using a stated preference web survey in selected cities of India, *Transportation Planning and Technology*, 39:5, 538-550, DOI: 10.1080/03081060.2016.1174368

70. Martin, E. W., and Shaheen, S. A. (2011). "Greenhouse gas emission impacts of carsharing in North America." IEEE transactions on intelligent transportation systems, 12(4), 1074-1086.

71. Martínez, L. M., Correia, G. H. d. A., Moura, F., & Mendes Lopes, M. (2016). Insights into carsharing demand dynamics: Outputs of an agent-based model application to Lisbon, Portugal. International Journal of Sustainable Transportation, 11(2), 148-159. doi:10.1080/15568318.2016.1226997.

72. MATSim Community. (n. d.) Multi-Agent Transport Simulation (MATSim). Retrieved on June 30, 2019 from www.matsim.org.

73. Merat, N., Madigan, R., & Nordhoff, S. (2016). Human factors, user requirements, and user acceptance of ride-sharing in automated vehicles. In *International Transport Forum Roundtable on Cooperative Mobility Systems and Automated Driving*. Retrieved from www.itf-oecd.org, human factors user requirements and user acceptance ride sharing automated vehicles

74. Müller, K., and Axhausen, K. W. "Population Synthesis for Microsimulation: State of the Art." *Proc., Transportation Research Board 90th Annual Meeting*, Transportation Research Board.

75. OpenAddresses. Retrieved from openaddresses.io

76. OpenStreetMap. (2018). Map features. Retrieved from wiki.openstreetmap.org

77. Otani, N., Sugiki, N., Vichiensan, V., and Miyamoto, K. (2018). "Modifiable Attribute Cell Problem and Solution Method for Population Synthesis in Land Use Microsimulation." *Transportation Research Record: Journal of the Transportation Research Board*, 2302, 157-163.

78. Papacostas, C. S., and Prevedouros, P. D. (2001). *Transportation Engineering and Planning*, Prentice Hall, Englewood Cliffs, NJ.

79. Paul, B. M., Doyle, J., Stabler, B., Freedman, J., and Bettinardi, A. (2018). "Multi-level Population Synthesis Using Entropy Maximization-Based Simultaneous List Balancing." *Proc., Transportation Research Board 97th Annual Meeting*, Transportation Research Board, Washington, DC.

80. Pritchard, D. R., and Miller, E. J. "Advances in Agent Population Synthesis and Application in an Integrated Land Use and Transportation Model." *Proc., Transportation Research Board 88th Annual Meeting*, Transportation Research Board.

81. Quarles, N. T. (2017). Americans' plans for acquiring and using electric, shared and self-driving vehicles. (Master's thesis, The University of Texas, Austin).

82. Ramadan, O. E., & Sisiopiku, V. P. (2019). A Critical Review on Population Synthesis for Activity-and Agent-Based Transportation Models Transportation: IntechOpen.

83. Ralph, K. M. (2017). Multimodal Millennials? The Four Traveler Types of Young People in the United States in 2009. Journal of Planning Education and Research, 37(2), 150-163.

84. Rayle L, Shaheen S, Chan N, Dai D, Cervero R (2014) App-based, on-demand ride services: Comparing taxi and ridesourcing trips and user characteristics in San Francisco University of California Transportation Center (UCTC). *Technical report, UCTC-FR-2014-08.*

85. Rayle, L., Dai, D., Chan, N., Cervero, R., & Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*, *45*, 168-178.

86. Rodier, C., Alemi, F., & Smith, D. (2016). Dynamic Ridesharing: Exploration of Potential for Reduction in Vehicle Miles Traveled. Transportation Research Record, 2542(1), 120-126.

87. Ronald, N., Thompson, R., & Winter, S. (2015a). Simulating Demand-responsive Transportation: A Review of Agent-based Approaches. *Transport Reviews,* 35(4), 404- 421. DOI: 10.1080/01441647.2015.1017749

88. Ronald, N., Thompson, R. G., & Winter, S. (2015b). Comparison of Constrained and Ad Hoc Demand-Responsive Transportation Systems. *Transportation Research Record: Journal of the Transportation Research Board,* 2536, 44-51. DOI: 10.3141/2536-06

89. Ronald, N., Yang, J., & Thompson, R. G. (2016). Exploring Co-Modality Using On-Demand Transport Systems. *Transportation Research Procedia,* 12, 203-212. DOI: 10.1016/j.trpro.2016.02.059.

90. Roy, S. (2016). The Impacts of gender, personality, and previous use on attitude towards the sharing economy and future use of the services. (Master's thesis, California State University, Fresno) Retrieved from repository.library.fresnostate.edu.

91. Ryan, J., Maoh, H., and Kanaroglou, P. (2009). "Population Synthesis: Comparing the Major Techniques Using a Small, Complete Population of Firms." *Geographical Analysis*, 41(2), 181-203.

92. Saadi, I., Eftekhar, H., Teller, J., and Cools, M. (2018) "Investigating the Scalability in Population Synthesis: A Comparative Approach." *Proc., Transportation Research Board 96th Annual Meeting*, Transportation Research Board.

93. Saadi, I., Mustafa, A., Teller, J., and Cools, M. (2016a). "Forecasting travel behavior using Markov Chains-based approaches." *Transportation Research Part C: Emerging Technologies*, 69, 402-417.

94. Saadi, I., Mustafa, A., Teller, J., Farooq, B., and Cools, M. (2016b). "Hidden Markov Model-based population synthesis." *Transportation Research Part B: Methodological*, 90, 1-21.

95. Saidallah, M., El Fergougui, A., & Elalaoui, A. E. (2016). A Comparative Study of Urban Road Traffic Simulators. In *MATEC Web of Conferences* (Vol. 81, p. 05002). EDP Sciences. Retrieved from www.matec-conferences.org.

96. Sakaria, N., & Stehfest, N. (2013). Millennials and Mobility: Understanding the Millennial Mindset and New Opportunities for Transit Providers. Washington DC: The National Academies Press.

97. Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S. H., & Ratti, C. (2014). Quantifying the benefits of vehicle pooling with shareability networks. Proceedings of the National Academy of Sciences, 111(37), 13290-13294.

98. Schaller, B. (2017a). *UNSUSTAINABLE? The growth of app-based ride services and traffic, travel and the future of New York City*. Brooklyn, NY: Schaller Consulting. Retrieved from schallerconsult.com.

99. Schaller, B. (2017b). Empty seats, full streets: Fixing Manhattan's Traffic Problem. Brooklyn, NY: Schaller Consulting. Retrieved from www.schallerconsult.com/rideservices/emptyseats

100. Schoettle, B., and Sivak, M. (2016). Motorists' Preferences for Different Levels of Vehicle Automation: 2016. Retrieved from: www.umich.edu.

101. Schrank, D., Eisele, B., Lomax, T., & Bak, J. (2015). 2015 Urban Mobility Scorecard. College Station, TX: Texas A&M Transportation Institute. Retrieved from: static.tti.tamu.edu/tti.tamu.edu/documents/mobility-scorecard-2015

102. Shaheen, S., Chan, N., Bansal, A., & Cohen, A. (2015). Shared mobility. Definitions, industry developments, and early understanding. *University of California Berkeley Transportation Sustainability Research Center, Berkeley.* Retrieved from innovativemobility.org/wp-content/uploads/2015/11/SharedMobility_WhitePaper_FINAL

103. Shaheen, S., Cohen, A., Zohdy, I. (2016). Shared Mobility: Current Practices and Guiding Principles. U.S. Department of Transportation, Federal Highway Administration. Report No. FHWA-HOP-16-022. Retrieved from ops.fhwa.dot.gov/publications/fhwahop16022/fhwahop16022

104. Shaheen, S. A., Cohen, A. P., Zohdy, I. H., and Kock, B. (2016). "Smartphone applications to influence travel choices: practices and policies." United States. Federal Highway Administration.

105. Shen, W., & Lopes, C. (2015,). Managing autonomous mobility on demand systems for better passenger experience. Paper presented at International Conference on Principles and Practice of Multi-Agent Systems (pp. 20-35). Retrieved from link.springer.com/chapter/10.1007/978-3-319-25524-8_2

106. Sisiopiku, V. (2018). "Travel Patterns and Preferences of Urban University Students". Athens Journal of Technology & Engineering, Vol. 5(1), March 2018, pp. 19-31.

107. Sisiopiku, V. P., and Ramadan, O. E. (2017). Understanding Travel Behavior and Mode Choice of Urban University Campus Employees. Proceedings of the 58th Transportation Research Forum Annual Conference, Chicago, IL.

108. Siuhi, S., & Mwakalonge, J. (2016). Opportunities and challenges of smart mobile applications in transportation. *Journal of Traffic and Transportation Engineering (English Edition),* 3(6), 582-592. DOI: 10.1016/j.jtte.2016.11.001

109. Sivak, M. (2014). "Has motorization in the US peaked? Part 4: Households without a light-duty vehicle." Transport Research Institute, University of Michigan, Ann Arbor, Report No UMTRI-2014-5.

110. Sivak, M. (2015). "Has motorization in the US peaked? Part 7: Update through 2013." Transport Research Institute, University of Michigan, Report No UMTRI-2015-10.

111. Smith, A. (2016). 2. On-demand: Ride-hailing apps. Chapter 2 in *Shared, Collaborative and On Demand*. Pew Research Center: Internet, Science & Tech. Retrieved from www.pewinternet.org/2016/05/19/on-demand-ride-hailing-apps.

112. Sochor, J., Strömberg, H., & Karlsson, I. C. M. (2015). Implementing Mobility as a Service: Challenges in Integrating User, Commercial, and Societal Perspectives. *Transportation Research Record: Journal of the Transportation Research Board,* 2536, 1-9. DOI: 10.3141/2536-01

113. Sperling, D. and Brown, A. (2018). *How Lyft and Uber Can Fix—Not Cause—Congestion*. Planetizen - Urban Planning News, Jobs, and Education. Retrieved from www.planetizen.com/features/97135-how-lyft-and-uber-can-fix-not-cause-congestion.

114. Srinivasan, S., Ma, L., and Yathindra, K. (2008). "Procedure for Forecasting Household Characteristics for Input to Travel-demand Models." Florida Department of Transportation, Tallahassee, FL.

115. Stepler, R., & Lopez, M. H. (2016). "U.S. Latino Population Growth and Dispersion Has Slowed Since Onset of the Great Recession".

116. Sun, C., & Edara, P. (2015). Is Getting an Uber-Lyft from a Sidecar Different from Hailing a Taxi? Current Dynamic Ridesharing Controversy. *Transportation Research Record: Journal of the Transportation Research Board,* 2536, 60-66. DOI: 10.3141/2536-08.

117. Sun, L., and Erath, A. (2015). "A Bayesian network approach for population synthesis." *Transportation Research Part C: Emerging Technologies*, 61, 49-62.

118. TIGERweb. Retrieved from tigerweb.geo.census.gov

119. Tomer, A. (2017). *America's commuting choices: 5 major takeaways from 2016 census data*. Brookings. Retrieved from www.brookings.edu/blog/the-avenue/2017/10/03/americans-commuting-choices-5-major-takeaways-from-2016-census-data.

120. US Census Bureau. Retrieved from www.census.gov

121. Villwock-Witte, N., & Clouser, K. (2016). Mobility Mindset of Millennials in Small Urban and Rural Areas.

122. Vovsha, P., Hicks, J. E., Paul, B. M., Livshits, V., Maneva, P., and Jeon, K. (2015). "New Features of Population Synthesis." *Proc., Transportation Research Board 94th Annual Meeting,* Washington, DC.

123. Weinhalt, B. (2016). Automotive Industry. World Economic Forum White Paper Digital Transformation of Industries: In collaboration with Accenture. Retrieved from reports.weforum.org/digital-transformation/wp-content/blogs.dir/94/mp/files/pages/files/wef-dti-automotivewhitepaper-final-january-2016.v1.

124. Wielinski, G., Trépanier, M., & Morency, C. (2015). What about Free-Floating Carsharing? Transportation Research Record: Journal of the Transportation Research Board, 2536, 28-36. doi:10.3141/2536-04.

125. Williamson, P., Birkin, M., and Rees, P. H. (1998). "The Estimation of Population Microdata by Using Data from Small Area Statistics and Samples of Anonymised Records." *Environment and Planning A: Economy and Space*, 30(5), 785-816.

126. World Health Organization (WHO) (2018). Global Status Report on Road Safety, 2018. Geneva, Switzerland: WHO. Retrieved from https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/

127. Xu, H., Ordóñez, F., & Dessouky, M. (2015). A traffic assignment model for a ridesharing transportation market. *Journal of Advanced Transportation*, 49(7), 793-816.

128. Ye, X., Konduri, K. C., Pendyala, R. M., Sana, B., and Waddell, P. "Methodology to Match Distributions of Both Household and Person Attributes in Generation of Synthetic Populations." *Proc., Transportation Research Board 88th Annual Meeting*, Transportation Research Board.

129. Zheng, H., Son, Y. J., Chiu, Y. C., Head, L., Feng, Y., Xi, H., Kim, S. & Hickman, M. (2013). *A primer for agent-based simulation and modeling in transportation applications* (No. FHWA-HRT-13-054). McLean, VA: Federal Highway Administration. Retrieved from www.fhwa.dot.gov/publications/research/ear/13054/13054

130. Zhu, M., Liu, X. Y., Tang, F., Qiu, M., Shen, R., Shu, W., & Wu, M. Y. (2016). Public vehicles for future urban transportation. *IEEE transactions on intelligent transportation systems*, *17*(12), 3344-3353. Retrieved from ieeexplore.ieee.org/document/7471482

131. Zhu, Y., and Ferreira, J. (2014). "Synthetic Population Generation at Disaggregated Spatial Scales for Land Use and Transportation Microsimulation." *Transportation Research Record: Journal of the Transportation Research Board*(2429), 168-177.

132. Zhuge, C., Li, X., Ku, C.-A., Gao, J., and Zhang, H. (2017). "A heuristic-based population synthesis method for micro-simulation in transportation." *KSCE Journal of Civil Engineering*, 21(6), 2373-2383.

133. Zohdy, I., Huang, N., Keegan, N., & Lukens, J. (2016). Impacts of Millennial Student Loan Debt on Transportation Choices. (FHWA-PL-17-015). Washington, DC: FHWA, Office of Policy.

8.0 APPENDIX

8.1 NC Survey instrument

Introduction and Consent Form

We are a team of researchers from University of North Carolina at Chapel Hill who study transportation. The following survey asks how you travel and will help us understand the travel needs of the future. To understand your travel needs, we'll also ask about where your live, work, and your use of smartphone app as they impact how you travel. This research is funded by the US Department of Transportation through funding for the Southeastern US region.

Before answering any question in this survey, you will need to read the consent form on the next page and check "consent" at the bottom if you agree to taking this survey.

Consent Form for North Carolina Millennial Study

What are some general things you should know about research studies? You are being asked to take part in a research study. To join study is voluntary. You may choose not to participate, or you may withdraw your consent to be in the study, for any reason, without penalty. Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study.

What is the purpose of this study? The purpose of this research study is to collect data about how millennials travel and use of social me in their daily lives, particularly in North Carolina.

How many people will take part in this study? If you decide to be in this study, you will be one of approximately 600 people in this research study.

What will happen if you take part in the study? Your part in this study will last approximately 15 minutes. During this study, you will comp an online questionnaire. The questionnaire will ask you to describe how you commute to work or school, types of social media apps you us are familiar with, and your housing situation.

What are the possible benefits from being in this study? Research is designed to benefit society by gaining new knowledge. You may n benefit personally from being in this research study.

What are the possible risks or discomforts involved from being in this study? We anticipate few risks in this study.

How will your privacy be protected? All of the data you provide will be stored securely. We will not store data and remove identifiers such an IP address after we have checked for any duplicates.

What if you want to stop before your part in the study is complete? You can withdraw from this study at any time, without penalty and any question for any reason. The investigators also have the right to stop your participation if you have an unexpected reaction, have failed follow instructions, etc.

Will you receive anything for being in this study? Will it cost anything? You will receive no monetary reward for participating in this study. There are no costs associated with being in the study.

What if you have questions about this study? You have the right to ask, and have answered, any questions you may have about this research. Contact the principal investigator listed above with any questions, complaints, or concerns you may have.

What if you have questions about your rights as a research participant? All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns, or if you would like to obtain information or offer input, plea contact the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu

Consent ○ Don't consent ○

Filter questions

What year were you born?

Please enter your zip code:

Please list the city or town you live in:

How long have you lived in North Carolina?

- ☐ Less than a year
- ☐ 1 -5 years
- ☐ over 5 years
- ☐ Born and raised

Are you Spanish, Hispanic, or Latino?

- ○ Yes
- ○ No

Ethnicity (choose all that apply):

- ☐ White
- ☐ Black or African American
- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Native Hawaiian or Pacific Islander
- ☐ Other

Sex
- ○ Female
- ○ Male
- ○ _____ Other

Housing

The following questions are about your housing and living arrangements

How would you describe your home?
- ○ Single family home
- ○ Apartment
- ○ Condominium/Townhouse
- ○ _____ Other

How long have you lived in your current home?
- ○ less than one year
- ○ 1-2 years
- ○ Over 2 years

Which of the following best describes your current living situation?
- ○ Married and living with my spouse
- ○ Living with significant other
- ○ Living with parents or other family members
- ○ Living with roommates or friends
- ○ Living alone
- ○ _____ Other

Reason for living with parent or other family members (choose all that apply):
- ☐ Rents are high in my area/I need to save money.
- ☐ I moved back to help my family and/or relatives.
- ☐ It is part of my culture to live with parents or relatives until I am married or older.
- ☐ I am going to a local school and living in a dorm or on my own doesn't make sense.
- ☐ Student loan payments make it difficult to live on my own or with roommates.
- ☐ I needed help with child care.
- ☐ I have mental or physical challenges that require assistance.
- ☐ I am taking time off from school.
- ☐ _____ Other

Including yourself, how many people live in your household?

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6 or more

Are there any children under the age of 18 in your household?

- ○ Yes
- ○ No

Are they currently attending school outside of the home? This includes daycare, preschool program as well as K-12 schools.

- ○ Yes
- ○ No

The following questions are about general travel and vehicle ownership.

In a typical week, which of the following forms of transportation do you use? This includes travel to school, work, errands, etc.

- ☐ Walk
- ☐ Bike
- ☐ Public Transit (Bus, Light Rail)
- ☐ Car
- ☐ Carpool/get a ride from someone
- ☐ Ridehailing service (Lyft, Uber)
- ☐ _____ Other

Taking yesterday as an example, what modes did you use? (Check all that apply)

- ☐ Walk
- ☐ Bike
- ☐ Public Transit (Bus, Light Rail)
- ☐ Car
- ☐ Carpool/get a ride from someone
- ☐ Ridehailing service (Lyft, Uber)
- ☐ _____ Other

Do you have a driver's license?

- ○ Yes
- ○ No. I don't want one.
- ○ No, but I'm planning on getting one within the next year.
- ○ [ ] Other, please explain.

How many vehicles (cars, trucks, SUVs) are available for you to use in your household?

- ○ None
- ○ 1
- ○ 2
- ○ 2
- ○ 4 and over

Which best describes your vehicle ownership status:

- ○ I currently **own** a vehicle
- ○ I currently **lease** a vehicle
- ○ I have regular access to a vehicle that someone else in my household **owns**.
- ○ I have regular access to a vehicle that someone else in my household l**eases.**
- ○ I don't currently own a vehicle but plan to **buy** one in 1-2 years.
- ○ I don't currently own a vehicle but plan to **lease** one in 1-2 years.
- ○ I don't currently own a vehicle and have no plan to **lease** or **buy** one in the immediate future.
- ○ I have regular access to a vehicle that someone else outside my household owns/leases.
- ○ I will have regular access to a vehicle that someone else outside my household owns/leases.

As far as transportation, is there anything you would like to improve in your community? (This could things such as improving sidewalks, building more parking facilities/parking spaces, reducing traffic congestion, etc.)

[ ]

Employment Status

The questions in the following section focus on your current employment and commuting.

Note: If you are a **full-time student who also works**, you should answer "**full-time student**". You will be asked later in the survey about your employment.

Which of the following statements best describes your current employment status?

- ○ Employed, Full-time
- ○ Employed, Part-time
- ○ Self-employed, work from home
- ○ Self-employed, work away from home

- ○ Full-time Student
- ○ Manage household
- ○ Retired
- ○ Unemployed
- ○ _____ Other

Highest education level to date:

- ○ Less than high school
- ○ High school graduate
- ○ Some college
- ○ 2 year degree
- ○ 4 year degree
- ○ Professional degree
- ○ Doctorate
- ○ _____ Other

How do usually commute to work?

- ○ Walk
- ○ Bike
- ○ Public Transit (Bus/Light rail)
- ○ Drive to work
- ○ Get a ride from someone
- ○ _____ Other

How many jobs do you have?

- ○ 1
- ○ 2
- ○ 3
- ○ 4

On average, how many hours per week do you work? (include all of your jobs)

_____

How long have you had your current job?

- ○ Less than one year
- ○ 1-2 years
- ○ 3-4 years
- ○ 4-5 years
- ○ over 5 years

How far do you travel from your current residence to work?
If you have more than one job, use the job where you work the most hours to answer this question.

- ○ less than 5 miles
- ○ 5-10 miles
- ○ 10-15 miles
- ○ 15-20 miles
- ○ over 20 miles

Could you use public transit to get to work?

- ○ Yes, but it is inconvenient (e.g., too many transfers, infrequent buses, slow).
- ○ Yes, but I can't because service hours don't work with my schedule.
- ○ I don't know which routes go near/to my work.
- ○ No, there is no public transit to my work site.
- ○ _____ Other

Does your employer offer any of the following benefits? Choose all that apply. (Or, If you own a business, do you offer any of these benefits?)

- ☐ Discounted transit pass
- ☐ Free employee parking/Discount parking
- ☐ Carshare membership
- ☐ Carpool
- ☐ _____ Other
- ☐ None of the above

Ridehailing services familiarity

The next questions focus on your use of cellphone, social media, and other types of apps

What type of cell phone do you have?

- ○ I have iPhone cell phone.
- ○ I have an Android cell phone.
- ○ I have a Blackberry cell phone.
- ○ I have a Windows cell phone.
- ○ I don't have a cell phone
- ○ _____ Other

What do you use your cell phone for? (choose all that apply)

- ☐ Social Media (Facebook, Twitter, Snapchat, Instagram, etc.)
- ☐ Shopping (Amazon, Macy's, etc.)
- ☐ Watching Videos (Youtube, Hulu, Netflix, etc.)
- ☐ Navigation (GoogleMaps, Apple Maps, Waze)

- ☐ Communication (texting, calls)
- ☐ Photos/documenting (taking videos/ photos of people, receipts, documents, etc)
- ☐ Listening to music (iTunes, Spotify, Pandora,
- ☐ Other, please explain:

Which of the following apps do you use and how often? (choose all

	Daily	Weekly	Once per month	Rarely	Never	Never heard of
Snapchat	○	○	○	○	○	○
Facebook	○	○	○	○	○	○
Instagram	○	○	○	○	○	○
Twitter	○	○	○	○	○	○
LinkedIn	○	○	○	○	○	○
Pinterest	○	○	○	○	○	○
Tumblr	○	○	○	○	○	○
WhatsApp	○	○	○	○	○	○
WeChat	○	○	○	○	○	○
QQ	○	○	○	○	○	○
Other, please list:	○	○	○	○	○	○

Which of the following navigation apps have you used and how often?

	Daily	Weekly	Once per month	Rarely	Never	Never heard of
Google (Google Maps)	○	○	○	○	○	○
Waze	○	○	○	○	○	○
ReadyNC	○	○	○	○	○	○
NextBus	○	○	○	○	○	○
TransLoc Rider	○	○	○	○	○	○
Other:	○	○	○	○	○	○

Which of the following ride services have you used?

- ☐ Lyft
- ☐ Uber
- ☐ Car2Go
- ☐ _____ Other, please list the company if it is not one of the above:
- ☐ None of the above.

How did you learn about them?

- ○ Friends, family, colleagues
- ○ Website Advertisement
- ○ Print Advertisement (magazine, newspaper, billboard)
- ○ Workplace encouraged us to use.
- ○ _____ Other, please specify:

Which of the following statements best describes how often you use these services?

- ○ Daily
- ○ Weekly
- ○ Monthly
- ○ A few times per year

Thinking back to the last time you used this service, why did you use a ride service?

- ○ To/from Work
- ○ To/from School
- ○ To/from Airport
- ○ To run an errand (e.g. shopping, medical/dental appointment, etc)
- ○ To go to the movies, concert, restaurant/bar, etc.
- ○ _____ Other, please specify:

When was this trip?

- ○ Within the past week
- ○ Within the past 30 days
- ○ Over a month ago

Have you ever used one of these types of services? Check all that apply.

- ☐ Bikeshare
- ☐ Carshare

☐ _____ Other

☐ None of the above.

When did you use this service?

○ Within the past 7 days
○ Within the past 30 days
○ Over a month ago

The following questions are about food delivery services.

Which of the following delivery services have you used? (Check all that apply and/or add any that a not included)

☐ UberEats
☐ DoorDash
☐ GrubHub
☐ InstaCart
☐ _____ Other
☐ Never used a food delivery service

Have you used a food delivery service in the past year?

○ Yes
○ No

When did you last use a food delivery service?

○ Within the past 7 days
○ Within the past 30 days
○ Over 30 days ago

On average, how often do you use a food delivery service?

○ Weekly
○ Monthly
○ A few times per year

Student life and work

What type of college or university do you attend?

- ☐ Public (State, Community college, etc)
- ☐ Private (Duke, Vanderbilt, etc.)
- ☐ For Profit (DeVry, University of Phoenix)

What is your current educational objective?

- ○ High School diploma/GED
- ○ Complete enough credits to transfer to university
- ○ Bachelor's degree (BA, BS)
- ○ Master's degree (MA, MS)
- ○ Professional degree
- ○ Doctorate

When do you plan to graduate or complete your program?

- ○ Within a year
- ○ in 2019
- ○ after 2019
- ○ _____ Other

How do you usually get to school?

- ○ Walk
- ○ Bike
- ○ Take Public Transit
- ○ Drive to school
- ○ Get a ride to school from a friend
- ○ Ride service --Lyft, Uber, etc.
- ○ _____ Other

How far do you travel from your current residence to school?

- ○ less than 5 miles
- ○ 5-10 miles
- ○ over 10 miles to 15 miles
- ○ 15+ to 20 miles
- ○ greater than 20 miles

How do you usually commute to work? (Note: If you have more than one job, choose the mode you use most often)

- ○ Walk
- ○ Bike
- ○ Public Transit (Bus/Light Rail)
- ○ Drive a car
- ○ Get a ride with a friend/coworker
- ○ Ride hailing service: Lyft, Uber, etc
- ○ _____ Other

Could you use public transit to get to school?

- ○ Yes, but it is inconvenient.
- ○ I don't know which routes go near/to my school.
- ○ No, there is no public transit to my school.
- ○ No because the service hours don't work with my school schedule (ex: I could take the bus there but there's no bus service after 8 pm. Or, I can't get to class on time; the buses don't start until 9 a.m.)

Previous to attending your current university/college, how did you usually get around:

- ○ Walked
- ○ Biked
- ○ Took public transit (bus, subway, light rail)
- ○ Drove a car
- ○ Got rides from other people
- ○ _____ Other

Do you work while attending school?

- ○ Yes
- ○ No

How many jobs do you have? (Include only the jobs you work over 5 hours per week.)

How many days a week do you usually work?

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7

On average, how many hours per week do you work? (include all of your jobs, if you have more tha one job)

Student Loan Debt

The following questions are about student loans.

Do you have student loans?
- Yes
- No

How much do you currently owe?
- $1-$5,000
- $5001-10,000
- $10,001-$20,000
- $20,001-$30,000
- $30,001-$40,000
- $40,001-$50,000
- $50,001-$70,000
- $70,001-$100,000
- Over $100,000

How much do you estimate you will owe in student loans by the time you graduate?
- $1-$5,000
- $5001-10,000
- $10,001-$20,000
- $20,001-$30,000
- $30,001-$40,000
- $40,001-$50,000
- $50,001-$70,000
- $70,001-$100,000
- Over $100,000

Are you currently paying your student loans?

- ○ Yes. I am on the standard payment plan.
- ○ Yes. I am on a plan that reduces my monthly loan
- ○ payment. I am not making payments because I have a forbearance.
- ○ I am not making payments because it is still deferred.
- ○ I am not making payments because I cannot afford them.
- ○ [ ] Other, please specify:

Income and marriage

The last two questions are about your marital status and income.

Current Marital Status:

- ○ Married
- ○ Divorced
- ○ Separated
- ○ Single
- ○ Decline to state

Yearly individual income (before taxes):

- ○ Less than $10,000
- ○ $10,000 - $19,999
- ○ $20,000 - $29,999
- ○ $30,000 - $39,999
- ○ $40,000 - $49,999
- ○ $50,000 - $59,999
- ○ $60,000 - $69,999
- ○ $70,000 - $79,999
- ○ $80,000 - $89,999
- ○ $90,000 - $99,999
- ○ $100,000 - $149,999
- ○ More than $150,000
- ○ Decline to state

8.2 AL Survey Instrument

Birmingham Region Travel Diary Survey

Q1 Welcome to the UAB travel diary survey!

Dr. Virginia Sisiopiku (UAB) invites you to be part of a research project that studies technology influence on travel demand and behavior. Your feedback is very important, as it will help UAB researchers to understand and model travel behavior in the Birmingham region.

If you agree to participate, you will be asked to complete a survey about your travel preferences and practices as you travel on a typical weekday in and around Birmingham. The survey should take approximately 10 minutes to complete and your participation is voluntary. Please be assured that your responses will be kept completely confidential and exempt from public disclosure by law. Please note that this survey will be best displayed on a laptop or desktop computer. While you can complete the survey on a mobile device, some features may be less compatible for use on a mobile device.

Your kind assistance in providing input through the completion of this survey is greatly appreciated. If you have questions about the survey or research study, you can contact Dr. Sisiopiku, UAB, Civil, Construction, and Environmental Engineering, Birmingham, AL 35294, or via email at vps@uab.edu.

If you have questions about your rights as a research participant, or concerns or complaints about the research, you may contact the UAB Office of the IRB (OIRB) at 205-934-3789 or toll free at 1-855-860-3789. Regular hours for the OIRB are 8:00 a.m. to 5:00 p.m. CT, Monday through Friday.

By clicking the consent button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

O I consent, begin the study

O I do not consent, I do not wish to participate

Q2 **Home ZIP Code**

_____

Q3 I have used the following in the Birmingham region at least once in the past year:
Check all that apply

- O Transportation Network Companies (Uber, Lyft, etc.)
- O Public Transit
- O Organized ride sharing program
- O Bicycle
- O ⊗None of the above

Q4 Last trip with Transportation Network Companies (Uber, Lyft, etc.)

- O Within the past 7 days
- O Within the past 30 days
- O Within the past two months
- O Within the past 6 months
- O Within the past year

Q5 Reason(s) for using Transportation Network Companies (Uber, Lyft, etc.) *Check all that apply*

O Convenience

O Cheaper than other alternatives

O Destination has little or no parking availability

O Parking at destination is expensive

O Safety/to avoid driving under the influence

O Car is not available

O Transit is not accessible

O Transit is not reliable

O Other modes are not available

O Other reason (fill in) _____

Q6 Trip purpose(s) for using Transportation Network Companies (Uber, Lyft, etc.) *Check all that apply*

O Commute to school/work

O Run an errand (e.g. shopping, medical/dental appointment, etc.)

O Special events where parking is an issue

O Nightlife (or any other activity impairing driving)

O Shopping

O Other (fill in) _____

Q7 Reason(s) for not using Transportation Network Companies (Uber, Lyft, etc.) *Check all that apply*

O Not convenient

O Expensive

O Not available / Area not serviced

O Safety concerns

O Other (fill in) _____

Q8 We care about the quality of our survey data and hope to receive the most accurate measures of the trips of your day. It is important to us that you thoughtfully consider and record each trip of your day over a 24-hour period.

Do you commit to providing your thoughtful and honest answers to recording all the trips of your day over a 24-hour period?

O I will provide my best answers

O I will not provide my best answers

O I cannot promise either way

Q9a Please tell us about your trips during a typical weekday

Considering your trips yesterday or on a typical weekday, indicate every place you visited from the beginning of the day and for a 24-hour period. For the purpose of this survey, the day starts at 12:00 AM (midnight). Please also list walk trips that are 10 minutes or longer.

a. Please provide address (or closest intersection) to your initial location at 12:00AM (midnight)

_____

(Google Map was inserted on the survey)

Q9b **Location Type**

- O Home
- O Work
- O School
- O Ear/ Get take-out
- O Nightlife/ Bar
- O Shopping- Grocery
- O Shopping- Retail
- O Services (e.g. Bank, post office)
- O Pick-up passenger
- O Drop-off passenger

Q10 Please tell us about your trips during a typical weekday.

Considering your trips yesterday or on a typical weekday, indicate every place you visited from the beginning of the day and for a 24-hour period. For the purpose of this survey, the day starts at 12:00 AM (midnight). Please also list walk trips that are 10 minutes or longer.

b. Trip/Place Visited (address or closest intersection)

_____

(Google Map was inserted on the survey)

Q11a Trip Start Time

	hh	mm	AM/PM
Trip Start Time	00 ▾	00 ▾	AM ▾

Q11b Trip End Time

	hh	mm	AM/PM
Trip End Time	00 ▾	00 ▾	AM ▾

Q12 **Trip Purpose**

- O Home
- O Work
- O School
- O Eat/ Get take-out
- O Nightlife/ Bar
- O Shopping- Grocery
- O Shopping- Retail
- O Services (e.g. Bank, post office)
- O Pick-up passenger
- O Drop-off passenger

Q13 Mode

- O Car
- O Carpool/Vanpool
- O Car rental
- O Taxi
- O Uber/Lyft
- O Transit
- O Bike
- O Walk

Q14 Please share your experience with Transportation Network Companies (Uber, Lyft, etc.)

For each location you normally Uber/ Lyft or similar rides, indicate the typical wait time and car availability. Car availability means the number of Uber/ Lyft cars you typically see swarming your location when using the mobile app.

	Wait Time in minutes	Company	Uber/ Lyft Car Availability
	0 - 5 minutes	Uber	1-2

Q15 Is this your last trip of the day (before you go to bed)?

O Yes, this was my last trip

O No, I took another trip

Q16 I would like to see more of the following where I live.

Check all that apply

O Public Transit (bus, light rail)

O Transportation Network Companies services (Uber/ Lyft, etc)

O Bicycle lanes

O Sidewalks

O Parking lots

Q17 Gender at birth

O Male

O Female

Q18 Age

O 18 to 24 years

O 25 to 34 years

O 35 to 44 years

O 45 to 54 years

O 55 to 64 years

O 65 to 74 years

O 75 years and over

Q19 Current employment status

O Full time

O Part-time

O Student

O Stay-at-home parent/caretaker

O Self-Employed

O Retired

O Unemployed

O Other

Q20 Occupation

- O Management, business, science, and arts occupations
- O Service occupations
- O Sales and office occupations
- O Natural resources, construction, and maintenance occupations
- O Production, transportation, and material moving occupations
- O Student
- O Unemployed

Q21 Industry

- O Agriculture, forestry, fishing and hunting, and mining
- O Construction
- O Manufacturing
- O Wholesale trade
- O Retail trade
- O Transportation, warehousing, and utilities
- O Information
- O Finance and insurance, and real estate and rental and leasing
- O Professional, scientific, management, administrative and waste management services
- O Educational services, and health care and social assistance
- O Arts, entertainment, and recreation, and accommodation and food services

O Public administration

O Other services (except public administration)

Q22 Annual Household Income

O Less than $10,000

O $10,000 to $14,999

O $15,000 to $24,999

O $25,000 to $34,999

O $35,000 to $49,999

O $50,000 to $74,999

O $75,000 to $99,999

O $100,000 to $149,000

O $150,000 to $199,999

O $200,000 or more

Q23 Highest Degree

O High school diploma

O Associate degree

O Bachelor's degree

O Master's degree

O Doctorate

Q24 Auto Ownership

O I own a car

O I have regular access to a vehicle that someone else in my household owns

O I do not own or have regular access to a car

Q25 Please provide home address or closest intersection

_____

8.3 A Critical Review on Population Synthesis for Activity- and Agent-Based Transportation Models

Ramada, OE. and Sisiopiku, V.P

Book Chapter published in *Transportation*, May 16th, 2019; IntechOpen; Stefano De Luca and Roberta Di Pace Editors, available at ISBN 978-953-51-6877-5, DOI: 10.5772/intechopen.86307.

Abstract

Traditional four-step transportation planning models fail to capture novel transportation modes such as car/ridesharing. Hence, agent-based models are replacing those traditional models for their scalability, robustness, and capability of simulating non-traditional transportation modes. A crucial step in developing agent-based models is definition of agents, e.g. household and persons. While model developers wish to capture typical workday travel patterns of the entire study population of travelers, such detailed data are unavailable due to privacy concerns, and technical and financial feasibility issues. Hence, modelers opt for population syntheses based on travel diary surveys, land use data, and census data. The most prominent techniques are iterative proportional fitting (IPF), iterative proportional updating (IPU), combinatorial optimization, Markov-based, fitness-based synthesis, and other emerging approaches. Yet, at present, there is no clear guideline on using any of the available techniques. To bridge this gap, this chapter presents a comprehensive synthesis of practice and documents available successful studies.

8.3.1 Introduction

Transportation simulation models are widely used for travel demand forecasting, testing design alternatives, or predicting travel behavior. In 1992, Axhausen and Gärling (1992) developed a comprehensive review of conceptualizations and approaches of activity-based transportation models with special regard to the validity of behavioral assumptions of modeled population. In the course of their review, they concluded that individual travelers and households, rather than aggregates, should be identified and considered. Nevertheless, detailed travel records for individuals have never been easily accessible for several reasons the most important being privacy issues and cost. Hence, individual travel diaries needed to be synthesized from travel surveys, Census data, and publically available records. That process has since been known as population synthesis.

Population synthesizers initially were used as feeder data avenues to travel demand models (Bowman and Rousseau 2006); however, recent shifts towards

activity- and agent-based models brought population synthesizers to the spotlight, as they became determinants to the success or failure of any transportation model of that kind. Fitting is the core of any population synthesizer, with the main focus on fitting disaggregate sample of agents (represented by tabulated demographics of a representative sample of household and individual data) to aggregate constraints (represented by available aggregate data, such as data available from Census). There are several approaches for fitting including iterative proportional fitting (IPF), iterative proportional updating (IPU), combinatorial optimization, Markov-based, fitness-based synthesis, and other emerging approaches (Müller and Axhausen 2011). The following sections present a critical review of each approach in the chronological order by which they were introduced to illustrate the progression and evolution of each approach, with emphasis on notable and well-established efforts.

8.3.2 Iterative Proportional Fitting (IPF) Approach

Iterative Proportional Fitting (IPF) has been first introduced in 1940 by Deming and Stephan (1940). Since then, it became the foundation of population synthesis for transportation models and sometimes referred-to as the Fratar technique (Papacostas and Prevedouros 2001). The most notable realization of the IPF technique is attributed to Beckman et al. (1996) who pioneered population synthesis efforts through their development of a methodology for creating a synthetic baseline population of individuals and households for microscopic activity-based models. Their technique relied on using Census data represented by a Census Standard Tape File and Public Use Microdata Sample (PUMS) for a given Public Use Micro Area (PUMA) of 100,000 individuals with matching variables. In their case, the marginal totals of a multiway table were known and a sample from the population which generated those totals was provided; thus, they applied the IPF technique to develop constrained maximum entropy estimates of the true proportions in the population multiway table. Their rationale was built upon the consensus that IPF estimates maintain the same odds ratios as those in the sample table in absence of any marginal information which was their case. To validate the population synthesis method, they compared demographic characteristics of the synthetic population with those of the true population using variables not involved in the population synthesis. Despite their pioneer effort, Beckman et al. (1996) did not provide an answer to the zero-cell problem in the PUMS; instead, they replaced it by 0.01 and imputed the corresponding household size. Müller and Axhausen (2011) illustrated this as computing a series of tabulations $n_{ij}^{(k)}$, starting with the seed at

$k \coloneqq 0$, thus $n_{ij}^{(0)} \coloneqq n_{ij}$ for all i rows and j columns. Furthermore, they illustrated how that series can be computed as represented by Eq. (2).

$$n_{ij}^{(k+1)} \coloneqq n_{ij}^{(k)} \cdot \begin{cases} r_i \div n_{i.}^{(k)} \\ c_j \div n_{.j}^{(k)} \end{cases} \qquad Eq.(2)$$

where $n_{i.}$ is the row sum

$n_{.j}$ is the column total

r_i is the control total for row i

c_j is the control total for column j

Almost a decade later, Arentze et al. (2007) addressed one of the limitations of the IPF method, that is generating synthetic households when the demographic data describes population in terms of individual counts. Their solution relied on developing a two-step IPF procedure where, first, known marginal distributions of individuals are converted to marginal distributions of households of similar attributes; then second, using the resulting marginal household distributions as constraints of a multiway household counts. Additionally, their approach aimed to assess the relevance of spatial heterogeneity across populations. The Dutch Albatros model was used as a case study and proof of concept. The validation results yielded sample biases in the synthetic population on the dimensions of socioeconomic class, the presence of children, and the availability of transport modes. However, they were able to resolve biases in over- or underrepresentation of groups that were related to age and work status by fitting the relevant tables on these dimensions.

Simultaneous to the efforts of Arentze et al. (2007), Guo and Bhat (2007) addressed the two main drawbacks of IPF approach, namely the zero-cell problem and the inability to control for statistical distributions of both household- and individual-level attributes. Additionally, their study aimed to enhance the scalability and generality of the IPF method as it required code-level changes that are cumbersome and skills that are not typically found within planning agencies, who are the typical users of such approach. The algorithm developed by Guo and Bhat (2007) featured generic data structures and accompanying functions to avoid the zero-cell problem; as well as revisions to the algorithm of Beckman et al. (1996) to allow simultaneous control of both household- and individual-level attributes. That generic algorithm was built upon an object-oriented architecture, and contained eight major steps and a recurring procedure for merging any two contingency tables with common variables. The

proposed approach was used to generate synthetic population for the Dallas-Fort Worth metropolitan area in Texas and the statistical comparison yielded results that were closer to true population than that of Beckman et al. (1996). In addition, Guo and Bhat (2007) concluded that a higher percentage deviation from target size (PDTS) yielded better balance at satisfying the household- and individual-level multiway distributions than lower values of PDTS.

Srinivasan et al. (2008) went a step further and attempted to fine tune existing efforts to accommodate the household- and individual-level controls as well as assess the significance of controlling individual-level attributes. That study was performed in support of Florida Department of Transportation (FDOT) efforts to incorporate socio-demographic attributes within the Florida Standard Urban Transportation Model Structure (FSUTMS). The research was motivated by the need for reduced aggregation errors, ensuring sensitivity to demographic shifts like that of ageing population, and the ability to accommodate population specific transportation modes. That fine-tuning effort mainly aimed to address individual-level attributes of age and gender through the means of a greedy-heuristic data-fitting algorithm that was implemented in the matrix programming language GAUSS. Validation of Srinivasan et al. (2008) algorithm yielded satisfactory distributions of household, size, age, gender, and employment status; however, the distributions for all other variables did not match well.

Given the limited number of attributes that can be synthesized per agent, researchers had to further improve the IPF approach to overcome this limitation. Pritchard and Miller (2009) introduced a method that implements IPF approach with sparse list-based data structure that allows more attributes per agent. Additionally, they used both the conventional Monte Carlo integerization procedure and the conditional Monte Carlo to synthesize a list of individual agents from fitted tables. Despite their thorough efforts, the study of Pritchard and Miller (2009) had only a minor impact on goodness-of-fit, relative to the conventional approach.

Auld and Mohammadian (2010) developed a methodology to improve the basic IPF population synthesis routine in a manner that accounts for multiple levels of analysis units–control variables, which was a limitation to the population synthesizers mentioned hereinabove. Their methodology, named multilevel control, allows population characteristics to be replicated for multilevel synthetic population with one level (such as households) serving as the base level of analysis. After a runtime of sixteen hours, the proposed method was able to synthesize a 7.9 million agent population for Chicago, IL, with an improved fit of the synthesized individual-level characteristics when compared with synthesis

procedures that do not account for individual-level controls. The study concluded that the improved fit comes at no cost to the fit against household-level controls. However, the developed methodology was never experimented as to synthesizing commercial or business related agents.

Lee and Fu (2011) realized that the IPF-based population synthesis approaches, specifically the original synthetic reconstruction method (Beckman et al. 1996) and the complimentary combinatorial optimization method (Williamson et al. 1998), are not generally applicable to all population synthesis scenarios. Based on a comparison by Ryan et al. (2009), Lee and Fu (2011) concluded that combinatorial optimization method produces more accurate demographic information for populations over a small area and that the population synthesis problem should be evaluated from an optimization point of view. In addition, they explored how the estimation of a multiway demographic table can be formulated and solved as a constrained optimization problem in full consideration of both household- and individual-level attributes. Accordingly, that study tackled the inconsistency problem through an approach that is based on the minimum cross-entropy theory. The validity of that model was confirmed through a case study in Singapore, through which results from a 10,641-household study area were superior to conventional IPF approaches. However, Lee and Fu (2011) did not provide a full-scale application which constrains the applicability of their model to theoretical applications only.

Zhu and Ferreira (2014) were intrigued by the inability of the standard IPF algorithm to fit marginal constraints on multiple agent types simultaneously. Hence, they developed a two-stage population synthesizer that utilized IPF on the first stage and then estimation of the spatial pattern of household-level attributes through a second stage IPF-based approach. Their two-stage algorithm consisted of four distinctive steps. The first step involved developing an estimate joint distribution of household- and individual-level attributes. In the second step, households and individuals were drawn from microdata samples. The third step consisted of a conventional IPF with household type and parcel capacity marginal constraints. The fourth and last step included an estimated marginal distribution of other attributes from the fitted model. To validate their approach, Zhu and Ferreira (2014) generated synthetic population for Singapore. Their evaluation approach involved four comparisons, namely: fitting only for households-level constraints; fitting for both household- and individual-level constraints; allocating households to buildings while constraining building capacity; and repeating the previous comparison with income level constrained. Validation results yielded realistic spatial heterogeneity while preserving some of the joint distribution of household and locational characteristics.

Choupani and Mamdoohi (2015) addressed the issue of integerization of IPF results in non-integer values instead of integers; for example, fractions of household- or individual-level attributes for zones. In doing so, they proposed a binary linear programming model for tabular rounding in which the integerized tables totals and marginals perfectly fit to input data obtained from the Census Bureau. The main advantages of using tabular rounding were that it did not bias joint or marginal distributions of socioeconomic attributes of minority demographic groups and it minimized the distortion to the correlations structure of household- and individual-level non-integer tables. Furthermore, the tabular rounding approach outperformed all other eight rounding approaches. In addition, sensitivity analysis of tabular rounding demonstrated that small and large values are equally significant when it comes to integerization. Their findings were confirmed by a comprehensive literature review (Choupani and Mamdoohi 2016) that they performed one year later, which concluded that IPF is the most feasible approach for synthesizing populations for agent- and activity-based transportation models, once integer conversion and zero-cell issues were resolved. In addition, they confirmed that tabular rounding is the most efficient and feasible solution for the integerization issue.

Most recently, in an effort to further enhance the IPF approach, Otani et al. (2018) identified an issue that they named the Modifiable Attribute Cell Problem (MACP) which arises from combining discrete categories of individual-level attributes or due to the contiguous nature of those attributes. The proposed solution to the MACP issue was identified as “the organized cell set” which is the best combination of categories. The procedure to identify the best organized cell set consists of five steps. The first step involves aggregation of the elemental cell set to find several cases of cell organization that generate large cells. The second step involves constructing base-year data using the conventional IPF approach. The third step focuses on forecasting using microscopic simulation. The fourth step involves identifying the statistically acceptable cell value using a Student’s t-test. The fifth and final step involves considering the case with minimum number of cells to be the best cell organization. This method is computationally complex and cannot be performed using conventional optimization algorithms. Yet, it is the sole identifiable solution to the modifiable attribute cell problem.

8.3.3 Iterative Proportional Updating (IPU) Approach

The Iterative Proportional Updating (IPU) approach is a heuristic approach that was developed by Ye et al. (2009) to address the drawbacks of the IPF approach. Specifically, the IPU approach addresses the issue of control for individual-level attributes and joint distributions of personal characteristics. The IPU algorithm matches both household- and individual-level attributes in a computationally efficient manner by iteratively adjusting and reallocating weights among

households of a specific type until both household- and individual-level attributes are matched. Another advantage of the IPU approach is its practicality from the implementation and computational points of view. Eq. (3) represents the mathematical optimization problem as addressed by the IPU approach. In addition, the IPU approach has been generally described in twenty-three computational steps that can be easily coded in most, if not all, programming languages.

$$Minimize \sum_j \left(\frac{\sum_i d_{i,j} w_i - c_j}{c_j}\right)^2 \; or \sum_j \frac{\left(\sum_i d_{i,j} w_i - c_j\right)^2}{c_j} \; or \sum_j \frac{\left|\sum_i d_{i,j} w_i - c_j\right|}{c_j} \qquad Eq.(3)$$

Subject to $w_i \geq 0$

where i denotes a household ($i = 1,2, \dots, n$)

j denotes the constraint or population characteristic of interest ($j = 1,2, \dots, m$)

$d_{i,j}$ represents the frequency of the population characteristic (household/person type j in household i

w_i is the weight attributed to the ith household

c_j is the value of the population characertistic j

Furthermore, Ye et al. (2009) proposed an alternative method to address the zero-cell problem that undermined the IPF practicality. Their method is based on borrowing the prior information for the zero cells from PUMS data for the entire region, where zero cells are not likely to exist as long as the control variables of interest and their categories are defined appropriately. However, that method has the inherent risk of over-representing the demographic group of interest. Despite their attempt to overcome the zero-cell problem, the researchers could not overcome the zero-marginal problem that may result due to non-existence of a certain attribute in households of a certain geographic area; for example, having no low-income households in a certain census block or tract. Furthermore, a review by Müller and Axhausen (2011) pointed to the lack of a theoretical proof of convergence.

8.3.4 Combinatorial Optimization Approach

The Combinatorial Optimization approach was materialized by Abraham et al. (2012) and is a versatile approach capable of matching targets at multiple agent levels for both household- and individual-level attributes. A combinatorial

optimization approach is generally simpler and more direct than IPF. Mostly, it starts by creation of a trial population from the disaggregate sample data, and then the overall level of fit is assessed across all marginal targets. Units from the trial population are swapped with units chosen from the disaggregate samples, and when the measure of fit improves, the swap is made. This is implemented through a proprietary computer program that first identifies a list of units whose aggregate attribute values match a pre-specified set of corresponding target values, then iteratively performs one of three operations, namely adding a unit from the sample to the list, subtracting a unit, or swapping a unit between the sample and the previously identified list. That process is performed on a zone-by-zone level with equal probability of the three actions (i.e. add, subtract, or swap) being considered. The developed algorithm was applied to California and Oregon to synthesize populations for their models. The California application served the California Statewide Travel Demand Model including short- and long-distance travel considering personal and commercial vehicles. The Oregon application served the Oregon Statewide Integrated Model, which included employment synthesis for 34 industries. Both model applications resulted in a near-perfect fit for synthesized populations. Generally, the population synthesis procedure using combinatorial optimization has proven to be fast and flexible with the possibility for application to both households and employment scenarios. However, this algorithm can be further improved by using multi-core and parallel computing techniques.

8.3.5 Markov Process-Based Approaches

As demonstrated, hereinabove, IPF, IPU, and combinatorial optimization approaches rely on cloning attributes that were captured in microdata. In addition, they all share key drawbacks including (a) fitting of a contingency table while ignoring other solutions matching the available data; (b) loss of heterogeneity that has been captured in the microdata due to cloning rather than true population synthesis; (c) dependency on the accuracy of captured data to determine the cloning weights which may replicate inherent inaccuracies; and (d) limited scalability, in terms of the number of attributes of synthesized agents. Hence, Markov process-based approaches were developed to overcome such drawbacks and to offer an approach that truly synthesizes populations instead of cloning them.

The earliest notable effort in this direction was pioneered in 2013 by Farooq et al. (2013) who developed a Markov Chain Monte Carlo (MCMC) simulation-based approach for synthesizing populations. The proposed approach is a computer-based simulation technique that can be used to simulate a dependent sequence of random draws from complicated stochastic models. To synthesize populations that approach uses three sources of data, namely (a) zoning systems

such as census blocks, census tracts, counties, and states; (b) sample of individuals such as the North American PUMS and the European Sample of Anonymized Records (SARs); and (c) cross-classification tables for socioeconomics and demographics like income by age at a certain zoning level. Assuming that in a given spatial region at any point in time there exists a true population, the MCMC simulation-based approach synthesizes that population by drawing the individual attributes from their uniquely joint distribution using the available partial views, while ensuring that the empirical distribution in the synthetic population is as close as possible to the unique actual distribution of that population. The proposed approach was applied to the Swiss census data and results were compared against those developed by a conventional IPF approach. Eq. (4) illustrates the Standardized Root Mean Square Error (SRMSE) based goodness-of-fit tests that were performed on each case and results indicated that MCMC simulation-based synthesis outperformed IPF synthesis while featuring a higher level of heterogeneity.

$$SRMSE = \frac{\left(\sum_{i=1}^{m} \cdots \sum_{j=1}^{n} \left(R_{i\ldots j} - T_{i\ldots j}\right)^2 / N\right)^{1/2}}{\sum_{i=1}^{m} \cdots \sum_{j=1}^{n} T_{i\ldots j} / N} \qquad \text{Eq. (4)}$$

where N is the total number of agents

$R_{i\ldots j}$ is the number of agents with attribute values $i \ldots j$ in the population synthesized

$T_{i\ldots j}$ is the number of agents with attribute values $i \ldots j$ in the actual population

Two years later, in 2015, Casati et al. (2015) proposed an extension of the MCMC simulation-based approach to simultaneously combine both individual- and household-level attributes in a process that was named Hierarchical MCMC. Furthermore, Generalized Raking was introduced as a technique to fit the simulated synthetic population to actual observed control totals. The Hierarchical MCMC is a combination of two methods: (a) an extension of the original MCMC method that allows producing hierarchies of persons grouped into households, and (b) a post-processing method to satisfy known control totals on both the individual- and household-level. That extension aimed to synthesize populations with a hierarchical structure that is based upon ordering the agents living in the same household according to their household roles. The general formulation of the extension is based upon the definition of three groups

of agent types (namely owners, intermediate, and others) running Gibbs Sampling on the three groups, and merging subpopulations. The proposed approach was applied to the 2008 Household Interview Travel Survey of Singapore. The application resulted in realistic synthetic populations, and SRMSE-based test confirmed the goodness-of-fit of synthesized populations and their generated hierarchical structures.

Saadi et al. (2016a) proposed an integrated MCMC approach and profiling-based methods to capture the behavioral complexity and heterogeneity of synthesized agents. This approach used two types of datasets, namely (a) aggregated socio-demographic and transportation-related variables derived from household travel surveys; and (b) individual activity-travel diaries collected from travel diary surveys. The integrated approach consists of six steps that run on those two data types. The first step involves performing a MCMC simulation on the socio-demographic dataset. The second step concerns synthesizing population by a Gibbs Sampling procedure. The third step selects socio-demographics to compare behaviors in the activity-travel patterns. The fourth step uses results from the previous two steps to cluster synthesized populations according to socio-demographics and related activity sequences. The fifth step utilizes multiple sequence alignments to estimate Hidden Markov Models (HMM) estimates. The final step characterizes clusters including mixed socio-economic effects. The integrated approach was applied to the 2010 Belgian Household Daily Travel Survey. Results indicated that the integrated approach effectively captured the behavioral heterogeneity of travelers. In addition, comparisons against IPF and IPU approaches demonstrated that the proposed integrated approach is adequately adapted to meeting the demand for large-scale microsimulation scenarios of urban transportation systems.

Realizing the advantages of Markov process-based approaches, Saadi et al. (2016b) developed an extended HMM-based approach which promised better alternatives than the existing ones. More specifically, the proposed HMM-based approach promised great flexibility and efficiency in terms of data preparation and model training while being able to reproduce the structural configuration of a given population from an unlimited number of micro samples and a marginal distribution. The HMM-based approach considers population synthesis as a variant of the standard decoding problem, at which the state sequences are supposed to be unknown. Accordingly, the maximum likelihood estimators related to the transition states were determined through the Viterbi algorithm. An important advantage of the HMM-based approach is its ability to handle both continuous and discrete variables, which addresses the inherent issue of loss of information due to aggregation of continuous variables like age. Also, the proposed HMM-based approach satisfies the need to discretize continuous

variables to meet the fundamental limitation of Markov process to discrete states. The Statistical and Machine learning toolbox of MATLAB was used to generate sequences from an estimated HMM that were applied to the 2013 Belgian National Household Travel Survey. Three simulations were run to illustrate the HMM-based approach. The first simulation tested the combined effects of scalability and dimensionality. The second simulation compared the HMM-based approach against IPF and the third demonstrated the advantage of the HMM-based approach over IPF using various samples. Simulation results indicated that the proposed HMM-based approach provided accurate results due to its ability to reproduce the marginal distributions and their corresponding multivariate joint distributions with an acceptable error. Furthermore, the HMM-based approach outperformed IPF for small sample sizes while using smaller amount of input data compared to IPF. In addition, simulation results demonstrated that the HMM-based approach can integrate information provided by several data sources to allow good estimates of synthesized population.

8.3.6 Fitness Based Synthesis (FBS) Approach

To address the inability of the IPF approach to deal with multilevel controls, Ma and Srinivasan (2015) developed the fitness-based synthesis (FBS) approach that directly generates a list of households to match several multilevel controls without the need for determining a joint multiway distribution. The FBS approach generally involves selecting a set of households from the seed data, like PUMS, such that tract-level controls are satisfied. The FBS approach starts with an initial set of households that can either be a null set or a random sample from the seed data. Then, population of each census tract is synthesized in an iterative fashion, with one household being either added or removed from the current list in each iteration. Count tables, defined in terms of control attributes, are used to track the number of households of each type that have already been included. The FBS approach implements an adding or removing procedure while swapping is not considered. The main criteria in the FBS approach is reduced sum of squared error for addition F_I^{in} and corresponding error for removal and corresponding error for removal F_{II}^{in} as illustrated by Eq. (5) and (6).

$$F_I^{in} = \sum_{j=1}^{J} \sum_{k=1}^{K_j} \left[\left(R_{jk}^{n-1}\right)^2 - \left(R_{jk}^{n-1} - HT_{jk}^{i}\right)^2 \right] \qquad \text{Eq. (5)}$$

$$F_{II}^{in} = \sum_{j=1}^{J} \sum_{k=1}^{K_j} \left[\left(R_{jk}^{n-1}\right)^2 - \left(R_{jk}^{n-1} + HT_{jk}^{i}\right)^2 \right] \qquad \text{Eq. (6)}$$

subject to $$F_I^{in} + F_{II}^{in} = -2\sum_{j=1}^{J}\sum_{k=1}^{k_j}\left(HT_{jk}^{i}\right)^2$$

where $$R_{jk}^{n-1} = T_{jk} - CT_{jk}^{n-1}$$

j is an index representing the control (and the corresponding count) tables

J is the total number of control (or count) tables

jk is an index representing the different cells in a table

T_{jk} represents the value of cell k in control table j

CT represents the value of cell k in count table j after iteration $n-1$

R_{jk}^{n} is the number of households/persons required to satisfy the target for cell k in control table j after iteration $n-1$

HT is the contribution of the ith household in the seed data to the kth cell in control table j

Three applications of the FBS approach were performed to demonstrate the feasibility of incorporating many controls at multiple levels in the synthesis and increased accuracy of synthesized population. The three applications were performed using the 2000 Census data for 12 census tracts in Florida. The first application involved population synthesis using the IPF approach with only household-level controls. The second application involved population synthesis using the proposed FBS approach with few household- and individual-level controls. The third application also involved population synthesis using the FBS approach but with significantly larger number of controls. Validation for the three applications was performed by comparing the mean absolute error against twenty-two artificial census tracts that were created by randomly selecting subsets of households from the 2000 PUMS. Validation results demonstrated that FBS outperformed IPF and demonstrated efficiency and scalability. In addition, FBS did not require many iterations as it required only one to three times the number of households to be synthesized. In addition, the proposed FBS approach addresses the notorious IPF issues of zero cell problems, computational resources (memory), and non-integers cell value in the joint-distribution tables.

Hafezi and Habib (2015) refined the FBS approach and the refined FBS population synthesizer was examined by three models. The first model used household-level control tables. The second model used individual- and household-level control tables and the third model used weighting individual- and household-level control tables. The models were applied to the province of Nova Scotia in Atlantic Canada using the 2006 Canadian Census and Public Use Microdata File (PUMF). The refined approach was implemented using the sparse matrix technique package in MATLAB that is based on high-level matrix programming for numerical computation. The three models were validated by error percentages and goodness of fit evaluation. Validation results indicated that the refined FBS approach can efficiently obtain a satisfactory result using both individual- and household-level control tables. However, higher homogeneity was achieved within the third model.

8.3.7 Emerging Approaches

Other emerging approaches have been developed in an attempt to replace the IPF approach or to overcome one or more of its drawbacks. Emerging approaches include Bayesian network, annealing algorithms, linear programming, heuristic-based, Copula-based, and entropy maximization approaches. The following paragraphs introduce each of the emerging approaches.

The Bayesian network approach was developed by Sun and Erath (2015) in 2015. The proposed Bayesian network approach is a probabilistic population synthesizer that is intended as an alternative to approximate the inherent joint distribution in a more efficient manner. Using a graphical model, the proposed Bayesian network approach encodes probabilistic relationships, like causality or dependence, among a set of variables. Advantages of Bayesian network models lie in their ability to learn the structure of population systems, particularly when the number of attributes of interest is large using limited amounts of microdata. The Bayesian network approach was founded on the inference of the joint distribution—that is perceiving the population synthesis problem as an inference of a multivariate probability distribution of demographic and socioeconomic household- and individual-level attributes. Like the Markov process-based approaches, the Bayesian network approach does not require marginals as input. In addition, it does not require any conditionals since structure learning and parameter estimation are inherently integrated in the learning model. The performance of the proposed Bayesian network approach was demonstrated through an application to the 2010 Household Interview Travel Survey of Singapore. The Bayesian network approach demonstrated good performance as illustrated by low SRMSE values. It also demonstrated good heterogeneity in synthetic population when size of PUMS is less than 70% of the full population.

The simulated annealing (SA) algorithm was developed by Kim and Lee (2015) to synthesize populations for activity-based models. The proposed SA algorithm is built upon the concepts of thermodynamics and metallurgy and was first introduced as a generic heuristic method for discrete optimization. The Metropolis-Hasting Algorithm was employed to solve the inherent problems of hill climbing and cooling schedule when applying SA to population synthesis. The proposed algorithm consists of seven steps. The first step concerns setting the maximum number of iterations. The second step sets up the total amount of columns and rows in the population and enters observed values of sample distribution. The third step sets up the before-distribution, which is composed by random numbers while satisfying the total amount of restrictive conditions. The fourth step sets up the after-distribution, which is also composed by random numbers that satisfy total amount restrictive conditions. The fifth step involves calculation of absolute error on the before/after distributions as well as observed data. The sixth step involves calculation of selection probability. The seventh and final step iterates steps four through six and ends the calculations when the absolute error (calculated in the fifth step) has the smallest value or satisfies ending conditions. The SA algorithm was implemented using the household travel diary survey from the Korean National Statistics Office. Results from the implementation indicated the need for further verification of the accuracy of this algorithm.

The linear programming (LP) approach was developed by Vovsha et al. (2015) to synthesize populations as part of an activity-based model developed for the Maricopa Association of Governments. The LP approach is an analytical method that balances a list or sample of household weights to meet the controls imposed at some spatial level; typically, for each traffic analysis zone (TAZ). Features of the LP approach include (a) general formulation of convergence of the balancing procedure with imperfect controls; (b) optimized discretization of weights while preserving the best possible match to the controls; and (c) ability to set controls at multiple spatial levels. In addition, the proposed LP approach featured an innovative discretizing method applied for the household weights and integrated with the balancing procedure. While validation of the proposed LP approach is questionable, it still demonstrates reasonable accommodation to various fine resolution spatial levels that are much needed by newer-generation activity- and agent-based models.

The heuristic-based approach was developed by Zhuge et al. (2017) to address two IPF limitations that received less attention from earlier studies. The first limitation stems from the existence of various solutions for one target marginal distribution. The second limitation stems from the optimization nature of population synthesis with the objective function being minimizing the mean

absolute percentage error (MAPE) of control variables. The proposed heuristic-based approach consists of eleven steps arranged in three parts. The first part, including steps one and two, is used to generate the initial household weights. The second part, including steps three through eleven, adjusts the household weights until a stop criterion is met. The third part, including steps ten and eleven, calculates the adjustment steps and adjustment range, which are two fundamental parameters of the approach. The 2007 Household Travel Survey data from Baoding, China were used as a case study. Results indicated that heuristic-based approach cannot perform as well as IPF-based on comparing MAPE values for both approaches.

Most recently, the copula-based approach was proposed by Kao et al. (2018) to address previously identified limitations of IPF approach. Copulas are joint probability distributions with uniform marginal, which are a relatively new statistical tool. Hence, the copula-based approach was designed to preserve marginal distributions and dependence structure between variables. The proposed method was tested for the state of Iowa, and the results were compared with the IPF approach using mean, median, and correlation matrices. The synthesized households resulted in the same local statistics at each block group; but having similar intervariable correlations as described in the PUMS suggest the applicability of the copula-based approach.

Another recent effort to develop an alternative to IPF approaches resulted in the development of entropy maximization-based population synthesizer by Paul et al. (2018) which handles multiple geographies and avoids algorithmic errors. The entropy maximization approach was developed as part of Oregon Department of Transportation (ODOT) effort to utilize an open source population synthesis platform. The approach consists mainly of two algorithms. The first algorithm, namely list balancing, finds weights that match the given marginal control distributions. The second algorithm, namely integerizing, implements a LP-based procedure to covert fractional weights to integers. The proposed entropy maximization-based approach was implemented in Python, and made heavy use of the Pandas and NumPy libraries, which allow for vectorization of operations to reduce overall runtime. Validation results against those of IPF approach were promising and demonstrated reasonable match to controls.

8.3.8 Conclusion

This study presented a critical, comprehensive literature review of population synthesizers starting from the early efforts through the most recent approaches. The review and synthesis indicated that, despite its identified limitations and drawbacks, IPF approach is the most feasible and widely used population synthesizer. All other studies and efforts used it as a reference for comparison

and produced similar or slightly improved results. Evidently, IPF has its drawbacks and limitations. Yet, reviewed literature indicates that there is no single approach that can result in an efficient and accurate population synthesizer. However, an integration of robust methods appears as the most promising approach, like the effort of Fournier et al. (2018) where the limitations of IPF are resolved by combining five methods into an integral framework for population synthesis. Table 1, in the Supplemental Information section, summarizes the advantages and disadvantages of the presented approaches.

Almost three decades old, yet the IPF approach is still being used in state-of-the-art simulation platforms like MATSim. Given that IPF is the most studied approach and the fact that none of the alternatives provided an out-of-the-box solution, IPF is preferred approach by modelers and practitioners. This conclusion is confirmed by the findings of Saadi et al. (2018), who investigated the influence of scalability on the accuracy of different population synthesizers using both fitting and generation-based approaches. Their results revealed that simulation-based approaches are more stable than IPF approaches when the number of attributes increases; however, IPF approaches are less sensitive to changes in sample size.

Overall, this study provides a critical review and comprehensive synthesis of population synthesis approaches that can serve as a valuable reference to future efforts focusing on population synthesis for activity- and agent-based transportation models.

8.3.9 Acknowledgments

This study was performed in support of the project “Technology Influence on Travel Demand and Behaviors” that was sponsored by the United States Department of Transportation Office of the Assistant Secretary for Research and Technology (OST-R) through the Southeastern Transportation Research, Innovation, Development, and Education (STRIDE) Center under Contract № 69A3551747104 with matching funds from the Alabama Department of Transportation (ALDOT).

TABLE 7-1 KEY ADVANTAGES AND DISADVANTAGES OF POPULATION SYNTHESIS APPROACHES

Approach	Advantage(s)	Disadvantage(s)
Iterative Proportional Fitting (IPF)	Synthesized estimates maintain the same odds ratios as those in the sample table Most studied and improved approach with more than 20 years of continuous refinements Widely available with ready to use implementations in several computer programming languages	Does not provide an answer to the zero-cell problem in the Public Use Microdata Sample (PUMS) Unable to control for statistical distributions of both household- and individual-level attributes Limited number of attributes that can be synthesized per agent
Iterative Proportional Updating (IPU)	Addresses the issue of control for individual-level attributes and joint distributions of personal characteristics Computationally efficient Described in twenty-three computational steps that can be easily coded in most programming languages	Cannot overcome the zero-marginal problem that may result due to non-existence of a certain attribute in households of a certain geographic area
Combinatorial Optimization	Generally simpler and more direct than IPF Fast and flexible with the possibility for application to both households and employment scenarios	Implementation is limited to a proprietary computer program Resource-demanding and needs multi-core, parallel computers
Markov Process-Based	Truly synthesizes populations instead of cloning them Meets the demand for large-scale microsimulation scenarios Can handle both continuous and discrete variables	Requires extensive knowledge of computer programming Difficult to trace errors Refinement for specific scenarios or locations requires substantial redevelopment of the computer algorithm
Fitness Based Synthesis (FBS)	No need for determining a joint multiway distribution Addresses the notorious IPF issues of zero cell problems	*Requires extensive knowledge of the sparse matrix technique package* in *MATLAB* that is based on *high-level matrix programming* for numerical computation
Emerging Approaches	Scalable and adaptive Addresses all disadvantages of IPF approach	Requires advanced expertise in Python, and makes heavy use of the *Pandas* and *NumPy* libraries Limited successful applications compared to IPF