

6-2-2017

Exploring Potentials in Mobile Phone GPS Data Collection and Analysis

Eazaz Sadeghvaziri

Florida International University, esade003@fiu.edu

DOI: 10.25148/etd.FIDC001961

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

 Part of the [Transportation Engineering Commons](#)

Recommended Citation

Sadeghvaziri, Eazaz, "Exploring Potentials in Mobile Phone GPS Data Collection and Analysis" (2017). *FIU Electronic Theses and Dissertations*. 3375.

<https://digitalcommons.fiu.edu/etd/3375>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

EXPLORING POTENTIALS IN MOBILE PHONE GPS DATA COLLECTION AND
ANALYSIS

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CIVIL ENGINEERING

by

Eazaz Sadeghvaziri

2017

To: Interim Dean Ranu Jung
College of Engineering and Computing

This dissertation, written by Eazaz Sadeghvaziri, and entitled Exploring Potentials in Mobile Phone GPS Data Collection and Analysis, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Mohammed Hadi

B M Golam Kibria

Yan Xiao

Seung Jae Lee

Xia Jin, Major Professor

Date of Defense: June 2, 2017

The dissertation of Eazaz Sadeghvaziri is approved.

Interim Dean Ranu Jung
College of Engineering and Computing

Andres G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2017

© Copyright 2017 by Eazaz Sadeghvaziri

All rights reserved.

DEDICATION

I dedicate this dissertation to my lovely parents, Hayedeh and Fariborz, and my lovely brothers and sister, Davoud, Faraz, and Forouz for their unconditional love and their endless support and encouragement. Without their kindness, understanding, patience, and most of all love, the completion of this work would not have been possible.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude toward my major advisor, Dr. Xia Jin, for her support, patience, guidance, and understanding at every stage of this dissertation. This dissertation would not have been possible without her support and mentoring.

My deepest appreciation is extended to the committee members Dr. Mohammed Hadi, Dr. Yan Xiao, Dr. B. M. Golam Kibria, and Dr. Seung Jae Lee for serving on my committee and for their invaluable input to my research work. I am thankful for their time and help in reviewing my work.

I experienced a great teamwork environment in the Geographic Information System (GIS) lab. A special credit belongs to my late mates: Hamidreza Asgari, Mohammad Lavasani, Kollol Shams, Seyedmirsajad Mokhtarimousavi, Mario B. Rojas IV, Md Sakoat Hossan, and Fengjiang Hu for being part of my journey. I would also like to thank Mojtaba Mohammadafzali, Arash Moshkforoush, Masood Moghaddami, Arman Sargolzaei, Homa Fartash, and Dr. Mehrdad Beladi for their support.

ABSTRACT OF THE DISSERTATION

Exploring Mobile Phone Data for Better Understanding of Travel Behavior and Mobility
Patterns

by

Eazaz Sadeghvaziri

Florida International University, 2017

Miami, Florida

Professor Xia Jin, Major Professor

In order to support efficient transportation planning decisions, household travel survey data with high levels of accuracy are essential. Due to a number of issues associated with conventional household travel surveys, including high cost, low response rate, trip misreporting, and respondents' self-reporting bias, government and private agencies are desperately searching for alternative data collection methods. Recent advancements in smart phones and Global Positioning System (GPS) technologies present new opportunities to track travelers' trips. Considering the high penetration rate of smartphones, it seems reasonable to use smartphone data as a reliable source of individual travel diary. Many studies have applied GPS-Based data in planning and demand analysis but mobile phone GPS data has not received much attention. The Google Location History (GLH) data provide an opportunity to explore the potential of these data. This research presents a study using GLH data, including the data processing algorithm in deriving travel information and the potential applications in understanding travel patterns. The main goal of this study is to explore the potential of using cell phone GPS data to advance the understanding in mobility and travel behavior. The objectives of the study include: a) assessing the technical

feasibility of using smartphones in transportation planning as a substitute of traditional household survey b) develop algorithms and procedures to derive travel information from smartphones; and c) identify applications in mobility and travel behavior studies that could take advantage of these smartphones GPS data, which would not have been possible with conventional data collection methods.

This research aims to demonstrate how accurate travel information can be collected and analyzed with lower cost using smartphone GPS data and what analysis applications can be made possible with this new data source. Moreover, the framework developed in this study can provide valuable insights for others who are interested in using cell phone data. GLH data are obtained from 45 participants in a two-month period for the study. The results show great promise of using GLH data as a supplement or complement to conventional travel diary data. It shows that GLH provides sufficient high resolution data that can be used to study people's movement without respondent burden, and potentially it can be applied to a large scale study easily. The developed algorithms in this study work well with the data. This study supports that transportation data can be collected with smartphones less expensively and more accurately than by traditional household travel survey. These data provide the opportunity to facilitate the investigation of various issues, such as less frequent long-distance travel, hourly variations in travel behavior, and daily variations in travel behavior.

TABLE OF CONTENTS

CHAPTER	PAGE
1. INTRODUCTION	1
1.1. Background	1
1.2. Research Goal and Objectives	3
1.3. Dissertation Organization	4
2. LITERATURE REVIEW	6
2.1. Passively Collected Data	6
2.1.1. Mobile Network Data	7
2.1.2. Global Positioning System Data	12
2.1.3. Mobile Phone GPS Data	15
2.1.4. Google Location History Data	19
2.2. Data Processing	20
2.2.1. Trip Ends Identification	21
2.2.2. Trip Purpose Identification	22
2.2.3. Trip Mode Identification	23
2.2.4. Location Identification	25
2.3. Data Applications	27
2.3.1. Traffic Flow Characteristics	27
2.3.2. Origin-Destination Matrix	29
2.3.3. Rout Choice Model	30
2.3.4. Daily Variability	32
2.3.5. Other Applications	36
2.4. Summary	38
3. SURVEY DESIGN AND IMPLEMENTATION	42
3.1. Google Location History Data	42
3.2. Survey Design	43
3.3. Instructions for Participation	45
3.4. Survey Recruitment	51
3.5. Sample Composition	53
4. DATA PROCESSING METHODS	62
4.1. Fixed Location Detection	63
4.2. Trip Purpose Detection	64
4.3. Data Validation	72
5. TRAVEL PATTERN ANALYSIS	75
5.1. Time-of-Day Dependence for Long-Distance Trips	76
5.2. Daily Variations for Long-Distance Trips	84
5.3. Demographic Analysis	88

6. CONCLUSIONS AND RECOMMENDATIONS	92
6.1. Conclusions.....	92
6.2. Study Limitations and Future Research	95
REFERENCES	96
VITA.....	108

LIST OF TABLES

TABLE	PAGE
Table 3-1 Descriptive Statistics	61
Table 4-1 Collected Data Using Different Methods.....	74
Table 5-1 An Example of Travel Information in a 24-Hour Timeframe.....	76
Table 5-2 ANOVA Test of Long-Distance Trip Frequency of Different Times of Day..	81
Table 5-3 Bonferroni Test of Long-Distance Frequency of Trips of Different Times of Day	82
Table 5-4 ANOVA Test of Extra-Long Distance Trip Frequency of Different Times of Day	84
Table 5-5 ANOVA Test of Long-Distance Trip Frequency of All Days of Week	85
Table 5-6 ANOVA Test of Long-Distance Trip Frequency of Different Days of Week.	86
Table 5-7 ANOVA Test of Extra-Long Distance Trip Frequency of All Days of Week.	87
Table 5-8 ANOVA Test of Extra-Long Distance Trip Frequency of Different Days of Week.....	88
Table 5-8 Model Summary of Created Models	89
Table 5-9 ANOVA Test of Models	89
Table 5-11 Coefficients of the Models	90

LIST OF FIGURES

FIGURE	PAGE
Figure 3-1 Questionnaire	44
Figure 3-2 Location Activation Process for Different Operation Systems.....	47
Figure 3-3 Downloading the GLH data	47
Figure 3-4 Instructions for uploading the GLH data	49
Figure 3-5 FAQ's Sample.....	50
Figure 3-6 Flyer for Recruitment Announcement	52
Figure 3-7 Message Used in Different Group Pages in Facebook	53
Figure 3-9 Histogram of Education	54
Figure 3-10 Histogram of Gender.....	55
Figure 3-11 Histogram of Ethnicity.....	56
Figure 3-12 Histogram of Marital Status.....	57
Figure 3-13 Histogram of Employment Status.....	58
Figure 3-14 Histogram of Driving License.....	59
Figure 3-15 Histogram of Age.....	59
Figure 3-16 Histogram of Work Type	60
Figure 4-1 GLH Data Trace for a Sample Daily Travel	65
Figure 4-2 Sample Travel (a) Distance-Time (b) Speed-Time (c) 3D Space-time	66
Figure 4-3 Verification Trips (a) Distance-Time (b) Speed-Time (c) 3D Time-space	68
Figure 4-4 GIS Interface of a Sample Location Detection	71
Figure 4-5 GIS Interface of Home Detection	71
Figure 4-6 Locations Detection (a) Home Detection (b) Workplace Detection.....	72
Figure 4-6 Benchmarks of Data Verification Trip.....	73
Figure 5-1 A Typical Long-Distance Trip.....	77

Figure 5-2 Location of Home, Work and other of the Participant.....	77
Figure 5-3 Two-Dimensional Diagram of Longitude and Latitude.....	78
Figure 5-4 Two-Dimensional Diagram of Longitude and Time.....	78
Figure 5-5 Three-Dimensional Diagram of Longitude, Latitude, and Time	79
Figure 5-6 Schematic Example of Time Spent in Different Locations	80
Figure 5-7 Long-Distance Trip Frequency of Different Time of Day.....	81
Figure 5-8 Extra-Long Distance Trip Frequency of Different Times of Day	83
Figure 5-9 Long-Distance Trip Frequency of All Days of Week.....	84
Figure 5-10 Long-Distance Trip Frequency of Different Days of Week	85
Figure 5-11 Extra-Long Distance Trip Frequency of All Days of Week	86
Figure 5-12 Extra-Long Distance Trip Frequency of Different Days of Week.....	87

ABBREVIATIONS AND ACRONYMS

ATB	Activity-Travel Behavior
CA	Call Activities
CDR	Call Detail Records
FDOT	Florida Department of Transportation
GIS	Geographic Information Systems
GPS	Global Positioning System
GLH	Google Location History
IRB	Institutional Review Board
OD	Origin Destination

1. INTRODUCTION

1.1. Background

In order to support transportation planning decisions, one must first have detailed travel information from people and households that can be used to understand travel behavior. Traditionally, these data were obtained from household travel surveys which suffered from low respondent rates, high respondent burden, and significant costs for survey implementation (Wolf et al., 2003; Pearson, 2004; Stopher and Greaves, 2007). Recent research has shown the promise of passively collected data to supplement/complement traditional household travel surveys. The term passively collected data means the data were collected without active engagement with the respondents, which could include GPS data, mobile phone network data, and etc.

In other words, as communication and information technologies advance, emerging data sources via passive collection methods have shown promises in helping transportation professionals better understand and map people's movement through space and time. Traditional travel surveys are plagued by low respondent rates, high respondent burden, and significant costs for survey implementation (Wolf et al., 2003). Passively collected data, such as Global Positioning Systems (GPS) data, mobile network data, and cellphone GPS data, may have the ability to supplement and/or complement traditional household travel surveys and overcome existing issues.

A significant body of research has examined the use of GPS data and their capability to serve as an alternative to household travel surveys (Auld et al., 2009; Stopher et al., 2008; Hossan et al., 2017; Jin et al., 2014). Generally, the GPS device was fixed to

a participant's vehicle or a participant was asked to carry the device daily (Boht and Maat, 2009; Pang et al., 2013; Vacca and Meloni, 2015; Dai et al., 2003). Although these studies indicated that the use of GPS could provide detailed travel trajectory data with sufficient level of accuracy, it had some limitations. The cost of purchasing the GPS units and administering the survey (mail the units to the participants and retrieve them back) could severely limit the size and length of this type of surveys. There is also certain level of respondent burden. For example, a participant could forget to charge the device or leave it at home which would render it useless.

Later studies started to explore the usage of cell phone data, particularly the Call Detail Record (CDR) data (Gonzalez et al., 2008). These data are produced and stored by the network providers whenever a subscriber uses a cellphone, such as making a call, sending a text, or browsing the internet. The CDR log records the time and location when that triggered; this location is calculated via triangulation. Due to the proliferation of cellular phones, a large sample of data could be obtained at minimum cost. Recent data indicated a penetration rate of 128% in developed, and 89% in developing countries (International Telecommunication Union, 2015). However, since the location information is indirectly measured, the CDR data tend to be less accurate than GPS data (Calabrese et al., 2013). These CDR data are measured via triangulation which means that during each usage activity, the signal of the phone is discovered by multiple cellular network towers. Therefore, the accuracy of this data can deteriorate for many reasons including a low number of available towers as well as tower switching to improve overall cellular network performance. In addition, the cellular phone must be in use to trigger recording, which

means no information is provided when the phone is inactive. This could leave significant gaps in the trajectory traces, which would compromise the use of the data.

In an effort to search for other data sources that overcome the current limitations, the goal of this thesis is to explore the potential of cellphone GPS data, which combines high accuracy with a high penetration rate and low respondent burden. These data can be easily obtained from cellular phones with an integrated GPS system. According to Wikipedia smartphone penetration was about 56% in the US in 2013, and is expected to increase every year (Wolf et al., 2001). Cellphone GPS data have an average accuracy of 31 feet, which holds great promises in travel pattern analysis with higher resolution rather than the CDR data with an accuracy of 164-984 feet (Lindsey et al., 2013).

1.2. Research Goal and Objectives

A quick review of the literature indicated that very few studies have focused on cellphone GPS data; most existing studies focused on GPS or CDR data. Nevertheless, these studies still provide valuable insights in terms of data processing algorithms and potential applications. A major challenge of passively collected data is that there is no participant input, so the researchers mainly rely on data mining techniques to derive useful information that describes mobility patterns.

The main goal of this study is to explore the potential of using smartphone data to collect transportation data to advance the understanding in mobility and travel behavior. The objectives of the study include: 1) explore the feasibility of the transportation data collection using the proposed method and develop algorithms and procedures to derive travel information from smartphones; and 2) identify applications in mobility and travel

behavior studies that could take advantage of these cell phone GPS data, which would not have been possible with conventional data collection methods. This research aims to demonstrate how accurate travel information can be collected and analyzed with lower cost using smartphone GPS data and what analysis applications can be made possible with this new data source. Moreover, the framework developed in this study can provide valuable insights for others who are interested in using cell phone data.

1.3. Dissertation Organization

This study is organized as follow: Chapter 1 introduces the study background, explains the problems, and states the goal and objectives.

Chapter 2 presents a literature review regarding different ways of collecting data such as CDR data, GPS, and mobile phone GPS data. In this chapter data processing methods and applications of the data are presented.

Chapter 3 elaborates on the survey design and data collection process for this study.. It describes the survey design and implementation process, including the questionnaire,, respondent recruitment, data collection, and results of the survey.

Chapter 4 describes the data processing algorithm to derive travel information from the GLH data, including trip ends identification, location identification, trip purpose identification, and validation of the results.

Chapter 5 presents further analysis of the travel behavior and mobility patterns based on the travel information derived, which demonstrates the advantages and potentials of incorporating smartphone GPS data into planning analysis.

Chapter 6 summarize the finding of this research and point out future research opportunities.

2. LITERATURE REVIEW

This chapter presents a detailed review of recent studies related to the applications of Global Positioning System (GPS) Data, Call Detail Records (CDR), and mobile phone GPS data; also known as passively collected data. Some of the applications of these data include trip end identification, trip purpose inference, traffic flow characteristics derivation, transportation mode estimation, location prediction, travel variability, Origin-Destination Matrix, Traffic Monitoring Traffic Monitoring, Rout Choice Model, and many more. All of these applications and other will be discussed throughout this section.

2.1. Passively Collected Data

The most important information contained within an individual's travel data is coordinates and timestamps. These three sources which will be explained in this section are Mobile Network, Global Positioning System, and Mobile Phone GPS.

Some studies have been conducted a comprehensive review of trajectory data sources. The most recent and comprehensive of these was Yue et al. (2014). He constructed a review of studies related to travel behavior and categorized them by trajectory data types. In early studies data were mostly collected from household interviews or censuses. With these data researchers focused more on temporal, rather than spatial relationships.

Smart card, first used for automatic fare collection in public transportation, could be used in demand forecasting and travel demand modeling. However, one of the first concern in this field of research has been privacy. The actual and perceived ability of these data to accurately capture human mobility patterns has caused public concern to grow. Some the protection of this type of private data, with special attention to location. By

replacing a participant's user name with a number, records became anonymous. Over the past decades, different types of trajectory data have begun to supplement conventional travel surveys as new collection methods have emerged. This renaissance of technology has also driven a new wave of travel behavior studies. "Big data" has brought opportunities and challenges for travel behavior studies. However, it seems that in future there will be more discoveries in trajectory-based travel behavior research based on "big data".

2.1.1. Mobile Network Data

Mobile network data is a relatively new data source. Other attempts to improve travel surveys saw the incorporation of mobile network data, most commonly the Call Detail Records (CDR). The mobile network is comprised of a series of towers. The spacing and number of the towers, as well as signal strength, will directly affect the accuracy of the data. These data are recorded when a cellphone is transmitting Radiofrequency waves with cellular towers. Simply, data are only recorded when the phone is active, such as during a phone call or sending a message. Similar to GPS, the location of the cellphone is calculated based on the cellphone's distance from surrounding towers. Using this method, it is possible to locate a phone within 50-300 meter accuracy (Federal Communication Commission, 2015).

In other word, this kind of data is recorded from cellphones, irrespective of the service provider. When a cellular device connects to the cellular network, the data is generated. A connection to the network can be established many ways but some of the most common include placing/receiving a call, sending/receiving a message, and connecting to

the internet. Typically, this source collects the device's estimated location (latitude and longitude) and a timestamp for the connection.

As a cellphone moves, the signal will switch to the nearest and strongest towers' signal. However, a phone does not need to move to switch towers. A phone can be observed to switch between towers, or oscillate, due to network policies regarding performance optimization or the proximity to competing cellular tower with equal strength. In respect to travel studies, oscillation can cause the data indicate false movements; a real movement could also be misinterpreted as the oscillation based on the repetitive nature of the movement.

A typical CDR dataset contains the caller ID, timestamp, duration of call or other activity, longitude, and latitude. Other data, such as the call-receiver's ID, may also be available. Due to privacy concerns, these ID's are always anonymized and the formatting varies across carriers.

Due to the proliferation of cellular phones, a large sample of data can be obtained while greatly reducing the cost. Recent data indicated a penetration rate of 120% in developed, and 92% in developing countries (International Telecommunication Union, 2015). It should be stated that caution should be taken, as sampling mobile phone network could introduce bias or over represent participants with multiple phones. Not only is the scale of it enormous, this dataset also eliminates the respondent's burden; most "respondents" are not even aware the data is being recorded. However, CDR data tend to be less accurate than GPS data. Since data are only recorded when the phones are in use, these data are less frequent and irregular. This could leave significant gaps in trajectory

traces, which complicates the application of the data. However, many studies have used these data successfully.

These data have been used for many applications including exploring mobility patterns and activity-travel behavior (ATB), estimating origin-destination (OD) matrices, developing procedures for generating mobile phone datasets, and location detection (Calabrese et al., 2013; Jarv et al., 2014; Iqbal et al., 2014; Chen et al., 2014; Cheng et al., 2012). Mobile network data has two types: cell phone user-based and cell-tower-based data. Cell phone user-based has anonymous user ID, cell tower ID, and information on phone call, data, time, and location. Cell-tower-based data has cell tower traffic information (Yue et al., 2014).

Calabrese et al. (2013) studied individual mobility patterns by using mobile network data. The authors suggested techniques to extract mobility information using cell phone tracing of millions of users within the Boston Metropolitan Area. The dataset included anonymous location estimations collected by AirSage from approximately one million cell phones.

The first step of the authors was to explore if individuals' calling activities were frequent enough and produced sufficient data to monitor users' movement. It was understood that this data would be less accurate than traditional GPS data. Additionally, network performance rules occasionally shift a caller to a further tower in order to maintain overall network performance. Due to this, it was noted that a user traveled several miles in just a few seconds. Although this did occur, it was not frequent enough to discount the entire dataset so the errors were rectified.

Following the error adjustments, trip lengths and home locations were estimated. The length of trips was calculated as the distance between consecutive network connections. Then, mobility measures were calculated and compared to existing mobility measures in order to check the validity of the data. Based on this, it was suggested that the data could serve as a representative source from which individual mobility could be calculated.

Jarv et al. (2013) studied activity-travel behavior (ATB) using call detail records (CDR), a form of mobile network data. Using CDR, their activity location and activity space for 12 months were examined. The object of the study was to reveal the variance in monthly behavior over a longer (twelve-month) study period and examine the factors that affect variability.

The files included information for all outgoing call activates (CA) of users in the network. The CDR data originated from EMT, an Estonian cellphone network operator, which had more than 500,000 active clients. The database had records of all outgoing CAs: text messages, calls, Internet connections, and data service initiated by the owner of the phone. The research focused on the people who were 20 to 64 years old and worked in Tallinn.

The results of the study of Jarv et al. (2013) showed that there was large monthly variance in the spatial distribution of activities. Moreover, ATB could be explained by individual (inter- and intrapersonal) factors and the impact of seasonality. It was understood seasonality had a marginal effect on the number of unique activity locations. Additionally, it was found that seasonality affects daily activities. One issue noted was that the frequency of phone usage had a significant impact on the results. It was concluded that

when using this type of data for human behavior research, the characteristics of individual phone use (calling, texting, or connecting to internet) should be taken into account.

Chen et al. (2014) developed a procedure for generating a simulated cellphone dataset which has the ground truth information. This study also sought to understand if a generated dataset could be representative of real-world mobility patterns. In the process of this, the researchers first had to identify activity locations from cellphone trace data; to do this, they were forced to develop several items. In order to identify clusters, they developed a model-based clustering model which distinguished between activities and travel clusters; they also developed a logistic regression model that did this. To detect the different types of locations visited, they developed a set of behavior-based algorithms. This study defined trajectory as a set of time-stamped locations describing an individual's movement in time, but other studies used drivers' route choice on a highway network. The proposed simulated data (Dataset S) was comprised of two independent datasets: Dataset Ds and Dc. Ds contained household travel survey data and Dc which is a real-world cellphone data. In Dc, there were traces with corresponding time and location information and Ds contained unknown trajectories.

A total of 39,264 respondents were used in Ds; these came from the regional travel survey. Dc was comprised of trajectories with known time and location information. Based on these two sources, Dataset S was generated. It was shown that the simulated data's distribution of the activity locations were more similar to the accurate than existing studies. For example, participants' home and work were identified within 100m with 70% and 65% accuracy, respectively. It was concluded that the proposed methodology could illustrate

spatial and temporal patterns of mobility. The results indicated the possibility of using this generated dataset to replace traditional household travel surveys.

Cheng et al. (2012) analyzed activity patterns using cellular phone data in Shanghai, China. This study proposed a new method of extracting activities of individuals and the recognizing home and workplace. Around 61 billion location data points from 324 individuals were collected over a two-month period. Home was identified as the location of a participant during the night (9:00 pm – 7:00 am), and work as the locations of a participant during the day (9:00 am – 12:00 pm and 2:00 pm – 5:00 pm). Next residents classified into three groups. Participants in Group 1 had their home identified but not work. Group 2 had both identified and the two are less than 5 Km apart. Similar to 2, group 3 had both locations identified but the two were far from each other.

The results showed that mobile network data can be used to analyze human activity patterns in space-time. It was also found that people that lived and work far apart, such as group 2 and 3, tended to have a time-varying activity area near the home or work. One of the major limitations of this study was the issue of privacy. The other was how the researchers defined home and work, which does not apply to every lifestyle or working schedule.

2.1.2. Global Positioning System Data

GPS devices communicate with a network, or constellations of around 30 satellites which orbit Earth at approximately 20,200 km. Although there are 30, a GPS device usually can only see about 12 due to the configuration of the satellites. Each satellite constantly

beams a signal towards Earth, but the position of a device is recorded every 1-4 seconds, depending on the device (United States of America Department of Defense, 2015).

To identify the location of a GPS device, the distance between a satellite and the device is calculated. Connecting to two satellites provides the longitude and latitude. Incorporating a third satellite facilitates the calculation of the altitude; any additional satellites can increase the accuracy. Today, GPS has a vertical accuracy of five meters (or better) and horizontal accuracy of three meters (or better) 95% of the time (United States of America Department of Defense, 2015). The accuracy will be highest when the device is in clear view of the satellites (such as in an open field) and will suffer when obstructed the lowest while the (between high-rise buildings or in a tunnel).

GPS datasets always contain a timestamp, longitude, latitude, altitude, and speed for each data point. It is not uncommon to see the heading and a measure of accuracy such as the number of satellites in communication with the GPS device, included in the dataset (Lindsey et al., 2013).

In the majority of studies, a GPS device was fixed to the vehicle or the participant was asked to carry the device daily. Although most studies related to GPS devices indicated that travel trajectories with a high level of accuracy can be provided, there are some limitations. The first issue is the monetary limitation. The cost of purchasing the devices and administering the survey (mailing and retrieving the units to participants) can severely limit the scale and duration of this type of survey. Moreover, there is certain level of respondent burden. For instance, a participant may forget to charge the device or leave it at home which would render it useless (Stopher et al., 2008). This source is always more accurate than mobile network data (Calabrese et al., 2013).

Studies which need very accurate position data tend to rely on GPS data. For instance, studies that focused on transportation mode detection and developing mobility models for walking have been accomplished with GPS data (Zheng et al., 2008; Lee et al., 2009). Its accuracy also lends it the ability to supplement previous data collection methods, specifically surveys. Studies, such as Yue et al., 2014, proved that GPS data could potentially replace traditional diaries.

Zheng et al. (2008) studied the mobility based on GPS data of people in China. The purpose of the study was to propose an approach to predict participants' transportation mode. The authors collected the data from 65 people over ten months. Features such as heading change rate (HCR), stop rate (SR), and velocity change rate (VCR) were identified. A graph-based post-processing algorithm was proposed to further improve the predictions. The approach had two parts, online inference and offline learning. It was observed that employing new features caused an eight percent improvement, over previous studies, of mode inference accuracy.

Lee et al. (2009) studied the mobility models for walking in South Korea. The main purpose of the research was to present a new mobility model which can produce trajectories for the walking mode. The authors presented the SLAW (Self-similar Least Action Walk) model. The model was generated for applications where people share common interests or are in a single community, such as a university. Developing SLAW relied heavily on GPS recorded human walk trajectories, including 226 daily traces collected in five different outdoor sites from 101 participants. It was concluded that SLAW could generate the unique performance features of different mobile network routing protocols; people within a single community or with similar interest had certain mobility patterns.

2.1.3. Mobile Phone GPS Data

The most recent efforts to improve travel surveys were made possible with the advent of cellphone GPS, or assisted GPS, data. This technology represents a merging of mobile phone network and traditional GPS data. Mobile phone apps have been used in previous transportation studies (Sunio and Schmocker, 2017; Massahi et al., 2016). Similar to mobile network data, cellphone GPS data have the potential for large scale applications and a reduced, but not eliminated, respondent burden. Due to the cellphone GPS chip's energy demand, recording data can reduce battery life. Also, depending on the source, retrieving this information may place a burden on the respondent; this sample could introduce bias. As a hybrid between the previous sources, cellphone GPS data have an accuracy level near that of GPS data (9 meters (Lindsey et al., 2013)) while maintaining the high penetration rate, low cost, and minimal burden similar to the CDR data.

A cellphones position is calculated in nearly the same manner as GPS and mobile network, via triangulation. Data points in this dataset can be recorded via Wi-Fi, GPS satellite, or mobile network; certain phones allow user control over how the data is recorded in an effort to conserve power. It is possible to track the phone even when it is not being used, and it can be tracked without the cellphone signal if the phone is in-view of the satellites.

With the advances of smartphone technology, GPS-like data can be collected from a smartphone rather than a GPS device. For instance, travel behavior characteristics such as the number of activity locations visited and size of individuals' activity space (Ythier et al. 2013), as well as transportation mode (Ansari and Golroo, 2015) can be studied. It was also demonstrated that this new technology can measure position and velocity of vehicles

and used in applications like the traffic monitoring (Herrera et al., 2010). Additionally, using GPS-enabled smartphone, the location of the users can be identified by 100% accuracy. GPS uses so much battery power, however, with new charging technologies, such as wireless chargers, it is not a significant problem to use GPS of smartphones (Hariri et al., 2016; Hariri et al., 2017).

Ythier et al. (2013) studied the influence of social contacts on travel behavior. The purpose of the research was to exploit individual-level data from smartphones to investigate the influence of calling/texting on spatial movement and study the effect of social contact/communication patterns on travel behavior. The authors described how variables can be processed from the data and estimated results using regression analysis. Three concepts were used in the study. Travel behavior which could be explained as travel intensity, which is the number of activity locations visited, and size of individual's activity space, which is the area that most activity of the traveler spends time. Five variables were used as travel behavior characteristics: number of trips, number of activity locations, number of occasional activates, average distance per trip, and maximum distance traveled between home and an activity location.

The study's sample was comprised on 111 participants in Switzerland over three months. In this study, participants carried a N95 mobile phone. Using GPS, longitude, latitude, and altitude with time step of 10 seconds were accessible. However, participants were allowed to turn on/off the device to save battery life which made the data incomplete. Moreover, information of calls, such as duration of outgoing, incoming, and missed calls, as well as outgoing and incoming of text messages, were gathered. It was proved that mobile phone data could be used for understanding the travel behavior, even once they

were collected for billing purpose. It was concluded that the travelers were influenced by their social context. Moreover, the socio-economic characteristics influenced the travel intensity and travel behavior.

Ansari et al. (2015) studied transportation mode choices from a smartphone GPS dataset. The main goal of the study was to identify transportation modes used by the participants via machine learning method. Another purpose was to define attributes for mode classification and determine influential factors. An application was installed on participants' mobile phone to record their data (GPS track). The sample was comprised of traces for 25 males and 10 females and was collected over two week between 6:00 AM until 9:00 PM. All participants were asked to keep a record of their modes to complement the data. The model could classified transportation modes with around 96% accuracy, which was higher than previous studies.

Herrera et al. (2010) studied a traffic monitoring system based on GPS-enabled smartphones. The study was a proof of concept of a system which was able to measure position and velocity accurately. This presented a field experiment nicknamed Mobile Century which includes 100 vehicle that carried a GPS-enable Nokia N95 phone driving loops on a 10-mile stretch in California for 8 hours. Data were collected using Virtual Trip Lines (VTLs). Virtual Trip lines are geographical markers stored in the handset that probabilistically trigger position and speed updates once the headset crosses them. It was possible to compare the velocity measurements collected by both loop detector and VTLs as well as the computation of penetration rate achieved during the day.

As it was very difficult to collect ground truth velocity information, loop detector velocities were used for comparisons against travel times in order to assess the accuracy of

the data. It was understood that VTL measurements exhibit more variability than loop detector measurements, so the study used the 5-min aggregation method. Based on observations, it was suggested that a 2-3% penetration for these mobile devices would be enough to have an accurate measurement of the velocity of the traffic flow.

The field operational test in the initial phase of the development, called the Mobile Millennium, consisted of the free distribution of traffic software such as the software presented in this paper to regular commuters, and the collection of travel time during the months, and would cover North California in its initial phase. The system had mobile sensors to detect the GPS-enabled phones and static (loop detectors) which were expected to provide a more accurate estimation of traffic. Aside from the accuracy of data collection, another advantage of the proposed traffic monitoring system is that the system needs very little installation and maintenance cost. Due to this, the method could potentially be used in developing countries where there is a lack of the resources and monitoring infrastructures.

Yin et al. (2014) evaluated the accuracy of vehicle positioning using GPS-enabled smartphones. They intended to compare the results of different methodologies of identifying vehicles' location. The four data sources which the authors used were professional GPS handset, GPS-enabled smartphone, cellular network positioning, and GPS-enabled smartphone with Geo-fence. In the first scenario GPS-enabled smartphone was compared to a professional GPS handset, Juno. Second scenario, compared cellular network estimated handover points. In the third scenario, GPS-enabled smartphone was compared with Juno. The data were collected from 1:00 PM to 4:00 PM on vehicles commuting on a road in Alberta with twelve Geo-fences. Analyzing the results of the three scenarios, it was

understood that GPS-enabled smartphone could identify the location of user by 100% accuracy; cellular positioning system was lower. Results of a combination of Geo-fence and smartphone positioning was the most accurate.

2.1.4. Google Location History Data

Another way to collect the data of smartphone owners is through Google Location History (GLH). Google's location service uses Wi-Fi, and other signals, to determine location more accurately and often with lower power usage comparing to GPS. GPS helps provide the device with a precise location, hence, it can be used for turn-by-turn navigation and routing. However, it consumes a ton of battery life. The location service circumvents this problem by using signals that the device is usually using in the first location. It hones in on cell sites and Wi-Fi signals to locate the device accurately.

Cellphone GPS data's recording frequency varies based on movement, with less data records when the phone is still. Using Google Location History (GLH) data as an example, while the cellphone was in movement, data were usually recorded every 30-60 seconds; while the phone was still, the recording rate increased to over 1 minute and rarely exceeded 5 minutes. Due to the varying recording interval, it is possible to record greater than 1,000 points per day. GLH data can be accessed as a KML or JSON file. Both files provide information on the timestamp, longitude, and latitude.

The locating method by cell phone is adjustable. For instance, for Android system, choosing the second Location option on the setup screen, which is Power saving, let the smart phone even scan Wi-Fi signals when Wi-Fi is off, and it can be done with the minimal hit to the battery. Every single point on the map indicate where Google used Wi-Fi

Positioning System (WPS) to locate the device. Each time the smartphone is within range of a Wi-Fi access point, it would send its MAC address and SSID to Google's servers. Using GPS (when available) and cell ID data, the Wi-Fi access point can be located, and based on the collected data the history on the map is created (How-To Geek, 2015).

Google operates many location-based services on smartphones. For these services to work best, they need access to the mobile phone owner's location data. Moreover, Google Maps may store places the traveler has been recently and use them to show more relevant search results. Google does not share the individual's location history with others or marketers without individual's permission (Androidcentral, 2015).

2.2. Data Processing

This section discusses an overview of the data processing methods done to the aforementioned data. As indicated by the literature, data processing involves pre-processing, trip end identification, trip purpose identification, and trip mode detection. The main intention of this section is to understand the scope and level of efforts required to use this data.

The first step is pre-processing, usually accomplished through criterion-based data elimination. Speed and location of data points were noted as commonly used criterion to eliminate data. The most common criteria was to remove the data points with unreasonably high speeds (Huntsinger and Ward, 2015; Wang et al., 2013; Sharman and Roorda, 2010). The capabilities of Geographical Information Systems tools (GIS) were also implemented. In one case, the data outside of the study area or outside the desired zones were removed (Zanjani et al., 2015). Moreover, considering time or distance between consecutive data

points has proven to be effective (Berlingiero et al., 2014). In some cases pre-processing involves the removal of trajectories when there was insufficient data (Stopher et al., 2008). Although labor intensive, manual checking was also noted.

At the disaggregate level, considering the distance or time between consecutive data points has proven to be effective (Zanjani et al., 2015). Sometimes pre-processing also involves the removal of data where insufficient data points can be obtained (Bohte and Maat, 2009); although labor intensive, manual checking may also be helpful.

2.2.1. Trip Ends Identification

Various attributes have been used in previous studies to identify trip ends. Many studies used time-constrained rules, however, it tended to work best when the dataset is complete (no signal loss). Trip ends could also be detected when the speed is zero or very low (Nour et al., 2015). Other researchers used the “dwell time” or the amount of time spent at a specific location. This duration was calculated as the difference between departure and arrival time (Berlingiero et al., 2014; Wang et al., 2015; Sadeghavziri et al., 2016). It was also noted that many of these studies used conditional statements whereby both conditions needed to be satisfied to identify a trip end. Some studies also used GIS to help identify trip ends. These cases used data clustering, geo-referencing, and referred to the first/last recorded point of a trajectory (Stopher et al., 2008; Zanjani et al., 2015; Lu et al., 2010).

Some researchers defined activity locations based on duration, frequency, and time-of-day (Fang et al., 2014), as well as point density or other clustering methods in which points were assigned to a cluster based on the relative distance (Zheng et al., 2008). Chen

et al. (2014) showed the potential for the application of a model-based clustering algorithm to identify clusters which were then divided into trip end clusters and travel clustering.

Many studies also used the capabilities of GIS to help identify trip ends, including data clustering (Bohte and Matt, 2009), geo-referencing (Sharman and Roorda, 2010), and retrieving of the first and last recorded points in a trajectory (Lu et al., 2010).

During instances when there is no signal, many studies used the time between consecutive points as a proxy dwell time to detect the trip end. For instance, it is possible to assume that if the location of point A and point B does not change significantly over a time interval when no signal is available, the location is a single trip end. Nonetheless, if the distance between the two points changes significantly during the time interval, most likely that missing data is best represented by a trip rather than a trip end (Schussler and Axhausen, 2009).

Some researchers used GIS (Jarv et al., 2014) and others used criteria to detect Home and Work. For instance, one study (Calabrese et al., 2013) divided the study area into a grid of 500 m by 500 m cells in order to help with home/work locations. The number of nights (6:00 PM - 8:00 AM) spent in each cell was tabulated and the cell with the highest frequency was identified as home.

2.2.2. Trip Purpose Identification

In most studies, trip purpose identification was the most difficult step. Some researchers used GIS to facilitate this process and incorporated the land use patterns. Generally, trip purpose identification was accomplished by three main methods: criterion-based, probabilistic, and machine learning methods.

Generally, the criterion-based method was applied with land use patterns (Bohte et al., 2009; Schonfelder et al., 2003; Pereira et al., 2013). General trip end rules, such as assigning a purpose as shopping if a trip end was located near a known shopping mall, were also used (Wang et al., 2010; Berlingerio et al., 2014; Fang et al., 2014). However, many time, these studies relied on information provided by the participants. Another method was to consider time-of-day and duration were after proximity rules were applied (Cheng et al., 2012; Wang et al., 2013). Some researchers applied a probabilistic approach. For instance, Chen et al. (2014) employed a Multinomial Logit Model in high density areas and a single deterministic matching method in low density areas (Bohte et al., 2009). Another explored the use of Nested Multinomial Logit Models (Wolf et al., 2014, Axhausen et al., 2003) to determine the purpose of the trip based on probabilistic calculations related to the trip distance.

Machine learning, specifically Decision Trees (Ansari and Golroo, 2015; Bohte and Maat, 2009), were employed to detect the purposes of trips. Compared to the first method, probabilistic and machine learning methods were used less due to their complexities. Of all the procedures, trip purpose identification is the area which has the most room for improvement.

2.2.3. Trip Mode Identification

There are different methods which can be used to detect the travel mode, but the most common criteria is speed. Average and maximum speed (Bohte and Maat, 2009; Sadeghvaziri et al., 2016; Alluri et al., 2017), the statistical mode of speed (Nour et al., 2015), a range of average and maximum speeds (Bohte et al., 2008), and the average

acceleration for each mode (Zanjani et al., 2015; Lee et al., 2009) have been applied successfully and applicable to most situation in order to detect the mode of travel.

Another method is using GIS to consider the built environment's characteristics (Hosseinelou et al., 2012). For instance, pedestrians can only walk on links which are accessible to pedestrians. In addition, other researchers used buffer zone in GIS. For instance, Bohte et al. (2009) created buffer zones around rail stations and bus stops. Being near a known point, or inside a buffer zone, allowed researchers to identify rail and bus modes.

There are other methods for travel mode detection. Some researchers developed a probability matrix (Huntsinger and Ward, 2015; Nitshe et al., 2012; Rojas et al., 2016; Abdi Kordan et al., 2014; Sadeghvaziri et al., 2016; Nitshe et al., 2014), and other researchers deployed fuzzy logic methods (Schüssler and Axhausen, 2008). Although these method were effective at identifying walking and cycling modes, they struggle to differentiate between motorized modes (Wolf et al., 2014).

As machine learning method has a high level of accuracy, it is an emerging approach for travel mode detection (Stenneth et al., 2012; Hosseinelou et al., 2012). In this area some common methods are Decision Tree (Reddy et al., 2008), Support Vector Machine (Pereira et al., 2013; Jahangiri and Rakha, 2014; Zhang et al, 2011), Bayesian Network (Moiseeva et al., 2010), Random Forest (Ansari and Golroo, 2015), and Multi-layer Perceptron Neural Network (Das et al., 2015; Gonzalez et al, 2010).

Ansari et al. (2015) detected the travel mode using a machine learning method which was fed smartphone GPS data; 35 participants over 14 days. They also studied attributes for mode classification and determined influential factors. There are two main

methods in mining data and to find the pattern of travels based on GPS traces: procedural and machine learning. In procedural method, based on logical assumptions, the mode of the travelers can be identify. For instance, availability of bicycle at home and low speed of the traveler are hints to interfere that the passenger used his/her bicycle. Machine learning method is very suitable to analyze data through complex approaches and it has been used to identify transportation mode of travelers. As machine learning methods could be used to forecast with high accuracy, it was used in this study. Speed, distance, acceleration, and heading differences were considered as the four main attributes to distinguish different modes. This study also considered new attributes: acceleration, acceleration changes, and heading changes. These attributes were defined based on some logical rules. For instance, acceleration and changes in acceleration would not be very large when walking. Also, walking allows greater freedom to change direction, or heading, than a motorized vehicle. The model was able to detect the travel mode with 96% accuracy. It was concluded that speed was the most dominant attribute in the process of transportation modes classification which was consistent with previous studies.

2.2.4. Location Identification

Chen et al. (2014) developed a set of behavior-based algorithms to identify locations types. They observed three main patterns of cell phone usage that helped identifying activity types at locations. First, as people spend more time at work and home than other places, more sightings were generated at work and home locations. Second, as most individuals depart from home to work at similar times, stay at work for a set period, and go home at similar time, these points could serve as a basis for comparison. Third,

sightings indicated distinct time frames. Most sightings generated at work were during daytime hours and most sightings generated at home were during evening because of individuals natural biological clock is to be at work during daytime hours and at home during evening. Moreover, Cheng et al. (2012) considered the places where the users spent time at night (9:00 PM-7:00 AM) as home and the places where they spent time during the day (9:00 AM-12:00 AM and 2:00 PM-5:00 PM) as work. It is worth mentioning that researchers commonly used timestamps in Unix format; this displays time as a digit which can be easily read by computer and easily converted to be read by humans. For instance, Feb-22-2015 at 13:20:00 is 1424568000 in Unix format. During two months, location traces of 774 individuals were gathered and then analyzed.

Few researchers studied the location prediction accuracy regarding the length of location history. For instance, the purpose of the study of Wang et al. (2014) was to investigate the relation between the length of location history and prediction accuracy for subpopulation which has different level of uncertainty in their trajectories. This study applied entropy S as an indicator of the amount of information input in the predictor. An individual's trajectory with $S = 3$ means ($2^S=2^3=8$) that an individual could be in 1 of 8 possible locations. Entropy was measured based on the temporal-uncorrected entropy with the Equation (1):

$$S_i^{unc} = -\sum_{j=1}^{N_i} P_i(j) \log_2 P_i(j) \quad (1)$$

In this formula $P_i(j)$ is the historical probability that location j was visited by the individual i . Mobile data set which was used in this research was obtained from mobile

phone positioning data. As soon as the mobile phone communicated with the tower, a unique ID number, time, and location was generated. It was found that the level of uncertainty in trajectories was a function of the amount of input information required for correct prediction of the location and low level of uncertainty resulted in a high level of predictability accuracy. For instance, given 100 historical location, for people with low and high level of uncertainty, prediction accuracy would be 80% and 50%, respectively. On the other hand, if a prediction accuracy level over 60% was required, just 10 historical location is enough.

This study found that based on a desired accuracy level, the size of location history could be determined. This implies that limited amounts of input is required for travelers who have a low level of uncertainty in their trajectory. Therefore, it is possible to discard data without compromising the model's prediction accuracy. Moreover, it was shown that for individuals who had high level of uncertainty trajectory, very long location history did not necessarily provide increased accuracy predictions. For example, for individuals who had highest level of uncertainty in their trajectories, accuracy of prediction stabilized between 50% and 60%.

2.3. Data Applications

This section discusses the transportation applications that can be obtained from these data sources.

2.3.1. Traffic Flow Characteristics

Having accurate data, the different flow can characteristics can be obtained. The three main characteristics that can be obtained with accuracy are travel time, travel speed,

and traffic volume (Bar-Gera, 2007; Liu et al., 2008; Massahi et al., 2016; Caseres et al., 2012; Massahi et al., 2017). For instance, Bar-Gera (2007) aimed to examine the performance of a new operational system to measure travel time and traffic speeds from cellular phone service provider data. They compared cellular measurements with those obtained from magnetic loop detectors. Their case study was a 14 km freeway with ten interchanges over two months. The cellular phone-based system recorded observations for about 2% of the total traffic from 10:00 AM to 2:00 PM, and generated 440331 (63%) travel time estimates for 27 sections. Travel times converted to an average section speed simply as the ratio of road length to estimated travel time. The main finding of their study was that there was a good match between two measurement methods and that the cellular phone-based system could be useful.

Liu et al. (2008) aimed to study travel time and speed using cell phone data. Traditionally, traffic condition could be measured using loop detectors. However, new technologies have become available recently that can assess traffic conditions by tracking vehicle trajectories and travel times. Among new technologies, smartphones have received strong interest from the transportation field. AirSage, Inc., has constructed a proprietary system that can track cellphone movement in Minnesota and deliver travel times for most of the urban roads, including both signalized arterial and limited access freeways. The performance of the system was evaluated, including system's accuracy during the peak hours over a 16-day period. Speeds were calculated by dividing travel times (for AirSage's reported and the observed conditions) by the ground truth segment length and converted to miles per hour. The technology produced results with varied accuracies.

Another application of these data is obtaining the traffic volume. Volume is commonly collected by using on-road (fixed) sensors such as cameras and inductive loops. The installation of fixed sensors to cover all roads is impractical, and not economically feasible. Due to this, they are only installed on subset of links. Cellular systems provide alternative methods to detect phones in motion without the cost and coverage limitations related to those infrastructure-based solutions. Caceres et al. (2012) proposed a set of models that could infer the number of vehicles moving from one cell to another by means of anonymous call data of phones. A set of inter-cell boundaries with different traffic background and characteristics were selected for the field test. The results showed that reasonable estimates were achieved by comparison with volume measurements collected by detectors that were located in the same area.

2.3.2. Origin-Destination Matrix

There are many studies which were able to apply the available data to produce OD table. It was proved that with both GPS (Thakur et al., 2014) and CDR (Wang et al., 2013) data, trips can be reproduced accurately. However, most studies used CDR data to infer trips and then estimate large scale OD matrices (Colak et al., 2015; Iqbal et al., 2014; Sharman and Roorda, 2010; Schussler and Axhausen, 2009). Later, Zanjani et al. (2015) were able to estimate Florida's statewide OD truck flows.

By incorporating time-of-day, another study found that it was possible to create OD sample characteristics, mobile OD flow distributions, directional patterns, and spatial analysis; flow analysis for each OD pair was also conducted (Li, 2015). Moreover, to

capture more detailed information, Rokib et al. (2015) used similar CDR data in combination with Foursquare check-in data to reproduce OD matrices.

Iqbal et al. (2014) developed origin-destination matrices using cell phone call data. The authors proposed a methodology for developing OD matrices and traffic counts, using cell phone CDR. The patterns of trips were extracted from cell phone and ground traffic scenario was derived from traffic counts. First, CDR, which had time stamped tower location with caller IDs, were analyzed and then trips were used for generating tower-to-tower OD matrices. Then these are associated with corresponding nodes of the traffic network and converted to node-to-node transit OD matrices. The CDR data was collected from Grameenphone Ltd, and it was consisted of call from 6.9 million users. A total of 13 locations of Dhaka were chosen and video data of them were collected over three days. The study was done in central part of capital of Bangladesh, Dhaka, and CDR from 2.87 million user over one month and traffic data from 13 key locations over three days were used. As CDR is recorded for billing purposes by cell phone companies, this way method is more economic than traditional ways which were based on household surveys. The ODs applied to simulate the traffic between 9:00-12:00 in MITSIMLab, then simulation's results were compared against the observed counts form 4 other locations. As a limitation, the transferability of the models in MITSIMLab was not tested in detail and just key constants of the model were updated to better match the patterns of traffic of Dhaka. The method is effective for places where land use pattern is heterogeneous and there is a limitation of traditional data sources.

2.3.3. Rout Choice Model

Many studies focus on route choices of individuals. For instance, Levinson et al. (2013) investigated the route which drivers choose for a period of 15 days using GPS. The results showed that the participants did not have a single dominant route.

In another study, which was also based on GPS data, the authors identified four types of heterogeneous driver learning and choice evolution patterns (Tawfik and Rakha, 2012). This study had three main findings. First, the observed route choice percentages varied from those derived by using stochastic user equilibrium expectations, but approached specific values. Second, the study identified four types of heterogeneous driver learning and choice evolution patterns. Third, driver and choice situation variables can predict the identified learning patterns.

Dhakar et al. used a combination of GPS and GIS data to develop models for route choice. The results showed statistically significant and reasonable effects of free-flow travel time, intersections, right turns, left turns, and circuitry on the attractiveness of different route alternatives (Dakhar and Srimivasan, 2014).

Casello and Udyukov (2014) also used GPS data to study effective path parameters. They used logit models to determine the relative importance of four statistically significant path parameters including grade, auto speed, length, and the presence of bike lanes. It was concluded that the possibility of generating a relative robust path and mode model which can be included into multimodal travel forecasting models.

In another study, Spissu et al. (2011) used GPS data to explore route choice models. They successfully converted GPS data into routes in order to characterize route choice variability and compare the least-cost route to the actual route. The study analyzed a GPS-based dataset of 679 routes, collected by a personal probe system called the activity

locator over two weeks for a sample of 12 students from University of Cagliari in Italy. It was shown that the higher levels of intra-individual variability were found for discretionary trips. Moreover, higher levels of inter-individual variability, as well as greater deviation from minimum-cost routes, are associated with study and work trips.

2.3.4. Daily Variability

Another common application encountered in the literature review was daily variability (Huff and Hanson, 1988; Hanson et al., 1982). Variability in travel behavior has received little attention in the literature review. The main reason is that most data sets used for analyzing and modeling urban travel comparison information for just a single day for each sampled individual and therefore preclude examination of variability (Pas and Koppleman, 1987; Baqersad et al., 2017). Having multiple days' data available, still the variability in travel pattern is not obvious. Although individuals' travels have a spatial nature, most studies focused on non-spatial aspects, such as daily travel distance, number of daily trip, and daily travel time (Wang et al., 2017). Not having a comprehensive knowledge on variability in spatial dimensions, such as activity locations, lead to a potential gap in current thinking on the relationship between activity-travel behavior and the consumption of urban space (Buling et al., 2008). For example, a traveler may have a same number of trips on two days, however, to completely different sets of activity locations.

Location variability characterize travelers' activity location choice behavior and activity location choice has been proved to be affected by time of day (Kitamura et al., 1998; Baqersad et al., 2017). Therefore, it is important to know whether the travelers'

location variability depends on the time of the day. Time-of-day dependence of location variability is relevant to the concept level of fixity. It is a measure reflecting the extent to which activities are constrained in space. Obviously, those kind of activities which have higher level of fixity are more difficult to relocate and reschedule. Activities with the highest level of fixity are compulsory activities such as school and work (Kitamura and Van, 1987; Shafieifar et al., 2017). Among all activities with different level of fixity that an individual performed, the individuals' locations during time between activities with higher level of fixity have less variability.

Generally, for evaluating location variability, the frequency of repeat visited to activity locations is measured. One study (Kitamura and Van, 1987) studied the individuals' activity locations and the results showed that the trips related to two to four most visited locations accounted for more than 70% of all trips. Based on the result of another study (Buliung et al., 2008), individuals did 72% of all their activities at repeated locations and 28% were carried out at location occurring only once during one-week period. In another study, Song et al. (2010) explored the fraction of time a mobile phone owners spent at their top-visited locations. It was found that individuals spent around 60% of their time at their top two locations. Lu et al. (2013) studied the predictability in human mobility and found that those who visited more than ten locations, spent around 75% of their time at the top two locations, meanwhile, this percentage is about 95% for those who only visited four distinct locations. Therefore, it can be concluded that the individuals' locations choices presented a significant amount of repetition which is same as conclusion of Schonfelder (2001).

In addition, level of fixity is a measure of the extent to which activities are constrained in time and space and deeply rooted in time-geography (Wang et al., 2015). A key idea in time-geography is that travelers' activities and related travel are subject to spatial and temporal constraints.

“Multiday GPS Travel Behavior Data for Travel Analysis” contained four case studies in which multi-day GPS data were analyzed (RGS, Multiday GPS Travel Behavior Data for Travel Analysis, 2015). The first case study explored how drivers' choices were affected by auto network reliability. OD pairs were estimated based on the GPS data and the day-to-day travel time variation was examined. The results indicated that reliability was significantly impacted by trip and household characteristics. Also, there was a lack of definitive proof that there is a direct correlation between reliability and travel frequency at the household level.

The second case study used a 3-day, person-based GPS dataset to detect day-to-day variations in the number, types, and level of dispersion (distance) of destinations visited. Patterns of variability were discovered through the use of latent class cluster analysis. Distance and location were identified as influential over the variation type. The types of variability were also a function of spatial attributes.

In the third case study the authors used the same dataset as the previous case study to analyze day-to-day variations in mode choice. The authors successfully classified individuals into groups based on mode changing and frequency. These groups were connected to specific personal and household characteristics. Relationships between participant characteristics and modality were also considered. Correlation between modality style and other characteristics (household income, number of workers, individual

education level, employment status, and gender) was confirmed, but to different extents. The use of multiple modes was noted for those with greater transit access (greater proximity). The presence of children in a household and age indicated there was a preference for only using one mode, specifically auto. However, statistical significance was not achieved because of the sample size (Baqersad et al., 2017).

The last study used a multi-day, person-based GPS dataset to study the deviation in travel time between the shortest and actual path, as well as the frequency of use of the shortest path for home-to-work trips (RGS, Multiday GPS Travel Behavior Data for Travel Analysis, 2015). It revealed that participants did not make the same home-to-work auto trip frequently over multiple days. It suggested that multi-day studies require a large sample size. It was also noted that most participants did not use the shortest path, which could be due to trip circuitry, number of turns, and age of the driver.

Several CDR studies focused on human mobility patterns. One presented a technique to use “social sensing” to gauge human mobility (Gkiotsalitis et al., 2014). Another study compared individual mobility and vehicular mobility to understand daily mobility; the major influences on total trip lengths were observed to be job accessibility and distance to non-work destinations (Calabrese et al., 2013). An optimized network design model for public transit which decreased system-wide travel times was also proposed (Berlingerio et al., 2014). Entropy was used as a measure of location variability to explore the effect of time-of-day on travel behaviors (Wang et al., 2015). From this, the authors determined that time-of-day effects location variability and that location variability is smallest during the morning hours.

2.3.5. Other Applications

Aside from the previous groups, other applications were noted. One study used anonymous GPS data to construct activity-travel pattern characteristics which were combined with land use data to estimate various models of demographic characteristics (Auld et al., 2015). The proposed methodology successfully identified several characteristics (work-status, education level, age, possession of license, and presence of children), but struggled with others (gender, household size, and number of vehicles). Based on this, the results were generally positive and demonstrated the ability to reconstruct some socio-economic demographic data.

CDR data (Calabrese et al., 2013) and GPS data (Liang et al., 2012) were also used to explore human mobility patterns. CDR data was used to demonstrate that variations in individual and vehicular mobility are due mainly to accessibility rather than population density and land use (Calabrese et al., 2013). A different study employed GPS taxi trajectories to serve as a proxy for individuals (Liang et al., 2012). It discovered that, unlike most models, the traveling distance and elapsed time of these data were best fit by an exponential distribution, and human mobility tends to be sporadic.

One year of large scale CDR data was used (Jarv et al., 2014) in an attempt to study human activity-travel behavior with respect to temporal trends. The results indicated that monthly variation in unique activity locations displayed seasonal trends, while the spatial distribution varied greatly. It also revealed that inter-personal and intra-personal factors were more influential than seasonal impacts. Also daily variation of activity locations

remained relatively constant throughout the study, while participants' activity more than doubled during summer months.

Another application encountered in the literature was traffic monitoring. Barth et al. (1996) explored the possibility of estimating emissions and fuel consumption of different transportation mode using GPS data and embedded traffic road sensors (Barth et al., 1996). Venter et al. used GPS data of private and commercial vehicles to better understand the fuel consumption and emissions by link and time-of-day (Venter and Joubert, 2013). In addition, a very frequented topic in this area is generating traffic performance measure (Bachman et al., 2012; Liao, 2014; Li et al., 2015; Yin et al., 2015; Ma et al., 2011).

Another researches aimed to develop flow and density models. Ranjitkar et al. (2004) tested six microscopic traffic flow models, using GPS data. They used the genetic algorithm-based approach to estimate model parameters for two different cases: headway and speed data. It is noticeable that all models perform better with speed rather than headway data.

Other researches focused on real-time monitoring. For instance, Demissie et al. (2013) attempted to use CDR data and traffic counts. Nonetheless, the results were not very accurate. In a similar study (Wang et al., 2015), mobile phone GPS data was employed to investigate traffic conditions in real-time. In terms of the privacy issue, this research used virtual trip lines which is a technology that just transmit data specific locations. The study concluded that 2-3% penetration rate of GPS-enable smart phone is sufficient to duplicate the results.

2.4. Summary

In this chapter, different methods for collecting the data were reviewed. Most researchers did their studies using CDR rather than GPS data and very few studies have been done with smartphone GPS data. The GPS data has higher level of accuracy than CDR data. Google can record the trajectory of individuals with higher level of accuracy, easier and cheaper. Google Location History has not been used for data collection in transportation studies so far. The first thing that can be identified is visited location. In other words, where and when the individuals visits different places. In next step, it was explained that how the visited locations' usage can be identified. Among all visited locations, the main effort was for identifying home and work place. Some researchers identified the locations based on logical criteria, while others did it using GIS.

Moreover, according to coordinates of data point, travel time and consequently travel speed were obtained. There were studies that investigate transportation mode detection based on the travel speed and GIS. Some researchers went one step further and explored the location variability of individuals and explored that in which part of the day; morning, afternoon or evening, visited location has more variability.

In terms of the limitations of the other data sources, it is noticeable that the CDR data cannot provide very accurate data. Sometimes, the cell phone owner is switching from one tower to another one, or there might be not enough towers to cover the whole network. Sometimes there is no signal at all, therefore, the cell phone cannot be tracked. On the other

hand, the advantage of the CDR is that almost everyone has a cell phone and carry it with themselves wherever they go.

The GPS device also has many limitations. For instance, participants may forget to charge the battery of the device and at the middle of the day it may turn off and no data would be collected. Also, some participant may forget to take the device with themselves and left it at home once they are leaving home at morning. Hence, there would be no data for the whole day. Another limitation of GPS device is the cost of purchasing and its maintenance. The researchers need to buy a GPS device for each single participant and need to take care of all maintenance issues. On the other hand, the advantage of the GPS is its high level of accuracy rather than CDR. The CDR data can be generated once the user connects to the network, which might be couple of hours gaps between two consecutive records, while GPS device records the data more frequently.

There are few mobile apps that can record the trajectory of the mobile phone owner. Collecting the trajectory data based on this method has its own issues. The main issue is that the GPS of the smartphone should be activated, otherwise the data cannot be recorded. During the time the GPS is activated the usage of the battery increased rapidly and results to unsatisfactory situation for participants and in some cases losing the trajectory data due to shutting down the cell phone. Another problem is that as all smartphones do not have the same operating system, including Android and iOS (originally iPhone OS), one specific app has different abilities for smartphones with different operating systems.

More and more studies have explored the potentials of supplementing traditional travel surveys with passively collected data, such as GPS and mobile data, this may be the beginning of a paradigm shift. Moving forward it will be possible to tap into the full

potential of these new data sources and supplement them with minor surveys. This approach has the potential to decrease respondent burden and cost while simultaneously improving data quality and prediction accuracy.

Irrespective of its accuracy, traditional GPS has proven to be less useful at a large scale due to its cost. Mobile network data is a much cheaper alternative to traditional GPS. This source can provide anonymous information for millions of users. Scale alone can be misleading though due to the fact that data is only recorded while the phone is in use. When this is considered, along with laws in certain parts of the United States concerning phone usage while driving, the potential of this data source may also be limited. Although cellphone GPS data has no direct cost associated with it, there may be some associated respondent burden for participants who lack computer literacy. However, this can easily be overcome by providing clear instructions, as was noted by the authors. Cellphone GPS data have the potential to provide nearly the same accuracy as traditional GPS with the large scale, low cost, high penetration rate, and low respondent burden of mobile network data.

Given the “passive” nature of these data, a considerable amount of efforts is needed to derive useful trip information out of those recorded traces, such as the trip ends, the modes, the location types, etc. In this regard, many studies have employed various data processing methods to transform the data points into a meaningful representation of people’s movement. Based on this information many applications have become feasible, including OD estimation, traffic monitoring, and understanding spatio-temporal human mobility patterns. These data sources have the level of detail that could afford researchers the opportunity to create real-time representations of congestion, and therefore emissions,

throughout the entire transportation network. These data hold the promise to help produce more accurate transportation measures, more representative models of human behavior, and usher in a new era for activity-based modeling.

Although very detailed, these data have a few limitations. The most impactful of these is the lack of demographic information, which is critical in travel studies and demand analysis. However, some studies have showed the potential to derive demographic information with the aid of supplement data (such as land use, Census data, etc). This area probably requires further exploration. Another limitation is the potential sample bias of these data, as not all people carry smart phones while others may possess multiple. This issue may diminish as technology advances and smartphones get more common, however, studies using these data still need to aware of the sampling issue.

All in all, all of the previous methods have their own advantages and disadvantages. In this study we use Google Location history (GLH) data. It has a couple advantages which none of the previous methods has them all together. Almost everyone has a cell phone and on the other hand, result of GPS is more accurate than CDR, therefore using GPS data of smartphone can be the best choice, however, GPS uses the battery of the smartphone a lot and the app has different abilities on different smartphones. GLH does not need the owner to activate their GPS device, hence, it does not use the battery so much. Moreover, the google appearance and abilities are the same for all smartphone since it is not related to the operating system of the phone. Additionally, it has no cost to install and activate it on smartphones.

3. SURVEY DESIGN AND IMPLEMENTATION

This chapter describes the approach for survey design and data collection. First GLH data were explored to understand the characteristics of the data and the procedures involved to obtain the data. Based on the understanding of the GLH application, instructions were prepared to provide detailed information for the respondents to ensure proper initiation of the application and management of the data. Then this chapter discusses participant recruitment strategies and results of the survey.

3.1. Google Location History Data

Many types of data, including travel information, can be collected using cellphone apps on smartphones or simply the smartphone itself. Travel information, as it relates to this study, would be GPS data that is stored in the phone. There are many different apps that can facilitate the retrieval of GPS data from smartphones. For instance, Ansari and Golroo (2015) asked the participants to install a specific mobile app on their smartphone to record their data. They later used that data to detect the mode of transportation.

Google Location History (GLH) was selected for this study because it is free and easy to access, and sufficiently accurate. Also, recording the data does not require installing additional apps, provided that the participant has a Google account; one simply must allow Google to track their locations through the Google application. GLH provides the data in two formats that can be accessed easily (.KML and .JSON). Both of these formats contain the same basic information (timestamp, longitude, and latitude) and record

at the same rate. However, the .JSON file provides some additional information (such as the accuracy level, and activity by mode). The .JSON file is also a comprehensive file, meaning that it contains all points every recorded by Google. On the other hand, the .XML file can only contain up to one month's worth of data.

GLH allows Google to store a history of an individual's location data. This data is collected from all devices where the individual's Google account is logged-in and have location services enabled. In terms of smartphones, GLH uses all available means including Wi-Fi Positioning System (WPS), GPS satellite, and mobile networks to locate the device. This allows for very accurate data, and improves Google's apps, but tends to consume the device's battery. For instance, GLH lets "Google Now" show the smartphone owner a card with their commute time before they leave the house (Support Google, 2015). Moreover, GLH allows the user to control how and when the data is recorded. It is not clear exactly how frequently GLH records data, but it seems that there is a relationship between the recording frequency and the device's movement. When the phone is still, data are recorded less frequently and ranges from 1 to 5 minutes; when moving, data are usually recorded every 30-60 seconds. It is common to record more than 1,000 data points per day.

3.2. Survey Design

A questionnaire was created to obtain socioeconomic information from the participants, as shown in Figure 3-1. The questionnaire had two sections; the first was required, but the second part was optional. In the first part information including education, gender, ethnicity, marital status, employment status, having driving license, and age was collected. In the second part non-essential information was collected; this included Hours

of working, if the participant is a full time worker, days of working, and address of home and office.

Education:			
<input type="checkbox"/> No schooling completed	<input type="checkbox"/> Nursery school to 8th grade	<input type="checkbox"/> Some high school, no diploma	<input type="checkbox"/> High school graduate, diploma or the equivalent
<input type="checkbox"/> Some college credit, no degree	<input type="checkbox"/> Trade/technical/vocational training	<input type="checkbox"/> Associate degree	<input type="checkbox"/> Bachelor's degree
<input type="checkbox"/> Master's degree	<input type="checkbox"/> Professional degree	<input type="checkbox"/> Doctorate degree	
Gender:			
<input type="checkbox"/> Female	<input type="checkbox"/> Male	<input type="checkbox"/> Other	
Ethnicity:			
<input type="checkbox"/> White	<input type="checkbox"/> Hispanic or Latino	<input type="checkbox"/> Black or African American	<input type="checkbox"/> Native American or American Indian
<input type="checkbox"/> Asian	<input type="checkbox"/> Pacific Islander	<input type="checkbox"/> Other	
Marital Status:			
<input type="checkbox"/> Single, never married	<input type="checkbox"/> Married or domestic partnership	<input type="checkbox"/> Widowed	<input type="checkbox"/> Divorced
<input type="checkbox"/> Separated			
Employment Status:			
<input type="checkbox"/> Employed for wages	<input type="checkbox"/> Self-employed	<input type="checkbox"/> Out of work and looking for work	<input type="checkbox"/> Out of work but not currently looking for work
<input type="checkbox"/> A homemaker	<input type="checkbox"/> A student	<input type="checkbox"/> Military	<input type="checkbox"/> Retired
Driving License:			
<input type="checkbox"/> Yes	<input type="checkbox"/> No	<input type="checkbox"/> Other (please specify)	
Age:			
<input type="checkbox"/> Under 12	<input type="checkbox"/> 12-17 years	<input type="checkbox"/> 18-24 years	<input type="checkbox"/> 25-34 years
<input type="checkbox"/> 35-44 years	<input type="checkbox"/> 45-54 years	<input type="checkbox"/> 55-64 years	<input type="checkbox"/> 65-74 years
<input type="checkbox"/> 75 years or older			
<u>(This part is arbitrary)</u>			
How many hours do you work per day?			
Are you a full time worker?			
If you are a part time worker, please specify which days of the week you work.			
Please write address of your home and office:			

Figure 3-1 Questionnaire

3.3. Instructions for Participation

In order to minimize participation burden and ensure a smooth experience for the participants, detailed instructions were prepared to guide the participants, in terms of how to activate the application, maintain and upload the data. The instructions were comprised of four parts: activation of location services, downloading participant data, anonymously uploading participant data, and completing the questionnaire. This set of instructions was emailed to each participant and as well others showed interest in participating in the study.

The method to enable GLH to track the phone varies slightly based on the settings of the particular device, but it generally involves few easy steps. An instruction was created to help the participants go through the steps including how to upload the data anonymously. All participants' operating systems were Android or iOS, as such the activation methods differed slightly. Due to this, separate instructions were compiled in order to serve both operating systems. All participants that desired to self-activate their smartphones were provided the document shown in Figure 3-2. Participants were asked to activate the GLH of their smartphone. Once GLH is enabled, there is minimal user burden; as long as the phone is on, location information will be recorded. All participants were checked frequently during the process to ensure that their data were recorded properly. Strategies for incentives to enhance participation rate were also sought, such as gift card or participation in lottery, etc.

How to Turn On the Google location history?

1- Android

1-1- First Method

1. From your device's apps menu, open Google Settings app  or the Settings menu .
2. Touch **Location** > **Google Location History** for the account you want to change. > ON and This Device.
3. Turn Location History on or off for your account or devices:
 - o **Google Account:** Move the top switch to turn Location History on or off for all devices associated with your Google Account.
 - o **Device only:** Move the switch next to a listed device to turn Location History on or off just for that device.

1-2- Second Method

1. Open the Google app .
2. In the top left corner of the home screen, touch the sign which is 3 lines upon each other. >  setting.
3. Account & privacy > Google account history > Google location history > On

2- iOS

2-1- First Method

1. Open the Settings app on your iPhone or iPad .
2. Touch **Privacy** > **Location Services**.
3. Make sure the **Location Services** switch is on.
Touch **Google** > **Always**.

2-2- Second Method

1. Touch  to open the Google+ application.
2. Touch  Menu >  Locations >  Settings.
3. Touch **Location Reporting & History**.
To **turn on** Location Reporting, touch Location Reporting > **Report from this device**.

2-3- Third Method

1. Open the Google app .
2. In the top left corner of the home screen, touch your photo. (If you're looking at Google Now cards, scroll up to the top of the page.)
3. Touch **Privacy** > **Location**.
4. Turn **Location Reporting** on (blue) or off (grey).
5. Touch **Done**.

3- General

1. Visit maps.google.com/locationhistory.
2. Sign in to the Google Account you want associated with your location.
3. Click the  gear icon, then select **History settings**.
4. Choose **Enable** or **Disable**, then **Save**.

Figure 3-2 Location Activation Process for Different Operation Systems

The instruction to download their data from GLH is shown in Figure 3-3.

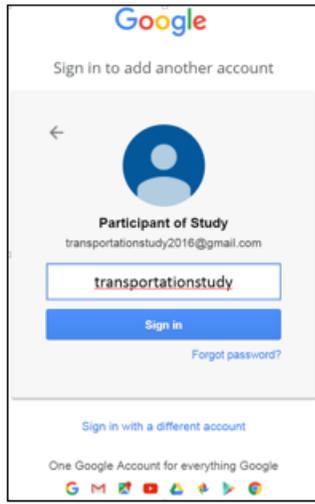
B- Downloading the Data
For downloading the data go to <https://maps.google.com/locationhistory> and on right bottom of the page click on , and choose Download a copy of all your data. Then, in Location History choose KML format and then Next, then Create Archive, and finally Download.

Figure 3-3 Downloading the GLH data

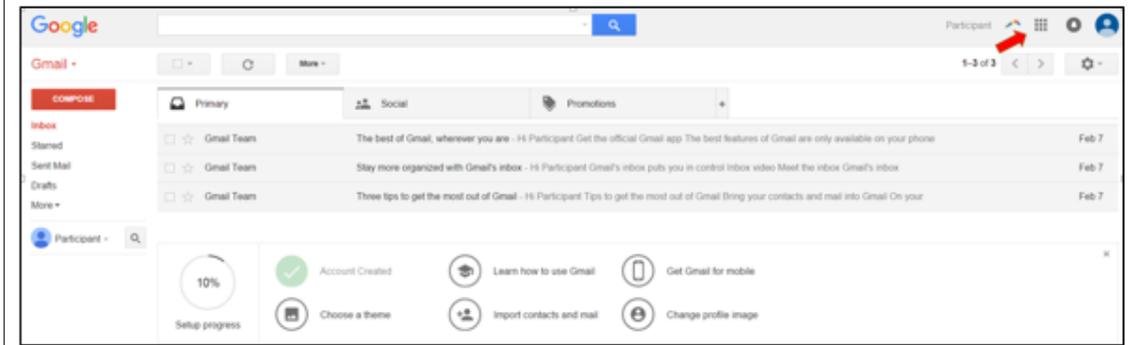
The researchers believed that one of the participants' main concerns would be privacy. Participants were assured that the analysis would be anonymous. As privacy was such a sensitive issue, each participant could create and send their data through an anonymous email address. As no real name was needed for creating the emails, once the participants sent their data via the anonymous emails to the researcher, the identity of the participants could not be detected. Therefore, the probability of detecting which data belonged to which participants would be zero. Moreover, in order to expedite the processes of anonymously data uploading, an anonymous Gmail account was created for each participant to upload his/her data. The process of uploading the GLH data is shown in Figure 3-4.

C- Uploading the Data anonymously

Step 1: Go to the following gmail account:



Step 2: Click on Google App, shown in the following figure



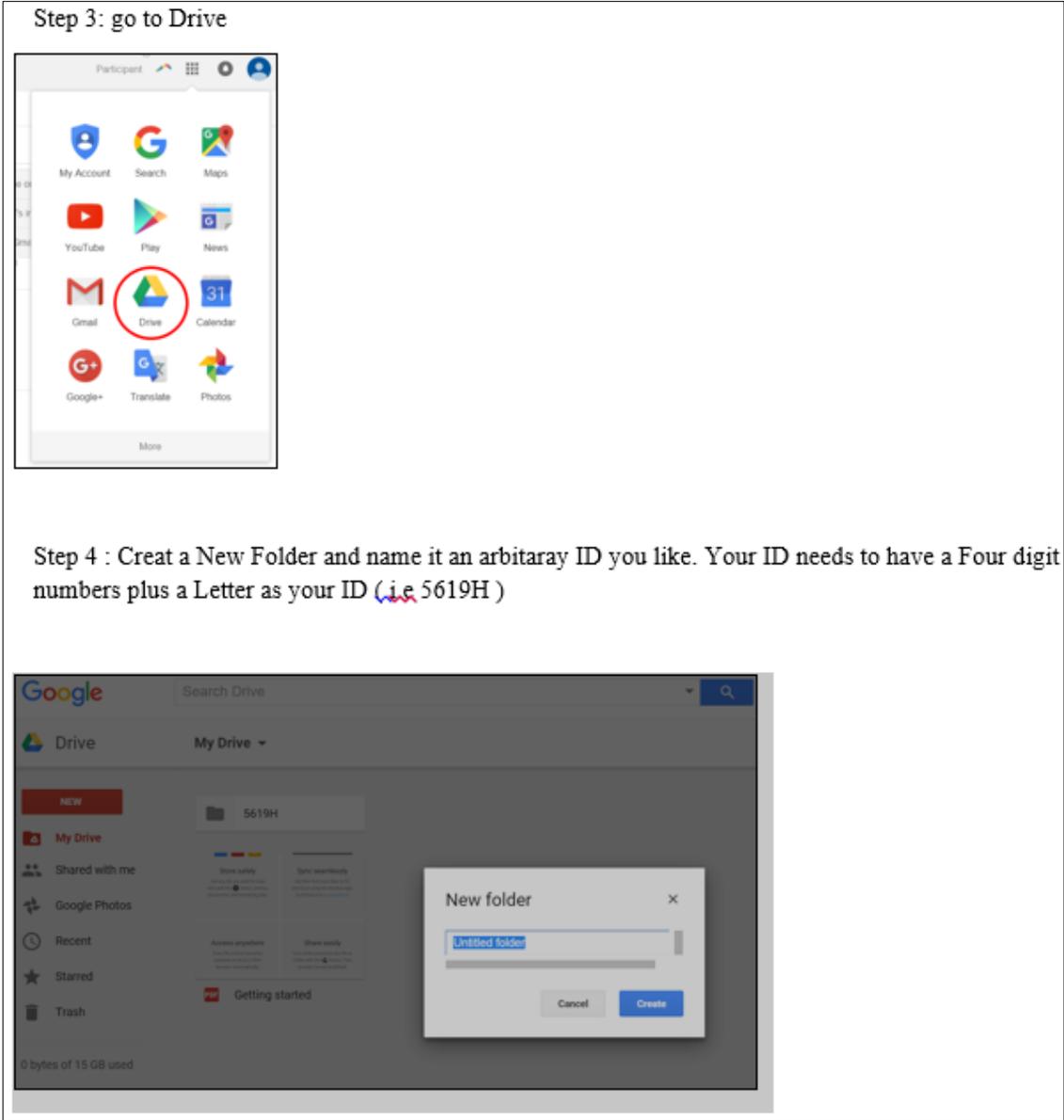


Figure 3-4 Instructions for uploading the GLH data

Frequently Asked Questions (FAQs) and their responses were compiled in anticipation of the participants having similar questions regarding the study. This was done to expedite the process of responding to participants and is shown in Figure 3-3.

FAQ's

Why do you need my travel data? Exactly how will the information be used?

- Your travel data is needed to fuel our research. From all participants' data, we will be able to determine trends in travel behavior which will help improve transportation planning.

What information will you get from the data?

- Each participant will record data using Google Location History. From the Google Location History we will be able to see latitude and longitude.

Is there a minimum amount of data I will be required to provide? If so, how much?

- In order to participate, we ask that you agree to provide us with at least 2 months' worth of daily travel data. We require Google Location History data for the entire duration of participation.

How will I be tracked? Do I need to download an app?

- To be tracked, a participant must enable Google Location History

Does this mean I have to keep my location services on all the time?

- Yes, you must keep you location services on during the time of participation.

How will I know if the data is being recorded properly?

- You can verify that Google Location History is working by following our walkthrough.

When do I have to record the data; are there certain times and days?

- Google Location History will record as long as your location services (GPS) is enabled.

Will my information be secured? Who will have access to my data?

- The information will be totally secure and only the researchers will have access to the data.

Figure 3-5 FAQ's Sample

A reminder email was sent out to each participant after one week, and again after two weeks. In an attempt to diversify the sample. Participants were encouraged to invite their family members and friends to take part in the study.

In cases where the participant did not have privacy concerns, the questionnaires were collected in-person and their data was emailed directly to the author.

3.4. Survey Recruitment

Multiple means were implemented to recruit faculty, staff and students to participate in the survey, which included distribution of recruitment flyers, promotion through classes, and advertisement through social media.

A recruitment flyer (Figure 3-6) was created. These flyers were printed and placed in Florida International University's (FIU) Modesto Maidique Campus and Engineering Campus. In total, more than 100 flyers were distributed throughout FIU. It is worth noting that the flyers were posted in several departments' buildings including engineering, statistics, modern language, history, psychology, economics, woman study, chemistry, physics, and several libraries.



Do you like to participate in a PhD transportation engineering research???

Study's Goal To better understand travel behavior and trends in order to improve transportation planning.

Participation: Daily travel information (**anonymity guaranteed**)
All data will be recorded via the smartphone.

If you are interested, just activate location of your Google app
and send us an email and say "I'm in!"

Or

Contact Us NOW!

Figure 3-6 Flyer for Recruitment Announcement

The author also prepared messages to share in several Facebook groups. The message (Figure 3-7) was posted in 5 Facebook groups totaling an audience of 12,543. The targeted groups were open to all FIU students, faculty, and staff. The themes of these groups were Psychology (548 members), Campus Tips (1,312 members), Free & For Sale

(2,142 members), Jobs & Internships (3,528 members), FIU Party Series (5,003 members). Approximately, 14 people replied to the posts and 3 agreed to participate in the study.

Dear all,

Do you like to participate in a PhD transportation engineering research???

For my doctoral dissertation, some participants can take part in the study. If you like, feel free to inbox me for more information or please give me a call (305) 305-1598.

Cheers!

Figure 3-7 Message Used in Different Group Pages in Facebook

Aside from this, several professors were contacted for potential recruitment through their courses. More than 12 classes were visited and over 300 students were invited to participate.

3.5. Sample Composition

In this section, different attributes of participants are discussed. The discussions include attributes such as age, education and other information. Survey recruitment process was started in December 2015 and the participants were asked to keep their google location history activated during January and February of 2016. 46 participants' data were collected successfully.

Education

The following choices were available for each participant: doctorate degree, some college credit or no degree, associate degree, master's degree, bachelor's degree, high

school graduate or diploma, no schooling completed, nursery school to 8th grade, trade or technical or vocational training, and professional degree.

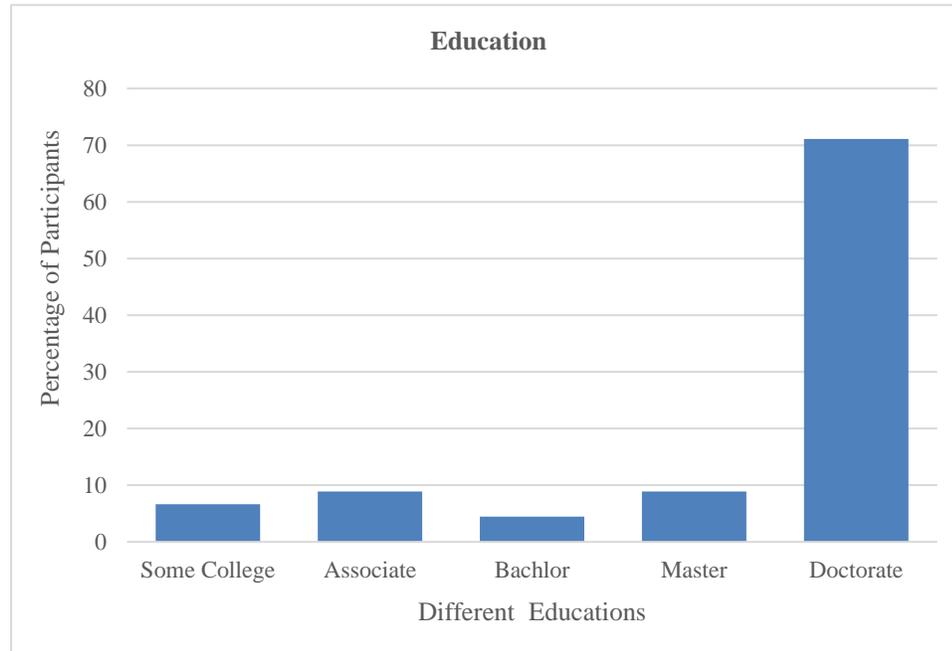


Figure 3-9 Histogram of Education

As Figure 3-9 shows, more than half of the sample was comprised of doctoral students. A potential explanation for this may be the of availability of doctoral students. These students are usually on-campus more often and were more likely to see the flyers. Moreover, as they are frequently on-campus, it was also more feasible for the researcher to speak with them in-person and convince them to participate. It is also possible that they understand the value of the study and its implication more than the other groups would.

Gender

The following choices were available for participants to choose: male, female, or other. In this sample, the majority of participants were male. This may be due to the fact

that females may be less likely to share their data, or that the majority of doctoral students in the sample were male.

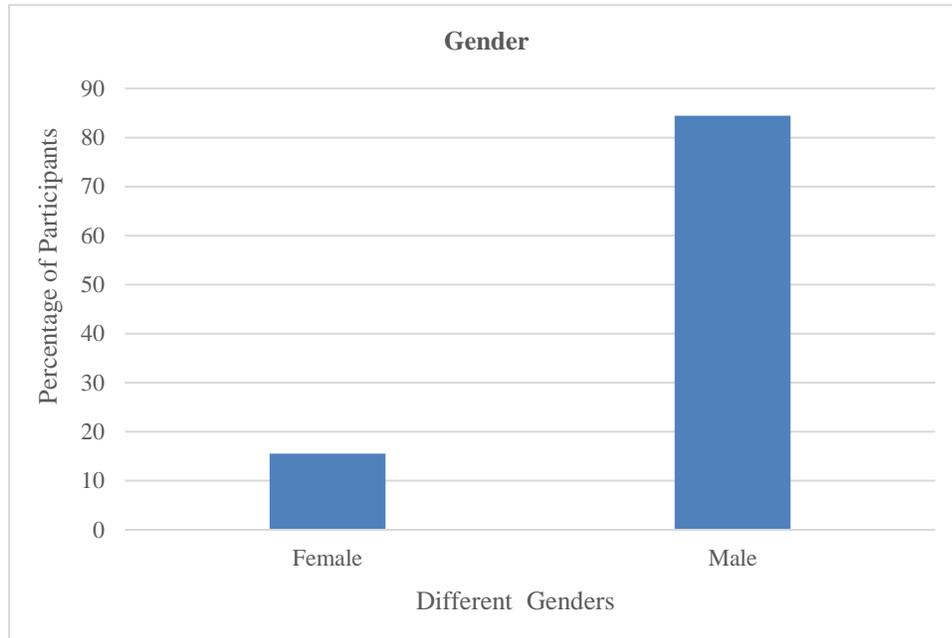


Figure 3-10 Histogram of Gender

Ethnicity

The following choices were available for participants to choose: White, Hispanic or Latino, Black or African American, Asian, Other, Native American or American Indian, as well as Pacific Islander.

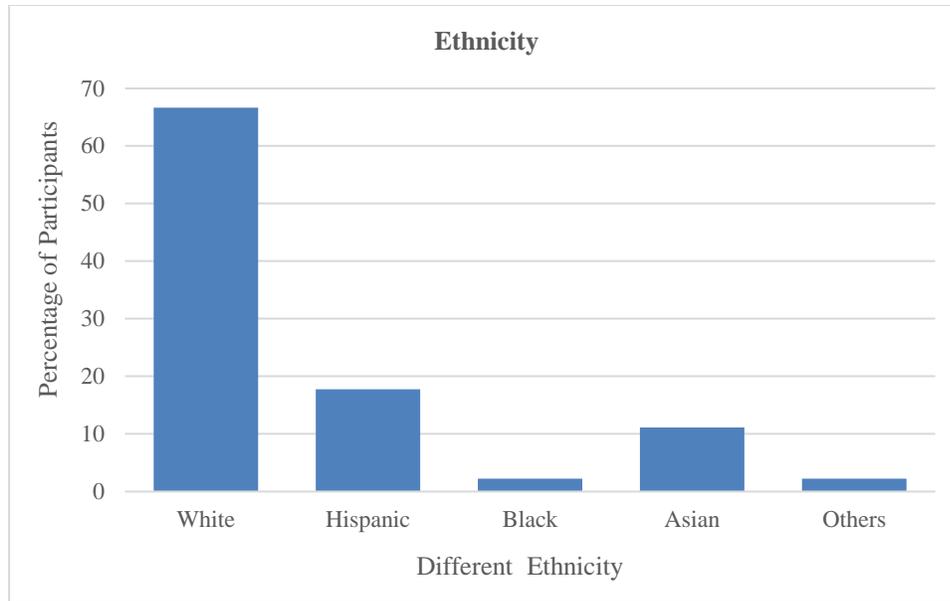


Figure 3-11 Histogram of Ethnicity

A survey in in 2014 revealed that 74.57% of people were white, 20.14% were black, 0.89% were Asian, 0.11% were Native American, and 3.18% identified as other. Also, 70.34% of the people in Miami, Florida, identified as Hispanic (meaning 29.66% are non-Hispanic) (Bestplaces, 2017). These categories considered people from North American, the Middle East, and Europe as White. Hence, White ethnicity has a wide variety and it is logical that White ethnicity has the highest percentage among other ethnicities.

Marital Status

The following choices were available for participants to choose: single or never married, married or domestic partnership, widowed, divorced, or separated. In 2012, the percentages of married people in Miami and US were 35.58% and 49.22%, respectively (Bestplaces, 2017). As the majority of participants of this study were young people, it makes sense that more participants were single.

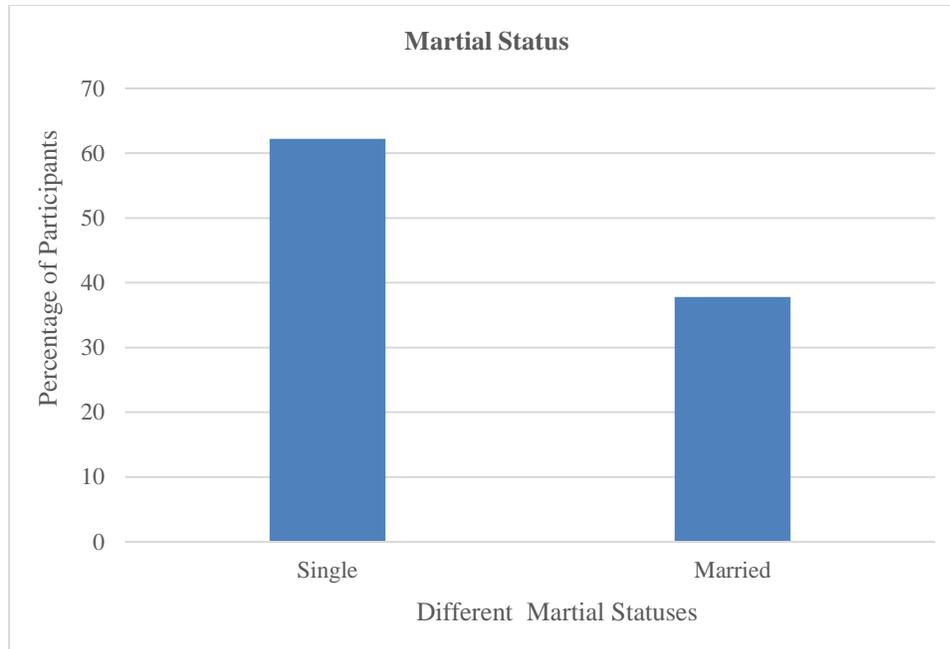


Figure 3-12 Histogram of Marital Status

Employment Status

The following choices were available: employed for wages, student, self-employed, out of work and looking for work, homemaker, out of work but not currently looking for work, military, and retired. Most participants in this study were 25 years of age or older so it is likely that the majority of them would be employed for wages.

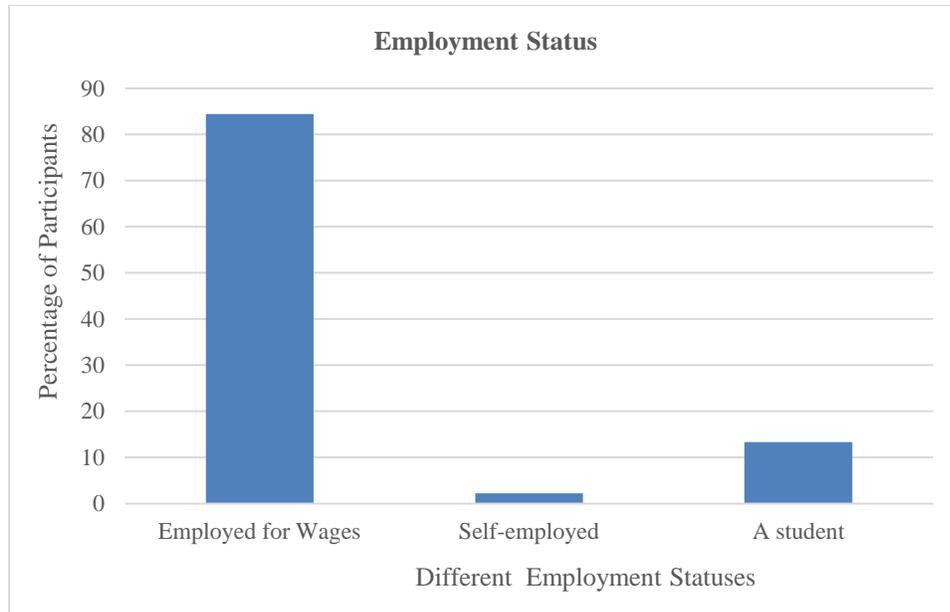


Figure 3-13 Histogram of Employment Status

Driving License

The following choices were available for participants to choose: yes or no. Most people in Miami use their private vehicle much more frequently than public transportations. In light of this, it reasonable that nearly all participants had a driver’s license as driving is the only reliable mode of transportation in the area.

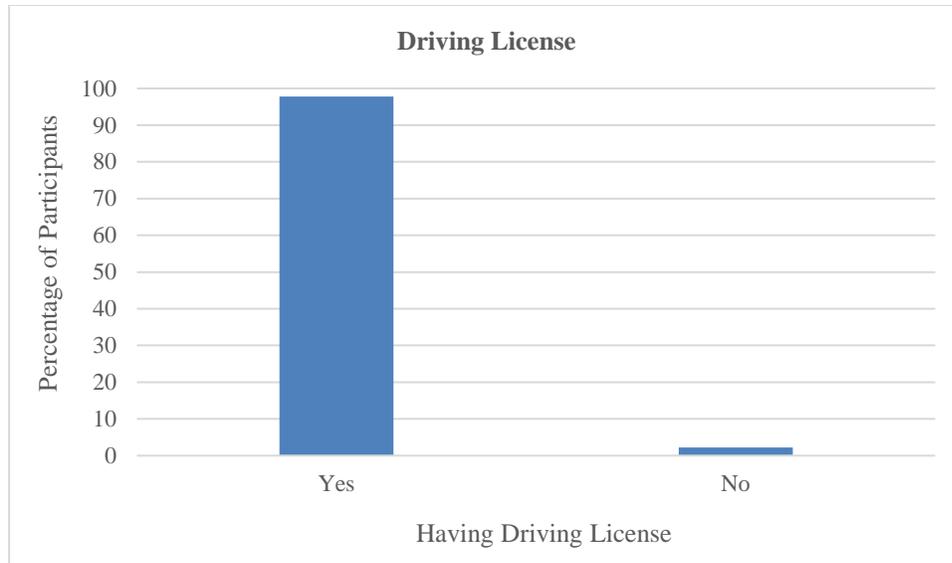


Figure 3-14 Histogram of Driving License

Age

The following choices were available for participants to choose: Under 12, 12-17 years, 18-24 years, 25-34 years, 35-44 years, 45-54 years, 55-64 years, 65-74 years, and 75 years or older.

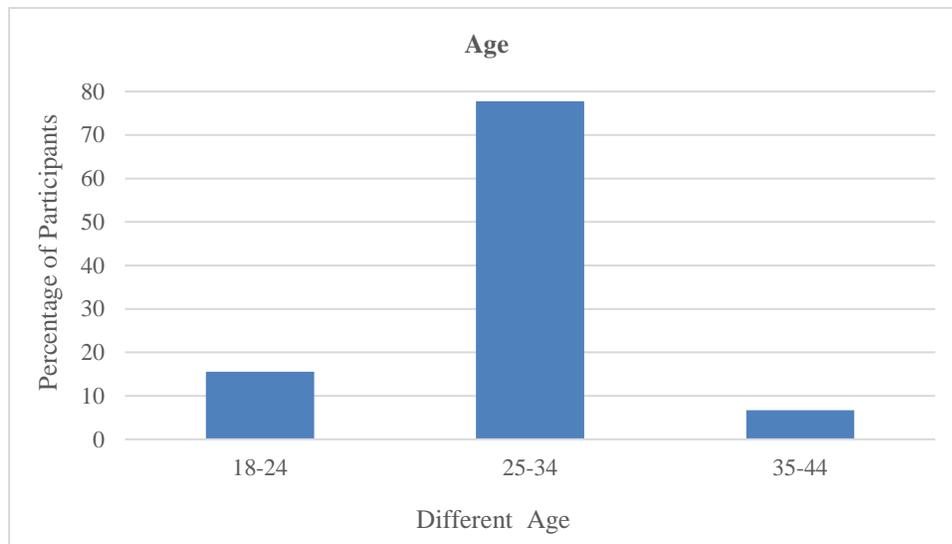


Figure 3-15 Histogram of Age

As the vast majority of this study’s sample was doctoral students, it is reasonable that the majority of the sample was between 25 and 34 years of age.

Being Fulltime Worker

The following choices were available for participants to choose: yes, or no.

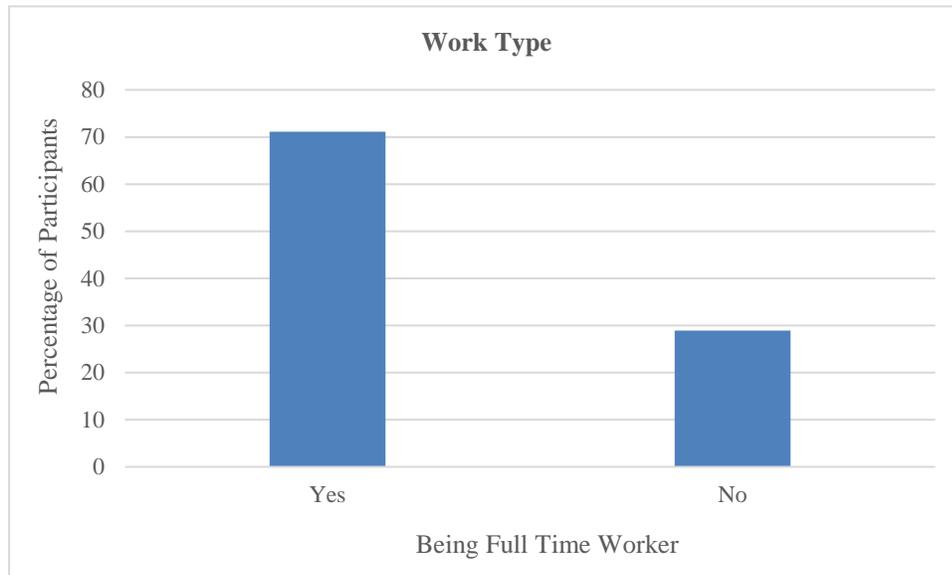


Figure 3-16 Histogram of Work Type

As participants are all between 25 and 44 years old, they are mainly full-time workers.

Table 3-1 shows the descriptive statistics of the entire sample compiled for this study.

Table 3-1 Descriptive Statistics

Education	Percentage	Number
Some college credit, no degree	6.67	3
Associate degree	8.89	4
Bachelor's degree	4.44	2
Master's degree	8.89	4
Doctorate degree	71.11	32
Gender		
	Percentage	Number
Female	15.56	7
Male	84.44	38
Ethnicity		
	Percentage	Number
White	66.67	30
Hispanic or Latino	17.78	8
Black or African American	2.22	1
Asian	11.11	5
Other	2.22	1
Marital Status		
	Percentage	Number
Single, never married	62.22	28
Married or domestic partnership	37.78	17
Employment Status		
	Percentage	Number
Employed for wages	84.44	38
Self-employed	2.22	1
A student	13.33	6
Having Driving License		
	Percentage	Number
Yes	97.78	44
No	2.22	1
Age		
	Percentage	Number
18-24 years	15.56	7
25-34 years	77.78	35
35-44 years	6.67	3
Being a Full-Time Worker		
	Percentage	Number
Yes	71.11	32
No	28.89	13

4. DATA PROCESSING METHODS

After the data were collected from the participants, the next step was data pre-processing. The raw data was in KML format, therefore, they needed to be converted to XML file which can be read by Microsoft Excel and MATLAB. Next, the data were cleaned and they were passed through multiple filters to detect any false record. Moreover, the raw data had information which were not related to the objectives of this study, hence, they were removed. The data processing has two main sections including fixed location detection and trip purpose detection.

In summary, to detect the fixed location the data went through the following steps; first, the whole data were separated to separate days. The separation was done based on the detection time of any data point which is less than the time of the previous data point. Second, the data were changed from degree to distance, using Spherical Law of Cosines. Third, if the participant is not moving for more than 6 minutes and the distance between two consecutive points is less than 150 feet a fixed location is detected. Fourth, based on the first fixed location and if the fixed locations are between 3:00 am and 4:00 am, the home, and if the fixed locations are between 11:00 am and 12:00 pm and 2:00 pm and 3:00 pm work place are detected. Fifth, frequency of the detected locations are calculated to see if the location which meet the criteria is meeting it not just one time and it is the most frequent one. Sixth, the provided results of the MATLAB export to the ArcGIS software and the coordinates of the detected locations are overlaid with the latest land use layer to see in which land use the detected locations are fallen (Baqersad et al., 2016) . The

coordinates of the detected locations were compared with the coordinates of the home and office address of the questionnaire.

4.1. Fixed Location Detection

For each single participant a file in xml extension was prepared for further analysis in MATLAB. Date, time, latitude, and longitude were the pertinent data for each point. Within MATLAB, the data were separated by day in order to capture daily variations. Next, distances between consecutive points were calculated based on their coordinates. As the coordinates were provided in degrees, the distance could be calculated from the Spherical Law of Cosines (Wikipedia, 2016).

$$d = A \cos[\sin(\varphi_1) \sin(\varphi_2) + \cos(\varphi_1) \cos(\varphi_2) \cos(\Delta\lambda)]R \quad (1)$$

Where,
 d = distance between records
 φ = Latitude
 λ = Longitude
 R = Earth Radius (3959 miles)

Then, based on distance between consecutive points and the corresponding time interval, speed was calculated. Based on the understanding of the data, a fixed location was identified when the distance between consecutive points was less than 150 feet and the dwell time (duration at the location) was more than 6 minutes. Meaning that if the data shows there is no movement in coordinates of the participants, they are still and not moving. This is the main criteria in the literature review as well. This distance, 150 feet, was chosen as the criteria due to the consideration that a person may move around the office or in the house.

4.2. Trip Purpose Detection

Once all fixed locations are detected, the duration of activities could be derived based on the time difference between the first and last point at a fixed location. Once duration is calculated, the activity purpose was inferred. In other words, what was the purpose of the visiting the location. This study considered three types of fixed locations: Home, Work, and Others. Different logics and assumptions were considered to identify these three categories; the most practical was used. Majority of people were asleep at night at home and the day's first trip could be identified. Due to this, the first fixed location was defined as Home; this was one of the criteria used to detect this type of fixed location in this study.

The next criterion was the number of times, or frequency, a participant visited a particular location. For instance, if a person's data show that he or she is at a particular location for most nights, but was at another location for only one night, that other locations cannot be his or her home. However, it is likely that the location where most nights were spent corresponds to his/her home. Similarly, a conclusion can be drawn for the workplace category. In the case that more than one location had the potential to be the participant's workplace, the one with a higher frequency was to be considered the workplace.

There were a few limitations when detecting the Work place. Inferring the workplace is generally more complicated and can result in less accurate final findings (Deng and Ji, 2010). Although the majority of people work during the day, such as from 9:00 AM to 5:00 PM, this is not true for all. It differs by job type and also lifestyle. Location detection can be based on the combination of hours. For instance, work place for a typical person can be identified if the person stays at the location between 11:00 AM and 12:00

PM and/or 2:00 PM and 3:00 PM during weekdays. Additionally, the frequency of visits to that location were checked. Therefore, in the first step of locations detection, coordinates of all fixed locations which can be home and workplace are detected.

Figure 4-1 illustrate an example of the data in a map format and Figure 4-2 shows its analysis. It can be seen that the first trip took place from home (H) to work (W) at 9:52:20 AM and the second trip was from work to home at 3:37:53 PM. In both trips there is a quick stop at traffic light (TL). The participant spent about five and a half hours at his work place and then headed back home.

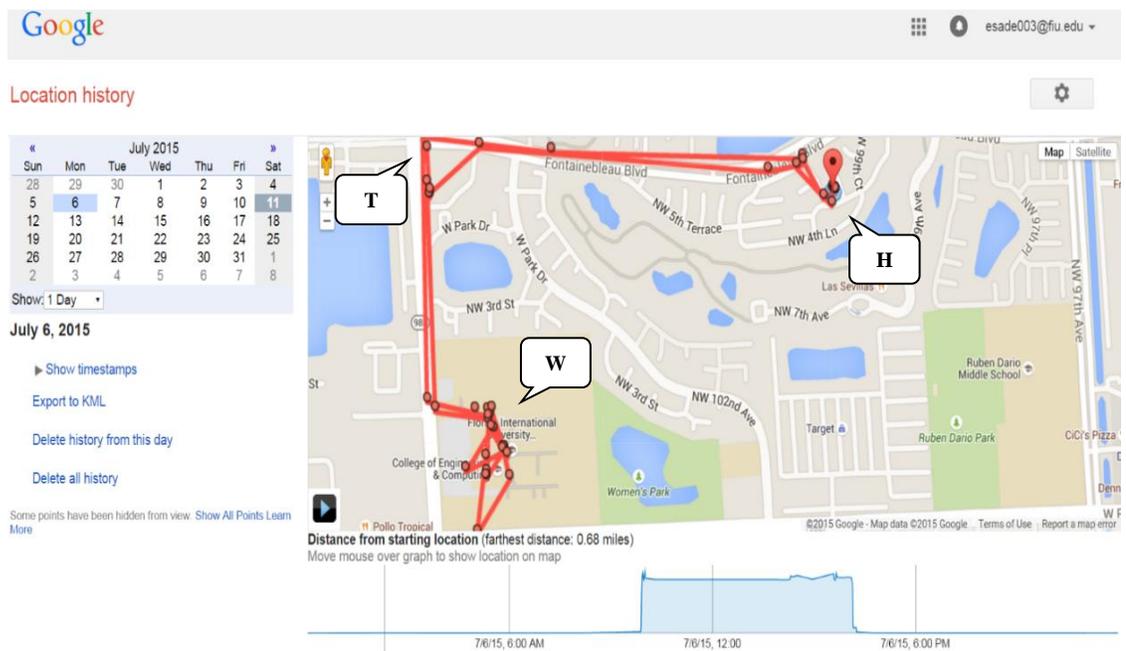
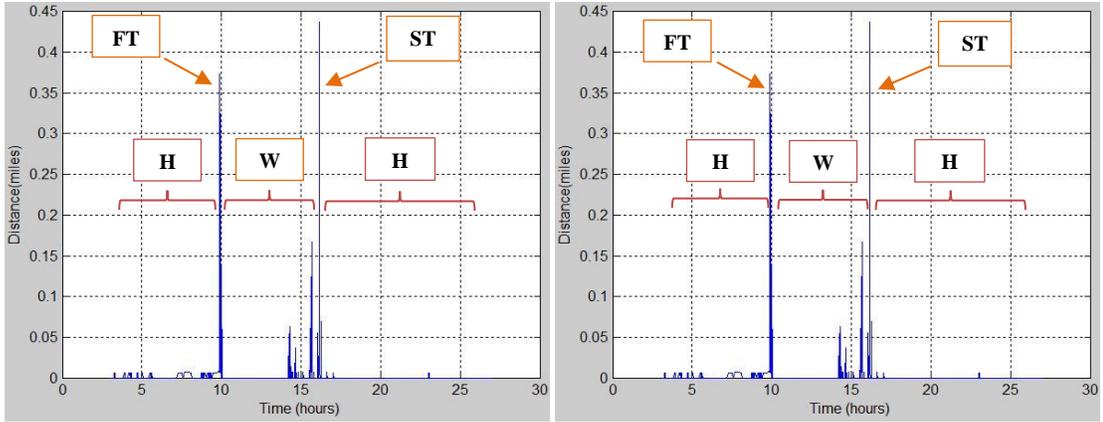
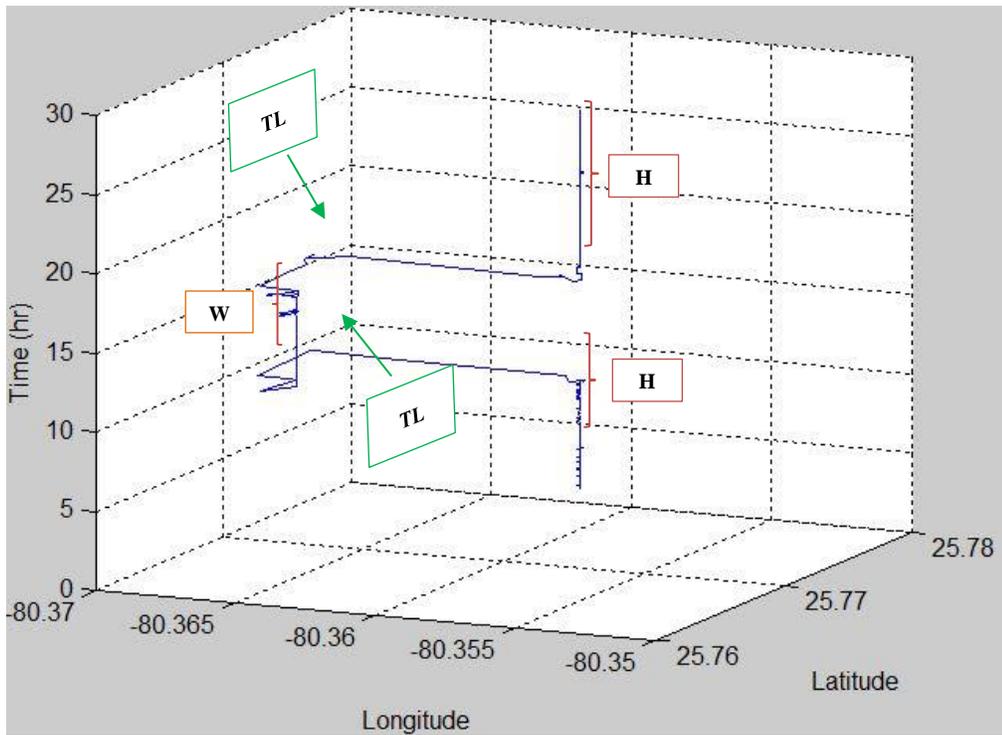


Figure 4-1 GLH Data Trace for a Sample Daily Travel



(a)

(b)



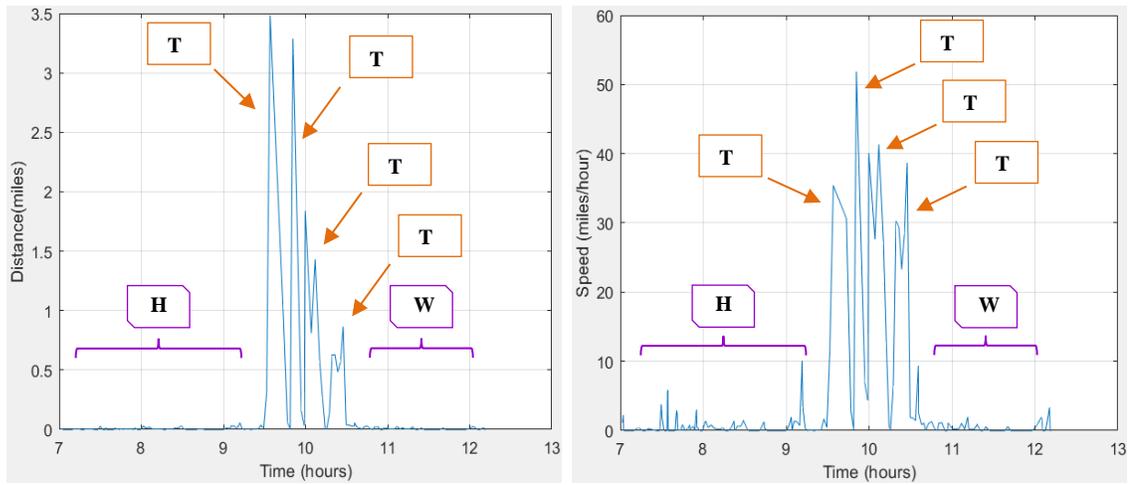
(c)

Figure 4-2 Sample Travel (a) Distance-Time (b) Speed-Time (c) 3D Space-time

As another example, the processing of trip of Figure 4-6 is explained. After cleaning the data and removing unnecessary points, each data point had three important pieces of

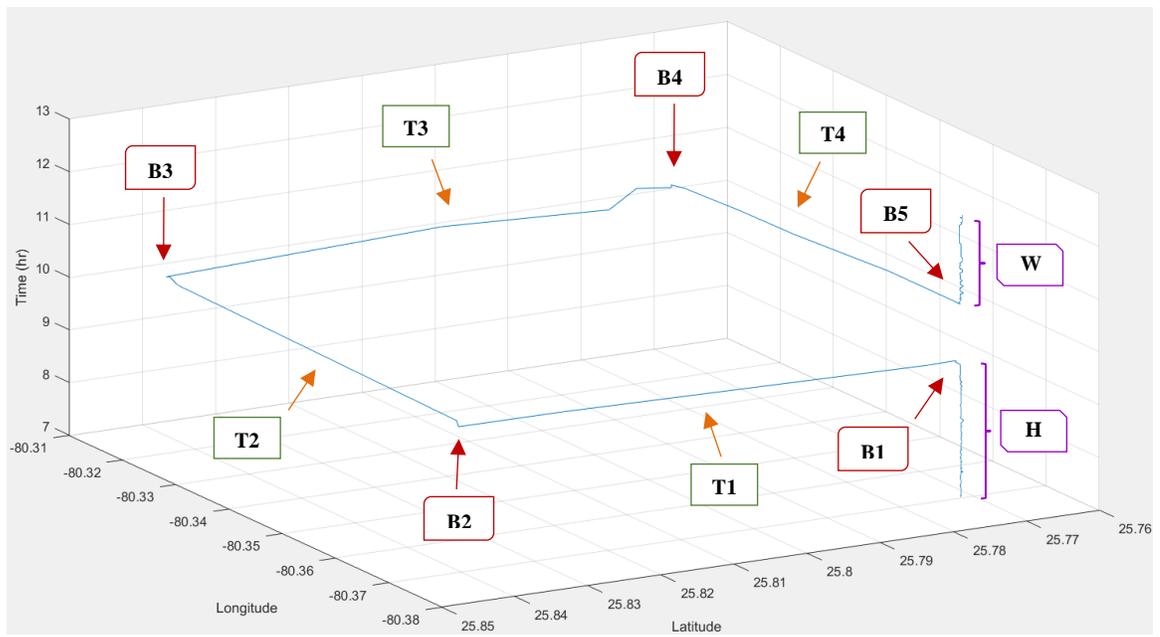
information: longitude, latitude, and the timestamp. The data were prepared to be read by MATLAB. As this analysis was based on daily travel, the data were separated by day. The coordinates were in the unit of degrees, therefore, the Spherical Law of Cosines was used to calculate the distance between two consecutive data point. Next, having distance between consecutive data points and their timestamps, the average travel speeds were calculated. The average speed and acceleration diagram for one trip is shown in Figure 4-3 (a) and 4-3 (b), respectively. It clearly visible that trip ends could be identified when the speed and acceleration were lowest. Conversely, when speed and acceleration were highest, it could be assumed that a trip was on-going.

This same trip was used to create Figure 4-3 (c). This figure represents the 3D time-space diagram, which resembles the benchmarks' location shown in Figure 5-1 but incorporates the time dimension. From this figure, it is possible to see that the researcher was at benchmark (B1) at 9:30 AM and then proceeded to B2, B3, B4, and B5. Again, the distance (and time) between the benchmarks correspond to his trips between benchmarks.



(a)

(b)



(c)

Figure 4-3 Verification Trips (a) Distance-Time (b) Speed-Time (c) 3D Time-space

The next criterion was the number of times, or frequency, a participant visited a particular location. For instance, if a person's data show that he or she is at a particular location for most nights, but was at another location for only one night, that other location cannot be his or her home. However, it is likely that the location where most nights were spent corresponds to his/her home. Similarly, a conclusion can be drawn for the workplace category. In the case that more than one location had the potential to be the participant's workplace, the one with a higher frequency was to be considered the workplace.

In this study three location activity categories were considered: Workplace, Home, and Other. In the MATLAB code, different assumptions and logics were considered to infer the location type. For instance, home can be assumed as the location from which the first trip of the day starts as most individuals sleep at home each night.-The majority of people work during the day and likely between 9:00 a.m. and 5:00 p.m., but this is not the case for everyone; different jobs or lifestyles may result in different working hours. In this study's MATLAB algorithms, the workplace was identified as a fixed location between 11:00 a.m. and 12:00 p.m. or 2:00 p.m. and 3:00 p.m. on weekdays. Therefore, in the first step of locations detection, coordinates of all fixed locations which can be home and workplace are detected.

The GIS land use layer was used as another criterion for the detection of activity locations. The coordinates of were imported into GIS. After doing this spatial analysis, GIS provided the detected locations' land use such as Medium Density Residential (MDR) 13-25 DU/AC, Institutions, Parks and Recreation, Industrial, Environmentally Protected Parks, Business and Office, Water, Open Land, Zoo Maimi Entertainment Area, and Agriculture. A layer of detected locations coordinates was imported to GIS and a spatial

analysis was conducted based on the land use. For instance, although there might be few exceptions, houses were commonly found in residential areas. Therefore, identification of activity could be accomplished with higher level of certainty. The updated land use layer used in this step had the following land uses: agriculture, business and office, environmental protection, environmentally protected parks, estate density residential (edr) 1-2.5 du/ac, estate density residential with density increase 1, high density residential (hdr) 60-125 du/ac, industrial and office, institutions, utilities and communication, low density residential (ldr) 2.5-6 du/ac, low density residential with density increase 1, low medium density residential w/ density increase 1, low-medium density residential (lmdr) 6-13 du/ac, medium density residential (mdr) 13-25 du/ac, medium density residential w/ density increase 1, medium-high density residential (mhdr) 25-60 du/ac, office/residential, open land, parks and recreation, restricted industrial and office, terminals, transportation (rail, Metrorail, etc.), water, and zoo Miami entertainment area.

An excel file containing all the detected home and workplace locations was imported into ArcMap. The researcher ensured that both, the land use layer and home/workplace coordinates had the same projection. After using the “intersect” command in the ArcMap software, the land use patterns were added to home/workplace coordinates. Finally, the coordinates were overlaid onto the land use layer. Figure 4-4 presents the GIS interface of location activity detection and Figure 4-5 shows the Home detection of participants.

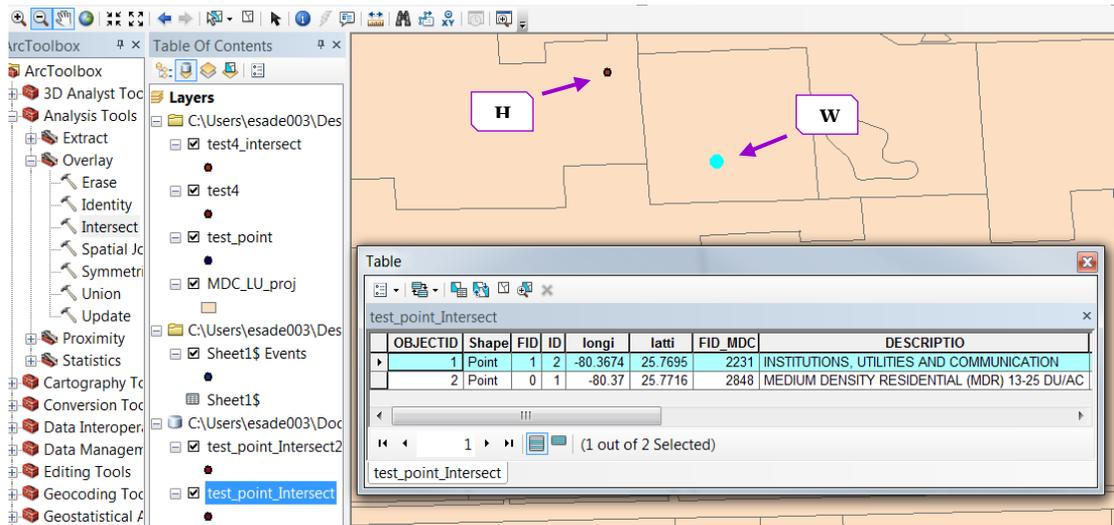


Figure 4-4 GIS Interface of a Sample Location Detection

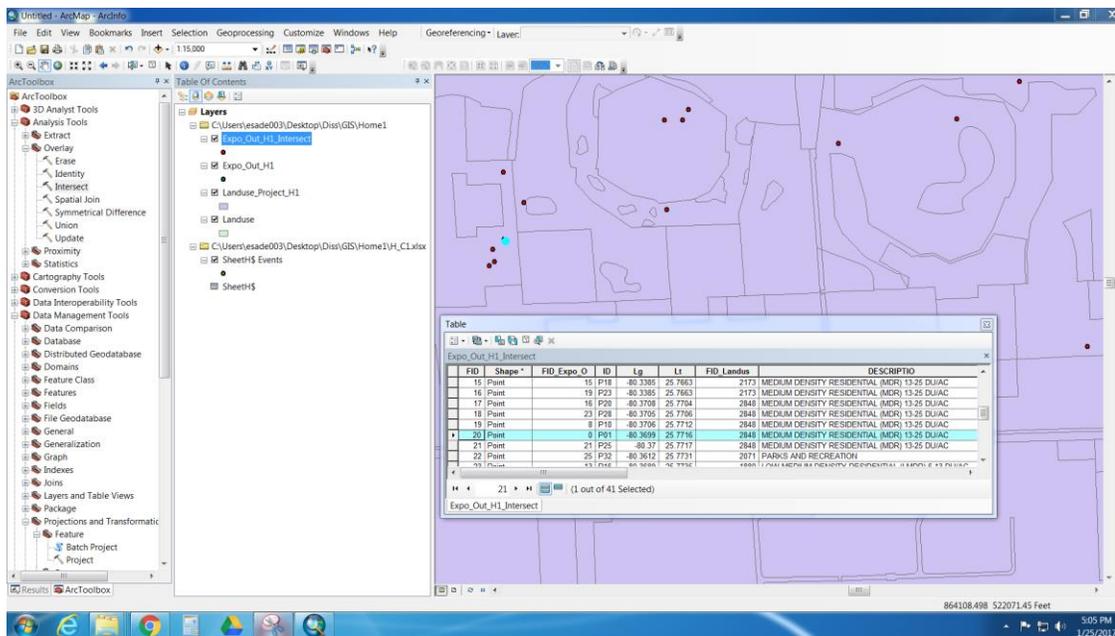


Figure 4-5 GIS Interface of Home Detection

As previously mentioned, during the questionnaire participants were asked to provide their home and workplace addresses. These addresses were compared with the data

obtained from each participant’s GLH data. The result of location detection suggested an improvement on similar location detection studies (Deng and Ji, 2010). As Figure 4-6 shows 93.48% of home and 70.27% of work places were detected correctly.



Figure 4-6 Locations Detection (a) Home Detection (b) Workplace Detection

4.3. Data Validation

In order to verify the accuracy of the data the result of the GLH and a GPS device were needed to be compared. If there was not a big difference between the GPS device data and GLH data, it could be concluded that the GLH data was good and useful for transportation purpose. First, a separate GPS unit was deployed and then a trip with certain number of benchmarks in a pre-decided path with both GPS device and GLH could showed how the output of the two methods were different. A trip was planned and five benchmarks were selected. The researcher traveled to each benchmark, rested for a few minutes, and then proceeded to the next. Overall, the verification trip took around 1.5 hours. The trip began in Doral just north of the intersection of NW 107 Avenue and West Flagler Street;

this point was labeled B1 (see Figure 4-5). From B1, the researcher traveled to B2 which was located at the intersection of NW 107 Avenue and NW 74 Street, then to the intersection of NW 74 Street and Milam Dairy Road (B3), then to the intersection of Milam Dairy Road and West Flagler Street (B4), and finally just east of the intersection of West Flagler Street and NW 107 Avenue (B5). The distance between consecutive benchmarks were noted as trips; for example, trip 1 (T1) was the distance between B1 and B2. Likewise, T4 was the trip between B4 and B5.

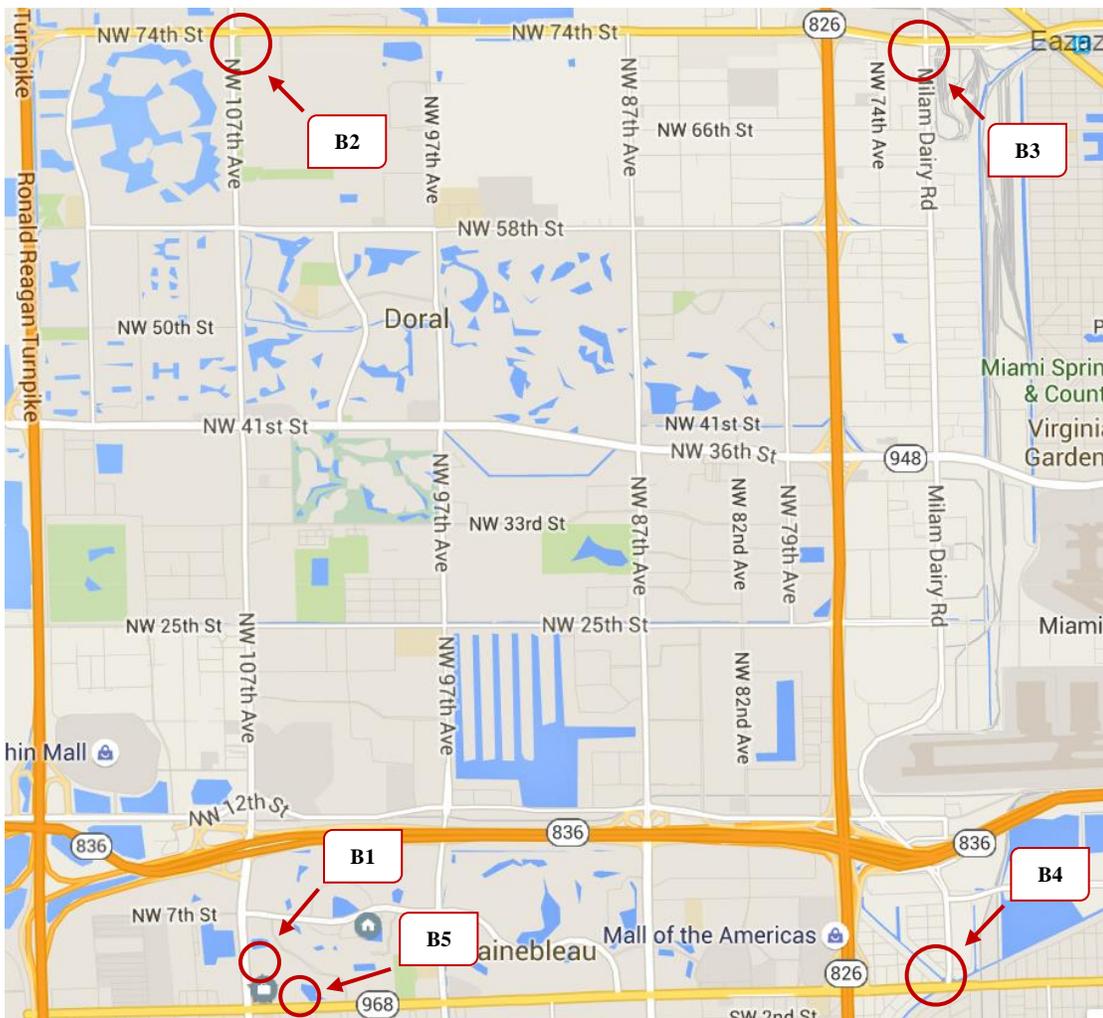


Figure 4-6 Benchmarks of Data Verification Trip

The exact benchmarks' coordinates were obtained from Google Maps and then compared to the data collected by the GLH and GPS device. This comparison showed that both methods detected the researcher's location accurately and therefore was acceptable to be in this study. Table 4-1 shows the similarity between the Google Maps, GLH and GPS data. Usually, the differences between the coordinates from GLH and Google Maps are around 30.23 feet (9.2 meters).

Table 4-1 Collected Data Using Different Methods

Benchmark	Google Maps		GLH		GPS	
	Latitude	Longitude	Latitude	Longitude	Latitude	Longitude
1	25.770865	-80.368721	25.7708099	-80.3687204	25.770850	-80.368742
2	25.840476	-80.369839	25.8406045	-80.3702609	25.840492	-80.369838
3	25.839715	-80.314308	25.8398448	-80.3143386	25.839730	-80.314322
4	25.770350	-80.313627	25.7703427	-80.3136940	25.770350	-80.313592
5	25.769309	-80.366480	25.7694927	-80.3668654	25.769330	-80.366512

Regardless of using which method is used, GPS or GLH, data accuracy can vary due to many factors (Wikipedia, 2016). A good antenna, or aerial, is required to detect information coming from GPS satellites. Also the more number of satellites in view, is also important. When more satellites are in view, the data will be more accurate. Generally, a direct line of sight, between the device and satellite is also important for data accuracy. Due to this, if a device lies outside the direct path, accuracy will degrade. In particular, this issue is important in urban environments, valleys and mountainous areas. Another issue arises from signals reflecting off surfaces when being sent and received to and from the device. these reflections generate multi-path signals arriving with a small time delays at the receiver which results in inaccurately calculated position.

5. TRAVEL PATTERN ANALYSIS

Various mobility and travel behavior analysis can be applied to the cleaned dataset. This section presents some examples of the applications, and mainly focuses on three perspectives: hourly variations in trip making, daily variations in mobility, and the influence of demographic attributes on travel behavior.

Studying the temporal variations of travel behavior is an important issue in transportation planning. Previous studies indicated that individuals were more likely to vary their location choices in the afternoon than in the morning and evening (Wang et al., 2015; Joh et al., 2005).

Cell phone GPS data obtained from the respondents were separated by days. It was possible to identify different trip characteristics that occurred on certain days of the week. More specifically, daily variations of travel time duration for different trips were obtained. Daily variations for maximum travel distance from home were also evaluated. Subsequently, different combinations of days were analyzed. This study compared weekdays to weekends, and midweek days to shoulder days. Midweek days include Tuesday, Wednesday, and Thursday; shoulder days include Monday and Friday.

Statistical tests were performed to understand whether the difference between days was statistically significant. To check for significance, the Analysis of Variance (ANOVA) test and the Bonferroni correction were conducted. In all statistical tests, 95% confidence intervals were assumed.

5.1. Time-of-Day Dependence for Long-Distance Trips

One of the most interesting applications of the collected data is the analyzation of the effect of the time of day on long-distance travel behavior. A long-distance trip is defined in this study as one-way and is 15 miles or longer. An example of a long-distance trip is shown in Figure 5-1. The participant started the trip from Home (H), then went to an Other (O) place, then back Home, then to the other place, then to Work (W), and finally went back to Home. Table 5-1 shows the timestamps and coordinates of the participant's movements in a 24-hour timeframe.

Table 5-1 An Example of Travel Information in a 24-Hour Timeframe

Arrival Time	Departure Time	Location	Latitude	Longitude
---	07:01	Home	25.772	-80.370
07:55	10:49	Other	25.911	-80.141
11:37	18:34	Home	25.772	-80.370
18:51	19:50	other	25.772	-80.370
20:37	22:14	Work	25.770	-80.367
22:22	---	Home	25.772	-80.370

Figure 5-1 shows the participant's trajectory. The collected data from the participant was converted into a layer in GIS software, and then it was overlaid with a road layer. Moreover, a Google Map photo of the area was used as the background. Figure 5-1 shows the whole trip's trajectory; however, the short-distance locations may not be clearly seen. Figure 5-2 shows the Home, Work, and Other places on a bigger scale.

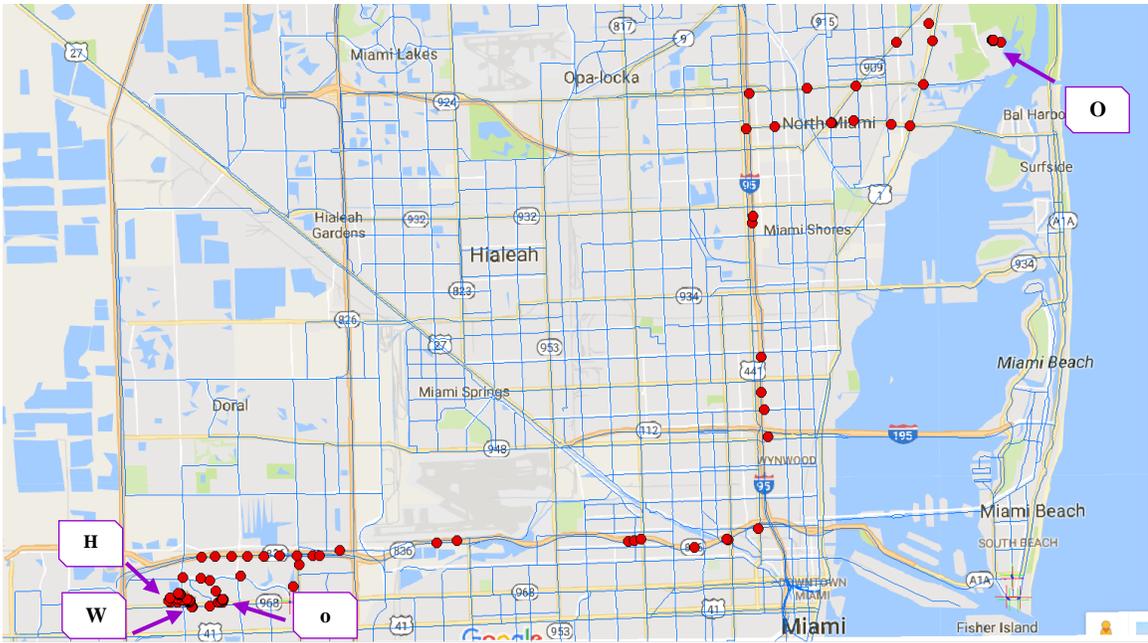


Figure 5-1 A Typical Long-Distance Trip

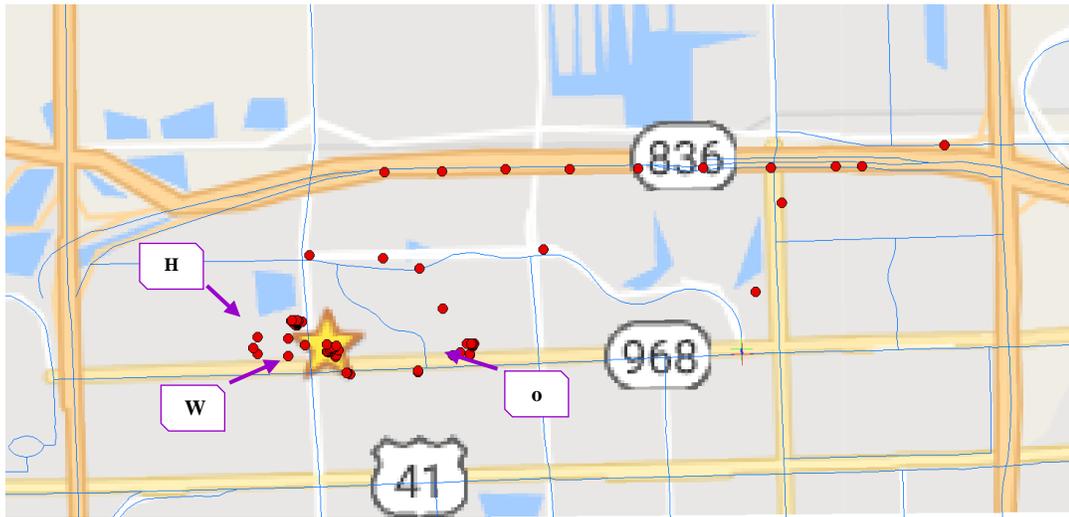


Figure 5-2 Location of Home, Work and other of the Participant

Analyzing the data in MATLAB shows the travel movement in two-dimensional and three-dimensional diagrams. Figure 5-3 shows the Longitude and Latitude diagrams of

the participant's movements, which is similar to Figure 5-1, which shows the trip's trajectory. Figure 5-4 shows the longitude and time of the trip.

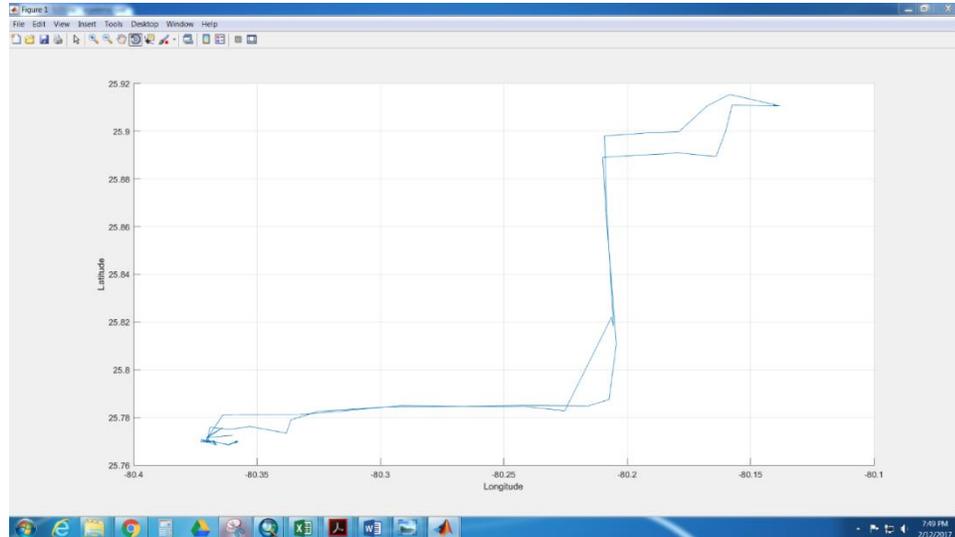


Figure 5-3 Two-Dimensional Diagram of Longitude and Latitude

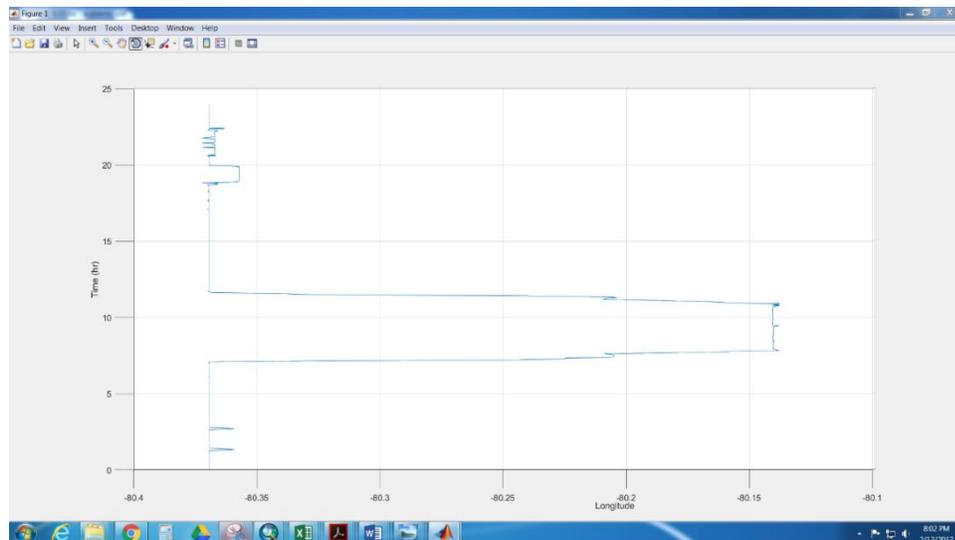


Figure 5-4 Two-Dimensional Diagram of Longitude and Time

Figure 5-5 shows the three-dimensional diagram of the trip. Figure 5-5 is also called the “time-space diagram,” which shows the trajectory of the trip and time.

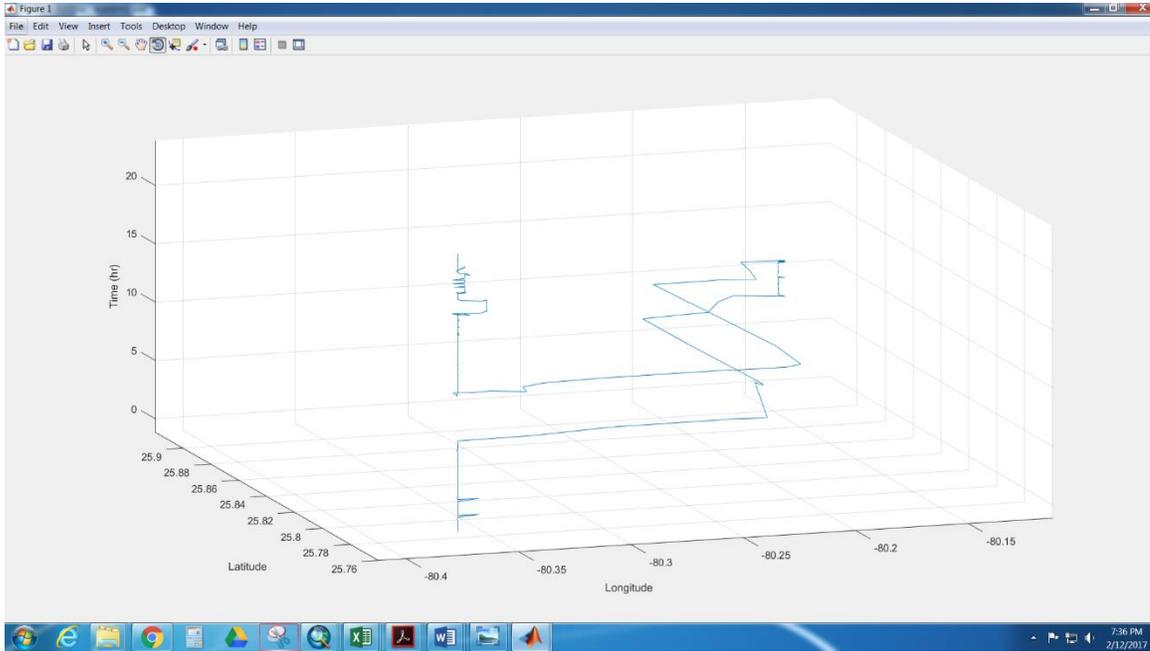


Figure 5-5 Three-Dimensional Diagram of Longitude, Latitude, and Time

Figure 5-6 shows a schematic example of time spent in different locations during a daily trip. Note that the sample subject of Figure 5-6 generated two HW trips, four HBO trips and zero NHB trips in one day.

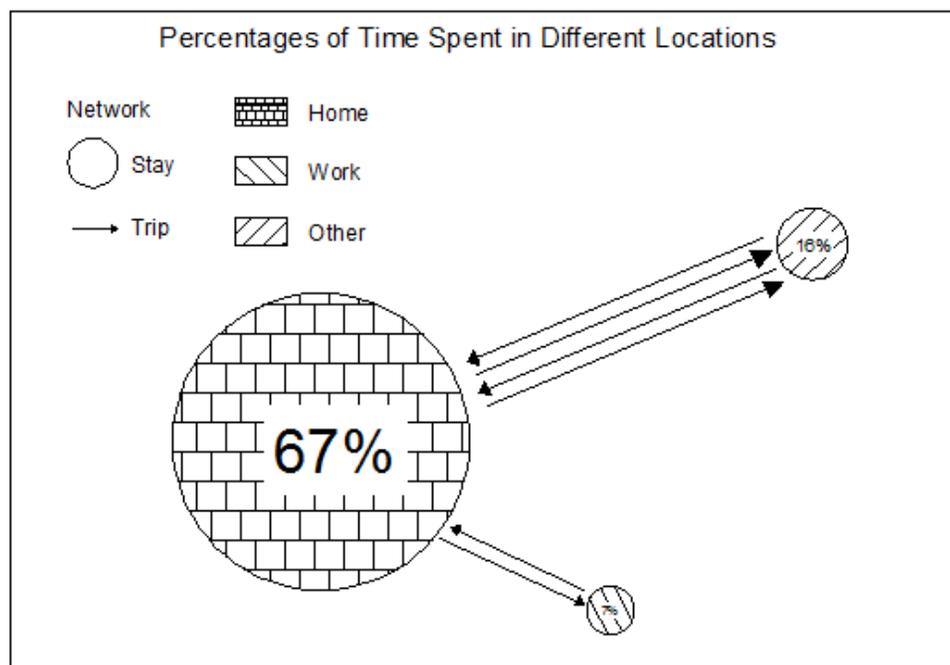
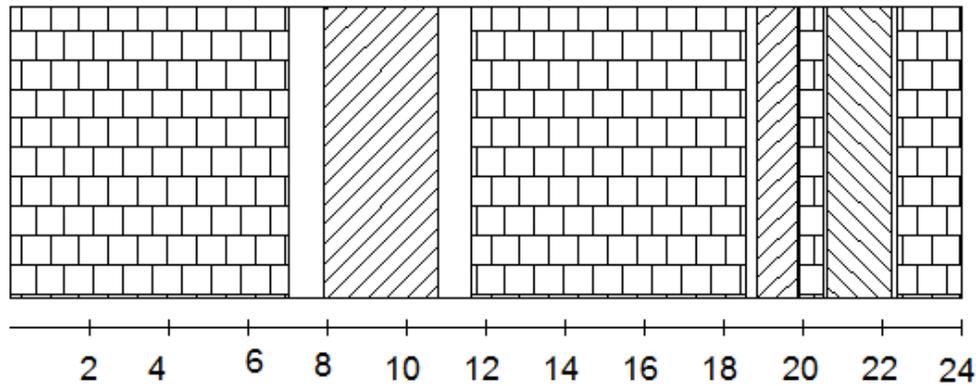


Figure 5-6 Schematic Example of Time Spent in Different Locations

In this study, the entire day was aggregated into six time periods, including early morning or “EM” (00:00–6:29), morning peak or “MP” (06:30–08:59), morning off-peak or “MOP” (09:00–11:59), afternoon off-peak or “AOP” (12:00–15:59), afternoon peak or “AP” (16:00–18:29), and evening or “EVE” (18:30–23:59). Figure 5-7 displays the frequency of long-distance trips by time of day.

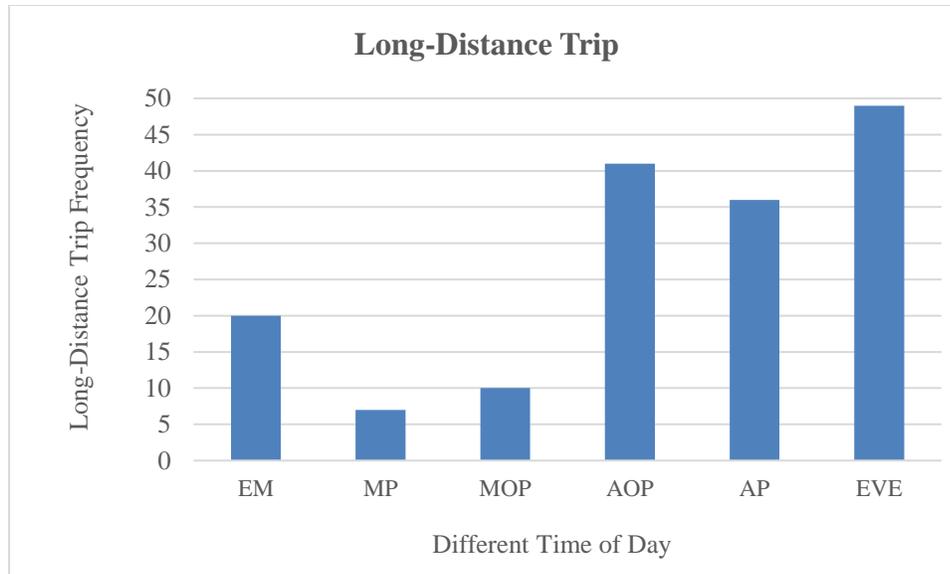


Figure 5-7 Long-Distance Trip Frequency of Different Time of Day

In terms of statistical analysis, the ANOVA (Table 5-2) and Bonferroni (Table 5-3) tests indicated that $F=3.0701$ (which is more than 1), and the p-value for testing equality of means is 0.003 (which is less than 0.05), showed strong evidence that frequencies of long-distance trips are different across time of day, at a 0.05 level of significance. In Table 5-2, different times of the day, which are EM, MP, MOP, AOP, AP, and EVE, are presenting as 1, 2, 3, 4, 5, and 6, representatively. In addition, it was found that the frequency of long-distance travel is significantly higher in the Evening than in the Morning Peak (p-value= 0.018).

Table 5-2 ANOVA Test of Long-Distance Trip Frequency of Different Times of Day

ANOVA					
LDT					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.290	5	1.658	3.701	.003
Within Groups	120.957	270	.448		
Total	129.246	275			

**Table 5-3 Bonferroni Test of Long-Distance Frequency of Trips of Different Times of Day
Post Hoc Tests**

Multiple Comparisons						
Dependent Variable: LDT						
Bonferroni						
(I) TimeOfDay	(J) TimeOfDay	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	.23913	.13956	1.000	-.1742	.6525
	3.00	.19565	.13956	1.000	-.2177	.6090
	4.00	-.08696	.13956	1.000	-.5003	.3264
	5.00	-.17391	.13956	1.000	-.5872	.2394
	6.00	-.21739	.13956	1.000	-.6307	.1959
2.00	1.00	-.23913	.13956	1.000	-.6525	.1742
	3.00	-.04348	.13956	1.000	-.4568	.3698
	4.00	-.32609	.13956	.303	-.7394	.0872
	5.00	-.41304	.13956	.050	-.8264	.0003
	6.00	-.45652*	.13956	.018	-.8698	-.0432
3.00	1.00	-.19565	.13956	1.000	-.6090	.2177
	2.00	.04348	.13956	1.000	-.3698	.4568
	4.00	-.28261	.13956	.658	-.6959	.1307
	5.00	-.36957	.13956	.129	-.7829	.0438
	6.00	-.41304	.13956	.050	-.8264	.0003
4.00	1.00	.08696	.13956	1.000	-.3264	.5003
	2.00	.32609	.13956	.303	-.0872	.7394
	3.00	.28261	.13956	.658	-.1307	.6959
	5.00	-.08696	.13956	1.000	-.5003	.3264
	6.00	-.13043	.13956	1.000	-.5438	.2829
5.00	1.00	.17391	.13956	1.000	-.2394	.5872
	2.00	.41304	.13956	.050	-.0003	.8264
	3.00	.36957	.13956	.129	-.0438	.7829
	4.00	.08696	.13956	1.000	-.3264	.5003
	6.00	-.04348	.13956	1.000	-.4568	.3698
6.00	1.00	.21739	.13956	1.000	-.1959	.6307
	2.00	.45652*	.13956	.018	.0432	.8698
	3.00	.41304	.13956	.050	-.0003	.8264
	4.00	.13043	.13956	1.000	-.2829	.5438
	5.00	.04348	.13956	1.000	-.3698	.4568

*. The mean difference is significant at the 0.05 level.

Another interesting application of the collected data is analyzing the effect of the time of day on extra-long distance travel behavior. An extra-long distance can be named as Irregular trip as well. An extra-long distance trip is defined in this study as a one-way trip that is 500 miles or longer. Since this study was conducted in Miami, Florida, 500 miles means a trip out of the state of Florida. Figure 5-8 displays the frequency of an extra-long distance trip by different times of day to determine whether there are any variations. Although Figure 5-8 shows that AP (Afternoon Peak) has the highest Extra-Long Trip Frequency, the ANOVA (Table 5-4) test indicated that there were no significant differences across the time of day periods (p-value is more than 0.05).

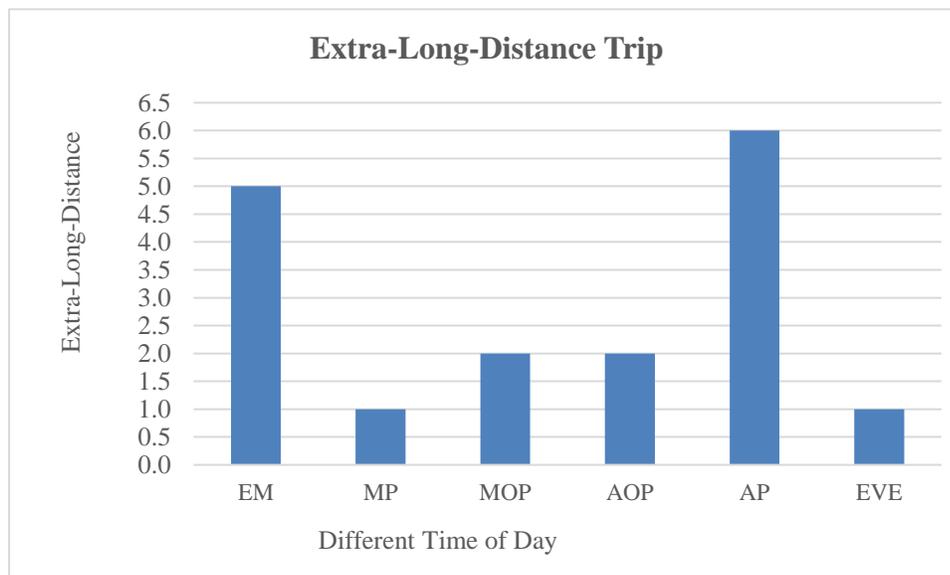


Figure 5-8 Extra-Long Distance Trip Frequency of Different Times of Day

Table 5-4 ANOVA Test of Extra-Long Distance Trip Frequency of Different Times of Day

ANOVA					
IT					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.641	5	.128	1.607	.158
Within Groups	21.543	270	.080		
Total	22.185	275			

5.2. Daily Variations for Long-Distance Trips

Figure 5-9 presents the frequency of long-distance trips by day of the week to explore whether there are any variations. Although Figure 5-9 shows that Friday has the highest Long-Distance Trip Frequency, the ANOVA (Table 5-5), indicated that there were no significant differences across all days of the week (P value is more than 0.05).

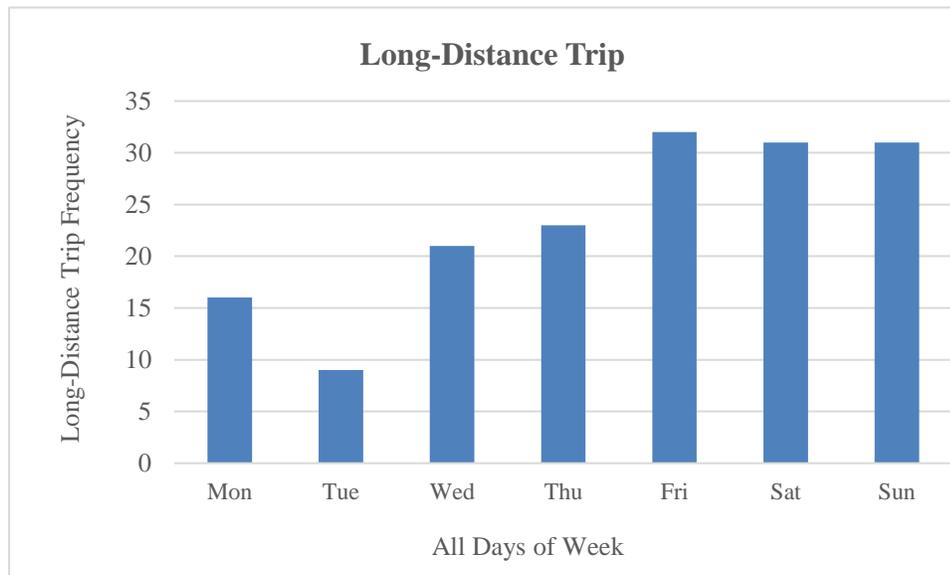


Figure 5-9 Long-Distance Trip Frequency of All Days of Week

Table 5-5 ANOVA Test of Long-Distance Trip Frequency of All Days of Week

ANOVA					
LDT					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.770	6	.795	1.204	.304
Within Groups	208.087	315	.661		
Total	212.857	321			

Figure 5-10 shows the frequency of long-distance trips by different days of the week. Although Shoulders and Weekend have higher long-distance trip frequencies than Midweek and Week Days, respectively, the ANOVA (Table 5-6) test indicated that the differences were not significant. The p-value is 0.429, which is not less than 0.05.

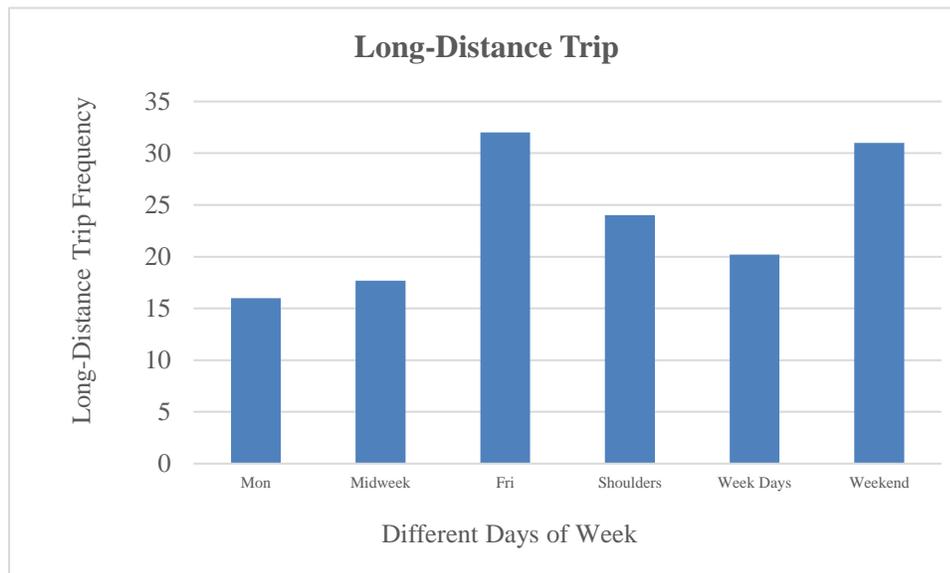


Figure 5-10 Long-Distance Trip Frequency of Different Days of Week

Table 5-6 ANOVA Test of Long-Distance Trip Frequency of Different Days of Week

ANOVA					
LDT					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.138	5	.428	.983	.429
Within Groups	117.482	270	.435		
Total	119.619	275			

Figure 5-11 presents the frequency of extra-long distance trips by day of the week to see whether there are any variations. Although Figure 5-11 shows that Tuesday has the lowest Extra-Long Trip Frequency, the ANOVA test (Tables 5-7) indicated that there were no significant differences across all days of the week (P value is more than 0.05).

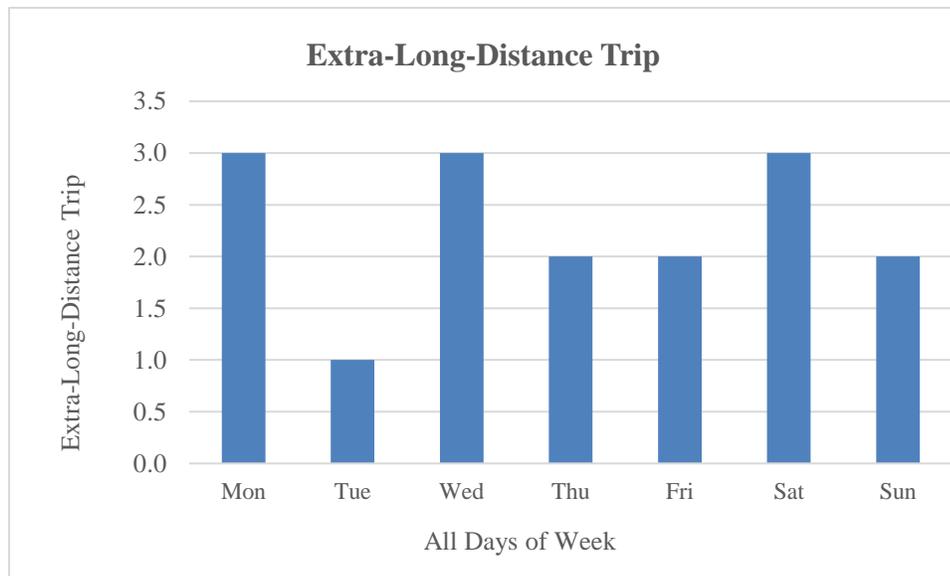


Figure 5-11 Extra-Long Distance Trip Frequency of All Days of Week

Table 5-7 ANOVA Test of Extra-Long Distance Trip Frequency of All Days of Week

ANOVA					
IT					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.062	6	.010	.201	.976
Within Groups	16.239	315	.052		
Total	16.301	321			

Figure 5-12 shows the frequency of Extra-Long trips by different days of the week. Although Shoulders and Weekend have higher long-distance trip frequencies than Midweek and Week Days, respectively, the ANOVA (Table 5-8) test indicated that the differences were not significant, or that no p-value less than 0.05 was found.

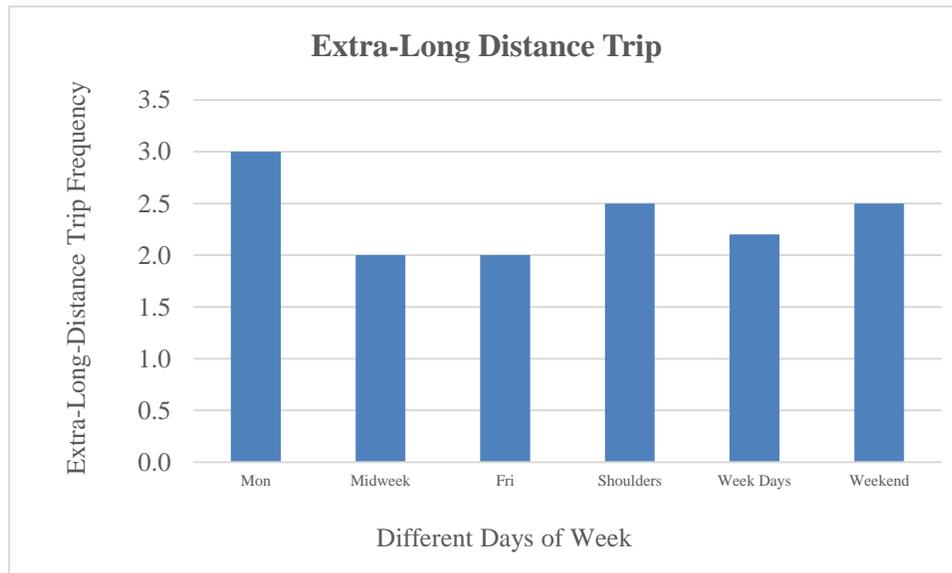


Figure 5-12 Extra-Long Distance Trip Frequency of Different Days of Week

Table 5-8 ANOVA Test of Extra-Long Distance Trip Frequency of Different Days of Week

ANOVA					
IT					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.005	5	.001	.032	.999
Within Groups	7.714	270	.029		
Total	7.719	275			

5.3. Demographic Analysis

In this section, a model for evening long-distance trip frequency was developed. Demographic information from each participant was collected to determine which independent variables significantly contribute to the dependent variables. The demographic data is categorical, and the dependent variable is number of trips; therefore, among all of the different choice models, the ordinal model was selected. First, all variables were recoded into different variables, zero and one, to develop an ordinal model in SPSS software. Then, an ordinal model, including all attributes, was developed, which showed that a couple of attributes do not have significant effects on the dependent variable. Next, a Stepwise method was used to remove the insignificant variables from the model. Tables 5-9 and 5-10 show the model's performance. The final model showed an R Square of 0.324, which is considered a good model. In addition, the ANOVA test showed that the model has a p-value less than 0.05.

Table 5-8 Model Summary of Created Models

Model Summary^d

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.411 ^a	.169	.151	1.25351	.169	9.356	1	46	.004
2	.518 ^b	.268	.236	1.18923	.099	6.107	1	45	.017
3	.606 ^c	.367	.324	1.11842	.099	6.879	1	44	.012

a. Predictors: (Constant), FullTime

b. Predictors: (Constant), FullTime, Asian

c. Predictors: (Constant), FullTime, Asian, From25To34

d. Dependent Variable: LDT

Table 5-9 ANOVA Test of Models

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14.700	1	14.700	9.356	.004 ^b
	Residual	72.279	46	1.571		
	Total	86.979	47			
2	Regression	23.337	2	11.669	8.251	.001 ^c
	Residual	63.642	45	1.414		
	Total	86.979	47			
3	Regression	31.942	3	10.647	8.512	.000 ^d
	Residual	55.038	44	1.251		
	Total	86.979	47			

a. Dependent Variable: LDT

b. Predictors: (Constant), FullTime

c. Predictors: (Constant), FullTime, Asian

d. Predictors: (Constant), FullTime, Asian, From25To34

Table 5-11 Coefficients of the Models

Coefficients ^a											
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	.200	.324		.618	.540					
	FullTime	1.194	.390	.411	3.059	.004	.411	.411	.411	1.000	1.000
2	(Constant)	-.197	.346		-.568	.573					
	FullTime	1.546	.397	.532	3.896	.000	.411	.502	.497	.871	1.148
	Asian	1.488	.602	.338	2.471	.017	.147	.346	.315	.871	1.148
3	(Constant)	.293	.376		.781	.439					
	FullTime	1.944	.403	.669	4.825	.000	.411	.588	.579	.748	1.338
	Asian	1.763	.576	.400	3.062	.004	.147	.419	.367	.842	1.187
	From25To34	-1.056	.403	-.340	-2.623	.012	-.098	-.368	-.315	.857	1.167

a. Dependent Variable: LDT

Table 5-11 shows the three created models with their attributes. The last model was represented with “FullTime,” “Asian,” and “From25To34,” and all three independent variables are statistically significant, meaning that their p-values were less than 0.05. The value of B for the FullTime variable is 1.944, which represents the amount of change in the dependent variable associated with a one-point increase in the predictor variable. Similarly, for Asian, a 1.763 change is expected in the dependent variable for every one point increase in the Asian. However, since the variables have different units, the variables with the highest contribution to the model cannot be selected with the sole consideration of B. Therefore, the Beta value should be considered, because the B value is standardized. It shows how many standard deviations the dependent variable changes for each one standard deviation increase in the independent variable. Beta values show that the FullTime variable has more of an effect on the model than the other variables. Part value, which is

also known as the “semi-partial correlation,” determines that the amount of variance removed would no longer be explained or realized in the dependent variable if the predictor variable were removed. This means that if the FullTime predictor is removed, the 0.579 of the variance will not be explained in the dependent variable. If the FullTime variable is removed, more variances are removed from the dependent variable rather than Asian, which is consistent with the Beta values. Tolerance should be greater than 0.1, as is found in all three models. Also, the VIF, which stands for Variance Inflation Factors, should be less than 10, which is also found in all three models.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

In order to make better transportation planning decisions, travel data that offers higher quality and lower costs than conventional household travel surveys are necessary. The Geographical Positioning System (GPS) has been used in many transportation studies; however, smartphone GPS data have received less attention. With a high penetration of smartphones, some researchers have collected GPS data through proprietary mobile applications that the participants were asked to install on their smartphones and activate during certain time periods. Google Location History (GLH) provides a similar opportunity, but there is no need to install or activate an additional application. All that is required for GLH to record data is that the service be turned on. GLH will continuously record the location of the device with no additional effort or input required from the participant. Simultaneously, this reduces the burden on participants and increases the quantity of data. The only other hurdle to accessing the data is the participant's ability to maneuver through GLH to transmit the data to the researcher.

Collecting and analyzing GLH data is a one of the newer methods being employed in transportation data collection studies. This research presented a data processing methodology which incorporated a data processing algorithm and GIS approach to infer activity locations from GLH data. The location detection process had two main parts, including an algorithm coded in MATLAB and spatial analysis in GIS. Two months of the participants' data were collected and analyzed. Not only was the accuracy similar to that of a traditional GPS, it was also less expensive than comparable datasets. Moreover, this

process improved the accuracy of the identification of home and workplace locations when compared with similar studies.

The following lessons were learned from this study:

- New technologies are substituting traditional methods. Previously, transportation data were obtained from household travel surveys, but new technologies are becoming more popular. This study showed that transportation data can be obtained from smartphones easily and at a reduced cost.
- Collecting the data using GLH had higher accuracy than CDR and comparable accuracy to a GPS device. In the case of CDR, if no phone calls are made then the location of the device cannot be detected. However, GLH records data constantly, and a data point is generally recorded every few minutes; some timestamps were noted to be approximately 5 minutes apart, but these generally corresponded to times when the devices were relatively motionless such as at night.
- GLH has the potential to reduce the burden imparted by carrying a GPS device. In studies where participants were required to carry a GPS device, they occasionally forgot the device or forgot to charge it. In both cases, this would result in a lack of accurate data. On the other hand, GLH does not impact such a burden. As people usually have their phones with them nearly the entire day, the vast majority of their locations will be recorded.

- Participants may avoid certain places when they are carrying the GPS device due to privacy concerns. However, they are unlikely to go many places without their smartphones. Hence, GLH has the potential to collect more data, unless the smartphone is turned off.
- Through this method of data collection, a robust and accurate dataset can be collected. This can be used for many things including transportation engineering, as was done in this study.
- The detection of various locations was one of the applications of this study's data. The two most frequently visited locations were home and work. Previous studies detected home more accurately than work.
- In this study, the location of home and work place were detected with 93.48% and 70.27% accuracy respectively which is an improvement over previous studies. One possible explanation is that the sample had similar demographics, such as students' sleeping patterns, since they usually sleep at home and during the night. Therefore, considering these two main criteria, the coordinates of home could be detected with a high level of accuracy.
- The location of work was detected with 70.27% accuracy. Although this is relatively a high percentage for work detection, there was definitely room for improvement. One potential barrier to the correct identification of workplaces is that some participants may have worked in multiple places. Also, the majority of participants were part-time workers. Some of the part-

time workers were students and at the university, where other participants attend for the purpose of learning and not working.

- It was found that the frequency of long-distance travel is significantly higher in the evening than that in the morning peak (p -value= 0.018).

6.2. Study Limitations and Future Research

There are several limitations of this study that need to be addressed in future research. Some potentially interesting topics to explore in future research include:

- Exploring the dynamic data collection using newer smartphone applications that work on all operating systems.
- Comparing the results of the GLH data with newer smartphone applications that are more useful for transportation purposes.
- As the popularity of smartwatches increases, a new application can be developed and adapted to transportation data collection.
- Transportation mode detection based on the travelers' speed and other criteria.

REFERENCES

- Abdi Kordan, A., A. Mehara Molan, S. Monajjem, E. Sadeghvaziri. Simulation Modeling of Dynamic Response of Vehicles to Different Types of Speed Control Humps. The 2nd Congress of Transportation and Development Institute (T&DI), June 8-11, 2014, Orlando, Florida., USA. American Society of Civil Engineers (ASCE) T&DI Congress 2014: pp 533-542.
- Alluri, P., A. Gan, D. Saha, L. Fernandez, and E. Sadeghvaziri. Evaluation of Signage Alternatives for Express Lane Facilities. The National Center for Transportation Systems Productivity and Management (NCTSPM), Contract # DTRT12GUTC12 with USDOT Office of the Assistant Secretary for Research and Technology (OST-R), Final Report, January 2017.
- Androidcentral, <http://www.androidcentral.com/understanding-googles-android-location-tracking> (access 10.04.2015)
- Ansari, Z., and A., Golroo. Automated Transportation Mode Detection Using Smart Phone Applications via Machine Learning: Case Study Mega City of Tehran. Transportation Research Board. Washington, D.C, 2015.
- Auld, J. A., C. Williams, A. Mohammadian, and P. Nelson. An Automated GPS-Based Prompted Recall Survey with Learning Algorithms. In Transportation Letters: The International Journal of Transportation Research 1, 2009, pp 59-79.
- Auld, J., A. K. Mohammadian, M. S. Oliveira, J. Wolf, and W. Bachman. Demographic Characterization of Anonymous Trace Travel Data. Submitted for presentation/publication at the 94th annual meeting of TRB. Washington, D.C., 2015.
- Axhausen, K. W., S. Schonfelder, J. Wolf, M. Oliveira, and U. Samaga. Eighty Weeks of GPS Traces: Approaches to Enriching the Trip Information. Transportation Research Record, No.1870, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp46-54.
- Bachman, W., M. Oliveira, J. Xu, and E. Sabina. Household-Level Global Positioning System Travel Data to Measure Regional Traffic Congestion. Transportation Research Record: Journal of the Transportation Research Board, No. 2308, Transportation Research Board of the National Academies, Washington, D.C.,2012, pp10-16.

Baqersad, M., A. E. Haghghat, M. Rowshanzamir, and H. M. Bak. Comparison of coupled and uncoupled consolidation equations using finite element method in plane-strain condition. *Civil Engineering Journal*, 2016, 2(8), 375-388.

Baqersad, M., A., Hamedi, M., Mohammadafzali, and H. Ali. Asphalt Mixture Segregation Detection: Digital Image Processing Approach. *Advances in Materials Science and Engineering*, 2017.

Baqersad, M., M. Mohammadafzali, B. Choubane, C. Holzschuher, A. Hamedi, and H. Ali. Precision Assessment of the Florida Texture Meter in Hot Mix Asphalt. *ASCE's Journal of Transportation Engineering, Part B: Pavements*, (2017).

Baqersad, M., E. A., Sayyafi, and H. M., Bak. State of the Art: Mechanical Properties of Ultra-High Performance Concrete. *Civil Engineering Journal*, 3(3), 2017, 190-198.

Bar-Gera, H. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from israel. *Transportation Research Part C*, 15, 2007, 380–391.

Barth, M., E. Johnston, and R. Tadi. Using GPS Technology to Relate Macroscopic and Microscopic Traffic Parameters. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1520, Transportation Research Board of the National Academies, Washington, D.C., 1996, pp89-96.

Berlingerio, M., F. Calabrese, G. Di Lorenzo, R. Nair, F. Pinelli, M. L. Sbodio, and IBM Research Ireland. AllAboard: A System for Exploring Urban Mobility and Optimizing Public Transport Using Cellphone Data. Submitted for presentation at the 93rd Annual Meeting of the Transportation Research Board, 2014.

Bestplaces, <http://www.bestplaces.net/people/city/florida/miami>, Accessed January 29, 2015

Bohte, W., and K. Maat. Deriving and Validating Trip Destinations and Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the Netherlands. In *Transportation Research Part C: Emerging Technologies*, Volume 17 Issue 3, 2009, pp 285-297.

Bohte, W., K. Maat, W. Quak, and T. Delefit, A Method for Deriving Trip Destination and Modes for GPS-Based Travel Surveys. *Research in Urbanism Series Urbanism on Track: Application of tracking technologies in urbanism*. 2008, pp127-143.

Buliung, R. N., Roorda, M. J. and Rimmel, T. K. Exploring spatial variety in patterns of activity-travel behaviour: Initial results from the toronto travel-activity panel survey (ttaps). *Transportation Research Part A*, 35, 2008, 697–722.

Caceres, N., Romero, L. M., Benitez, F. G. and Castillo, J. M. D. Traffic flow estimation models using cellular phone data. *IEEE Transactions on Intelligent Transportation Systems*, 13, 3, 2012, 1430-1441.

Calabrese, F., M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti. Understanding Individual Mobility Patterns from Urban Sensing Data: A Mobile Phone Trace Example. In *Transportation Research Part C: Emerging Technologies Volume 26*, 2013, pp 301-313.

Casello, J. M., and V. Usyukov. Modeling Cyclists' Route Choice Based on GPS Data. *Transportation Research Record: Journal of the Transportation Research Board*, No.2430, Transportation Research Board of the National Academies, Washington,D.C., 2014, pp155-161.

Chen, C., L. Bian, and J. Ma. From traces to trajectories: How well can we guess activity locations from mobile phone traces. *Transportation Research Part C*, 46, 2014, pp. 326-337.

Cheng, X., W. Li, F. Jia, D. Yang, Z. Duan. Analyzing human activity patterns using cellular phone data: a case study of Jinhe new town in Shanghai. *Transportation Research Board*. Washington, D.C, 2012.

Çolak, S., L. P Alexander, B. G. Alvim, S. R. Mehndiretta, and M. C. Gonzalez. Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations and Opportunities. Submitted for Publication and Presentation at the 94th Annual Meeting of the Transportation Research Board, 2015.

Dai, X., M. A. Ferman, and R. P. Roesser. A Simulation Evaluation of a Real-Time Traffic Information System Using Probe Vehicles. In *Proceedings of IEEE Intelligent Transportation Systems 1*, 2003, pp 475-480.

Das, R. D., Nicole Ronald, and S. Winter. A Simulation Study on Automated Transport Mode Detection in Near-Real Time using a Neural Network. In *Proceedings of Research@Locate'15*, Brisbane, Australia, 2015.

Demissie, M. G., G. H. de Almeida Correia, C. Bento. Intelligent Road Traffic Status Detection System through Cellular Networks Handover Information: An Exploratory Study. *Transportation Research Part C: Emerging Technologies*, 2013, pp76-88.

Deng, Z., and M. Ji. Deriving Rules for Trip Purpose Identification from GPS Travel Survey Data and Land Use Data: A Machine Learning Approach. In *Traffic and Transportation Studies 2010*, 2010, pp 768-777.

Dhakar, N. S., and S. Srinivasan. Route Choice Modeling Using GPS-Based Travel Surveys. *Transportation Research Record: Journal of the Transportation Research Board*, No.2413, Transportation Research Board of the National Academies, Washington,D.C., 2014, pp65–73.

Fang, J., M. Xue, and T. Z. Qiu. Anonymous Cellphone-Based Large Scale Origin-Destination Data Collection: Case Studies in China. Submitted for Presentation at the 93rd Annual Meeting of the Transportation Research Board, 2014.

Federal Communication Commission. 911 Wireless Services. 2014. <https://www.fcc.gov/guides/wireless-911-services>. Accessed July2015.

FIU Research. Institutional Review Board (IRB), 2015, <http://research.fiu.edu/irb/index.html>. Accessed on 09/24/2015.

Gkiotsalitis, K., F. Alesiani, and R. Baldessari. Educated Rules for the Prediction of Human Mobility Patterns based on Sparse Social Media and Mobile Phone Data. Presented at the 93rd Transportation Research Board Meeting, Washington DC, 2014.

Gonzalez, M., C. Hidalgo, and A. L. Barabasi. Understanding Individual Human Mobility Patterns. In *Nature* Volume 453, 2008, pp 779–782.

Hanson, S. and Huff, J. Assessing day-to-day variability in complex travel patterns. *Transportation Research Record: Journal of the Transportation Research Board*, 891, 1982, 18-24.

Hariri, A. O., A. Elsayed and O. A. Mohammed. An Integrated Characterization Model and Multiobjective Optimization for the Design of an EV Charger’s Circular Wireless Power Transfer Pads. *IEEE Transactions on Magnetics*, vol. 53, no. 6, June 2017pp. 1-4.

Hariri, A. O., T. Youssef, A. Elsayed and O. A. Mohammed. A Computational Approach for a Wireless Power Transfer Link Design Optimization Considering Electromagnetic Compatibility. *IEEE Transactions on Magnetics*, vol. 52, no. 3, March 2016pp. 1-4.

Herrera, J., D. Work, R. Herring, X. Ban, Q., Jacobson, and A. Bayen. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C*, 18, 2010, pp. 586-583.

Hossan, M., H. Asgari, and X. Jin. Trip Misreporting Forecast Using Count Data Model in a GPS Enhanced Survey. *Transportation*, 2017, pp.1-14. DOI: 10.1007/s11116-017-9782-2.

Hosseiniou, M. H., A. Massahi, and M. V. Aliabadi. Developing Tehran Vehicles' Air Pollution Macroscopic Models and Presenting the Method for Estimating Pollutants' Emission Rates in Urban Networks. *Science Series Data Report*, 2012, 4(4).

Hosseiniou, M. H., E. Balal, A. Massahi, and I. Ghiasi. Developing optimal zones for urban parking spaces by Arc GIS and AHP. *Indian Journal of Science and Technology*, 2012, 5(11), 3618-3622.

How-To Geek, <http://www.howtogeek.com/195647/googles-location-history-is-still-recording-your-every-move>, Accessed October 4, 2015

Huff, J. O. and Hanson, S. Repetition and variability in urban travel. *Geographical Analysis*, 18, 2, 1988, 97-114.

Huntsinger, L. F. and K. Ward. Using Mobile Phone Location Data to Develop External Trip Models. Submitted for Publication and Presentation at the 94th Annual Meeting of TRB, Washington, D.C., 2015.

International Telecommunication Union. ITU Statistics. 2015. <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>. Accessed July2015.

Iqbal, M. S., C. F. Choudhury, P. Wang, and M. C. González. Development of Origin-Destination Trip Matrices Using Mobile Phone Call Data. *Transportation Research Part C: Emerging Technologies*, 2014, pp63-74.

Jahangiri, A. and H. Rakha. Developing a Support Vector Machine (SVM) Classifier for Transportation Mode Identification using Mobile Phone Sensor Data. Submitted for

presentation/publication at the 93rd Annual Meeting of the Transportation Research Board, 2014.

Jarv, O., R., Ahas, and F., Witlox. Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. In *Transportation Research Part C* 38, 2014, pp 122-135.

Jin, X., H. Asgari, and M. Hossan. Understanding Trip Misreporting in Household Travel Surveys by Comparing GPS-Assisted and Diary-Based Samples. *CICTP 2014*: pp. 3401-3412. DOI: 10.1061/9780784413623.326.

Joh, C.-H., S.T. Doherty, and J.W. Polak. Analysis of Factors Affecting the Frequency and Type of Activity Schedule Modification. *Transportation Research Record: Journal of the Transportation Research Board*, 1926, 2005, 19-25.

Kitamura, R. and T., Van, T. Regularity and irreversibility of weekly travel behavior. *Transportation*, 14, 1987, 227-251.

Kitamura, R., Chen, C. and Narayanan, R. Traveler destination choice behavior: Effects of time of day, activity duration, and home location. *Transportation Research Record: Journal of the Transportation Research Board*, 1645, 1, 1998, 76-81.

Lee, K., S. Hong, and S. Kim. SLAW: A Mobility for Human Walks. Proceeding for publication in the IEEE INFOCOM 2009.

Levinson, D. and Z. Shanjiang. A Portfolio Theory of Route Choice. *Transportation Research Part C: Emerging Technologies*, 2013, pp232-243.

Li, H. Using Mobile Phone Data to Analyze Origin-Destination Travel Flow Dynamics for City of Pasadena, CA and Surrounding Area. Submitted for Publication and Presentation at the 94th Annual Meeting of the TRB, 2015.

Li, W., X. Cheng, D. Yang, Z. Duan, and G. Guo. Mining Rail-transit Use Patterns from Mobile Phone Data: Case Study of Three Communities in Shanghai. Submitted for Presentation and Publication to the 94th Annual Meeting of TRB, 2015.

Liang, X., X. Zheng, W. Lv, T. Zhu, and K. Xu. The Scaling of Human Mobility by Taxis is Exponential. *Physica A: Statistical Mechanics and its Applications*, 2012, pp2135-2144.

Liao, C. Generating Reliable Freight Performance Measures with Truck GPS Data: Case Study in Twin Cities Metropolitan Area, Minnesota. Transportation Research Record: Journal of the Transportation Research Board, No. 2410, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp21–30.

Lindsey, G., S. Hankey, X. Wang, J. Chen, and A. Gorjestani. Feasibility of Using GPS to Track Bicycle Lane Positioning. CTS Report No.13-16. University of Minnesota Intelligent Transportation Systems Institute Center for Transportation Studies. 2013.

Liu, H. X., Danczyk, A., Brewer, R. and Starr, R. Evaluation of cell phone traffic data in minnesota. Journal of the Transportation Research Board, No. 2086, 2008, 1-7.

Lu, H., J. Yang, Z. Liu, N. D. Lane, T. Choudhury, and A. T. Campbell. The Jigsaw Continuous Sensing Engine for Mobile Phone Applications, 2010. <http://www.cs.dartmouth.edu/~campbell/papers/jigsaw.pdf>. Accessed July2015.

Lu, X., Wetter, E., Bharti, N., Tatem, A. J. and Bengtsson, L. Approaching the limit of predictability in human mobility. Scientific Reports, 2013, DOI:10.1038/srep02923.

Ma, Y., J. van Dalen, C. de Blois, and L. Kroon. Estimation of Dynamic Traffic Densities for Official Statistics: Combined Use of Data from Global Positioning System and Loop Detectors. Transportation Research Record: Journal of the Transportation Research Board, No.2256, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp104-111.

Massahi, A., H. Ali, F. Koochifar, and M. Mohammadafzali. "Analysis of pavement raveling using smartphone." Transportation Research Board 95th Annual Meeting. 2016, No. 16-6155.

Massahi, A., M. Hadi, Y. Xiao, and X. Chen. Improved Model for Estimating Incident Impact on Urban Street Travel Time with Consideration of Upstream Intersection Capacity Reduction. Paper presented at the 95th Annual Meeting of the Transportation Research Board, National Research Council, Washington, D.C., January, 2016.

Massahi, A., M. Hadi, M. Adriana Cutillo, and Y. Xiao. Estimating the Capacity Impacts of Urban Street Incidents. Transportation Research Record (TRR) Journal, Journal of the Transportation Research Board. No. 2615, 2017, pp. 55–61

Moiseeva, A., J. Jessurun, and H. Timmermans. Semi-Automatic Imputation of Activity-Travel Diaries Using GPS Traces, Prompted Recall and Context Sensitive Learning

Algorithms. Presented at the 89th Annual Meeting of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., 2010.

Nitsche P., P. Widhalm, S. Breuss, and P. Maurer. A Strategy on How to Utilize Smartphones for Automatically Reconstructing Trips in Travel Surveys. *Procedia Social and Behavioral Sciences*, 2012, pp1033-1046.

Nitsche, P., P. Widhalm, S. Breuss, N. Brändle, and P. Maurer. Supporting Large-Scale Travel Surveys with Smartphones—A Practical Approach. *Transportation Research Part C: Emerging Technologies*, 2014, pp212-221.

Nour, A., J. Casello, and B. Hellenga. Developing and Optimizing a Transportation Mode Inference Model Utilizing Data from GPS Embedded Smartphones. Submitted for presentation/publication at the 94th Annual Meeting of the Transportation Research Board, 2015.

Pang, L. X., S. Chawla, W. Liu, and Y. Zheng. On Detection of Emerging Anomalous Traffic Patterns Using GPS Data. In *Data & Knowledge Engineering*, 2013, pp 357-373.

Pas, E. I. and Koppelman, F. S. An examination of the determinants of day-to-day variability in individuals' urban travel behavior. *Transportation*, 14, 1987, 3-20.

Pereira, F., C. Carrion, F. Zhao, C. D. Cottrill, C. Zegras, and M. Ben-Akiva. The Future Mobility Survey: Overview and Preliminary Evaluation. *Proceedings of the 10th International Conference of Eastern Asia Society for Transportation Studies*, Taipei, Taiwan, 2013.

Ranjitkar, P., T. Nakatsuji, and M. Asano. Performance Evaluation of Microscopic Traffic Flow Models with Test Track Data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1876, TRB, National Research Council, Washington, D.C., 2004, pp90–100.

Reddy, S., J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Determining Transportation Mode on Mobile Phones. *Proceedings of the 12th IEEE International Symposium on Wearable Computers*, 2008, p25-28.

Rojas, M., Sadeghvaziri, E., Jin, X., “A Comprehensive Review of Travel Behavior and Mobility Pattern Studies Using Mobile Phone Data,” *Transportation Research Record: Journal of the Transportation Research Board*, No. 2563, Washington, D.C., 2016, pp. 71–79.

Rokib, S. A., M. A. Karim, T. Z. Qiu, and A. Kim. Origin-Destination Trip Estimation from Anonymous Cell Phone and Foursquare Data Submitted for Presentation at the 94th Annual Meeting of TRB, 2015.

RSG. Multiday GPS Travel Behavior Data for Travel Analysis. Publication FHWA-HEP-015-026. FHWA, U.S. Departments of Transportation, 2015.

Sadeghvaziri, E., M. Rojas, and X. Jin. Exploring the Potential of Mobile Phone Data in Travel Pattern Analysis. Transportation Research Record: Journal of the Transportation Research Board, No. 2594, Washington, D.C., 2016, pp 27–34.

Sadeghvaziri, E., K. Haleem, P. Alluri, H. Fartash, and A. Gan. Effects of Different Guide Signs on Driver Behavior at Express Lane Entrance: A Driving Simulation Study. Transportation Research Board 95th Annual Meeting. 2016, No. 16-5555.

Schönfelder, S., and U. Samaga. Where do you want to go today? More observations on daily mobility. From proceedings of 3rd Swiss Transport Research Conference, Monte Verita/Ascona, Italy, 2003.

Schüssler, N., and K. Axhausen. Identifying Trips and Activities and Their Characteristics from GPS Raw Data Without Further Information. Presented at the 88th Annual Meeting of the Transportation Research Board. Transportation Research Board of the National Academies, Washington, D.C., 2008.

Shafieifar, M., M. Farzad, and A. Azizinamini. Experimental and numerical study on mechanical properties of Ultra High Performance Concrete (UHPC). Construction and Building Materials. 2017, 156, 402-411.

Sharman, B., and M. Roorda. Analysis Tool to Process Passively Collected GPS Data for Commercial Vehicle Demand Modeling Applications. Innovations in Freight Demand Modeling and Data Symposium, Transportation Research Board of the National Academies, Washington, D.C., 2010.

Song, C., Qu, Z., Blumm, N. and Barabási, L.-L. Limits of predictability in human mobility. Science, 327, 5968, 2010, 1018-1021.

Spissu, E., I. Meloni, and B. Sanjust. Behavioral Analysis of Choice of Daily Route with Data from Global Positioning System. Transportation Research Record: Journal of the Transportation Research Board, No.2230, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp96-103.

Stenneth, L., O. Wolfson, P. S. Yu, and B. Xu. Transportation Mode Detection Using Mobile Phones and GIS Information. GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems, 2012, pp54-63.

Stopher, P., C. FitzGerald, and J. Zhang. Search for A Global Positioning System Device to Measure Person Travel. Transportation Research Part C: Emerging Technologies Volume16, 2008, 350-369.

Sunio, V., and J. Schmöcker. Can we promote sustainable travel behavior through mobile apps? Evaluation and review of evidence." International Journal of Sustainable Transportation 11.8, 2017, 553-566.

Support Google, <https://support.google.com/gmm/answer/4426813?hl=en> (access 10.04.2015)

Tawfik, A. M., and H. A. Rakha. Network Route-Choice Evolution in a Real-World Experiment: Necessary Shift from Network- to Driver-Oriented Modeling. Transportation Research Record: Journal of the Transportation Research Board, No.2322, Transportation Research Board of the National Academies, Washington,D.C., 2012, pp70-81.

Thakur, A., A. B. Zanjani, A. R. Pinjari, and J. Short. Development of Algorithms to Convert Large Streams of Truck GPS Data into Truck Trips. Presented at the 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.

United States of America Department of Defense. Global Positioning System Standard Positioning Service Performance Standard. September2008. <http://www.gps.gov/technical/ps/2008-SPS-performance-standard.pdf>. Accessed July 2015.

Vacca, A., and I. Meloni. Understanding Route Switch Behavior: An Analysis Using GPS Based Data. In Transportation Research Procedia 5, 2015, pp 56-65.

Venter, C. J., and J. W. Joubert. Use of Multisource Global Positioning System Data to Characterize Multiday Driving Patterns and Fuel Usage in a Large Urban Region Transportation Research Record: Journal of the Transportation Research Board, No. 2338, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp1-10.

Wang M. E., S. D. Schrock, N. V. Broek, and T. Mulinazzi. The Feasibility of Using Cellular Phone Location Data in Traffic Survey on Inter-City Trips. Submitted for Presentation at the 92nd TRB Annual Meeting, 2013.

Wang, M., C. Chen, and J. Ma. On Making More Efficient Location Prediction. Transportation Research Board. Washington, D.C, 2015.

Wang, M., C. Chen, and J. Ma. Time-of-Day Dependence of Location Variability: An Application of Passively-Generated Mobile Phone Dataset. Transportation Research Board. Washington, D.C, 2015.

Wang, X., H. Dong, Y. Zhou, K. Liu, L. Jia, and Y. Qin. Travel distance characteristics analysis using call detail record data. In Control And Decision Conference (CCDC), IEEE, 29th May, Chinese, 2017, pp. 3485-3489.

Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/List_of_countries_by_smartphone_penetration. Accessed July 30, 2015

Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/Doral,_Florida . Accessed July 23, 2016

Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/Spherical_law_of_cosines . Accessed June 30, 2016

Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/Spherical_law_of_cosines . Accessed June 30, 2016

Wolf, J., M. Oliveira, and M. Thompson. Impact of Underreporting on Mileage and Travel Time Estimates: Results from Global Positioning System Enhanced Household Travel Survey. Transportation Research Record: Journal of the Transportation Research Board, No.1854, Washington, D.C., 2003, pp189-198.

Wolf, J., W. Bachman, M.S. Oliveira, J.A. Auld, A. Mohammadian, and P. Vovsha. Applying GPS Data to Understand Travel Behavior: Background, Methods, and Tests. NCHRP-Report 775, Transportation Research Board,20

Yin, E. and T. Z. Qiu. Using GPS-Enabled Smartphones and Geofence to Capture Traffic Data on Urban Freeways and Arterials. Submitted for presentation/publication at the 94th Annual Meeting of the Transportation Research Board, 2015.

Yin, E., P. Li, J. Fang, and T. Qiu. Evaluation of Vehicle Positioning Accuracy using GPS-Enabled Smartphones. Transportation Research Board. Washington, D.C, 2014.

Ythier, J., J., Walker, and M. Bierlaire. THE INFLUENCE OF SOCIAL CONTACTS AND COMMUNICATION USE ON TRAVEL BEHAVIOR: A SMARTPHONE-BASED STUDY. Transportation Research Board. Washington, D.C, 2013.

Yue, Y., T. Lan, A. Yeh, and Q. Li. Zooming into individuals to understand the collective: A review of trajectory-based travel behavior studies. *Travel Behaviour and Society*, 1, 2014, pp. 69-78.

Zanjani, A. B., A. R. Pinjari, M. Kamali, A. Thakur, J. B. Short, V. Mysore, and F. Tabatabaee. Estimation of Statewide Origin-Destination Truck Flows Using Large Streams of GPS Data: Application for Florida Statewide Model. Presented at the 94th meeting of Transportation Research Board, Washington, D.C., 2015.

Zhang, L., S. Dalyot, D. Eggert, and M. Sester. Multi-Stage Approach to Travel-Mode Segmentation and Classification of GPS Traces. International Society for Photogrammetry and Remote Sensing workshop on geospatial data infrastructure: From data acquisition and updating to smarter services, 2011, pp87-93.

Zheng, Y., Q. Li, Y. Chen, X. Xie, and W.Y. Ma. Understanding Mobility Based on GPS Data. Proceedings of the 10th International Conference on Ubiquitous Computing, Seoul, Korea, 2008, pp312-321.

VITA

EAZAZ SADEGHVAZIRI

EDUCATION

- 2004 - 2008 B.Sc., Civil Engineering
Azad University, Tehran Central Branch,
Tehran, Iran
- 2009 - 2012 M.Sc., Civil Engineering (Transportation Planning)
Azad University, Science and Research Branch,
Tehran, Iran
- 2013 – Present Ph. D. Student
Department of Civil and Environmental Engineering
Florida International University
Miami, Florida

- Secretary, ITE Student Chapter at FIU, 2013 – 2014
- Membership Chair, ITE Student Chapter at FIU, 2015 – 2016
- Vice-President of ISO at FIU, 2015-2016

PUBLICATIONS AND PRESENTATIONS

- **Sadeghvaziri, E.**, and Jin, X., “*Driving Activity Locations Inferred from Smartphone Data*,” Oral presentation at The 16th TRB National Transportation Planning Applications Conference, Raleigh, North Carolina, May 14-18, 2017.
- **Sadeghvaziri, E.**, Rojas, M., and Jin, X., “*Exploring the Potential of Mobile Phone Data in Travel Pattern Analysis*,” Transportation Research Record: Journal of the Transportation Research Board, No. 2594, Washington, D.C., 2016, pp. 27–34.
- Rojas, M., **Sadeghvaziri, E.**, and Jin, X., “*A Comprehensive Review of Travel Behavior and Mobility Pattern Studies Using Mobile Phone Data*,” Transportation Research Record: Journal of the Transportation Research Board, No. 2563, Washington, D.C., 2016, pp. 71–79.

- **Sadeghvaziri, E.**, Haleem, K., Alluri, P., Fartash, H., and Gan, A., “*Effects of Different Guide Signs on Driver Behavior at Express Lane Entrance: A Driving Simulation Study*,” Poster presentation at The Transportation Research Board, 95th Annual Meeting, Washington D.C., January 2016.
- **Sadeghvaziri, E.**, Ortega, L., Raihan, A., and Jin, X., “*Application of GIS in Road Crash Analysis: Two Case Studies of Miami*,” Poster presentation at The Graduate Student Appreciation Week (GSAW), Graduate & Professional Student Committee (GPSC), Florida International University (FIU), Florida, April 2015.
- Abdi, A., Mehara, A., Monajjem, S., and **Sadeghvaziri, E.**, “*Simulation Modeling of Dynamic Response of Vehicles to Different Types of Speed Control Humps*,” American Society of Civil Engineers (ASCE) T&DI Congress 2014: pp. 533-542.
- Lavasani M., Fartash H., **Sadeghvaziri, E.**, “*Adjusting HCM 2010 Freeway Density Formula to Consider Ramp Metering in Merge Areas*,” Oral presentation at The 12th International Conference on Traffic Engineering, International Conference Center of Tehran, Iran, Mar 1-2, 2016.
- Mokhtarimousavi, S., Rahami, H., Saffarzade, M., and **Sadeghvaziri, E.**, “*Multi-Objective Mathematical Modeling for Aircraft Landing by Simultaneous Minimization of Time and Cost*,” Oral presentation at The International Conference on Operation Research – OR2013, Rotterdam, The Netherlands, September 2013.
- Rahimof, K., Mahmoudi, A., Enayati, I., **Sadeghvaziri, E.**, “*Assessing Parking Space Demand for different Land use of City of Oroumie*,” Oral presentation at The 7th National Congress on Civil Engineering, Zahedan, Iran, May 2013.
- Seyedhoseini, M., **Sadeghvaziri, E.**, Rahimof, K., and Mokhtarimousavi, S., “*Financial Analysis of Overhead Bridges on Decreasing Air Polluters and Fuel Consumption*,” Oral presentation at The 12th International Conference on Traffic and Transportation Engineering, Tehran, Iran, February 2013.
- Izadpanah, H., Abdi, A., Javanshir, H., **Sadeghvaziri, E.**, “*Locating and Prioritizing Construction of Pedestrian Crossings Paths Using Non-Coplanar*

Meta-heuristic Methods,” Poster Presentation at The 12th International Conference on Traffic and Transportation Engineering, Tehran, Iran, February 2013.