ASSESSING TRANSPORTATION EQUITY CONSIDERING INDIVIDUAL TRAVEL DEMAND AND THE FEASIBILITY OF TRIP MODE ALTERNATIVES

by

Utkuhan Genc

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Industrial Engineering

School of Industrial Engineering

West Lafayette, Indiana

May 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Hua Cai, Chair

School of Industrial Engineering

Environmental and Ecological Engineering

Dr. David R. Johnson

School of Industrial Engineering

Political Science

Dr. Konstantine "Nadia" Gkritza

Lyles School of Civil Engineering

Agricultural and Biological Engineering

Approved by:

Dr. Barrett S. Caldwell

To my family, Sharon, and the loveliest cat, Truffle

ACKNOWLEDGMENTS

Foremost, I would like to thank my thesis advisor Dr. Hua Cai of the School of Industrial Engineering at Purdue University. Her expertise and sincere compassion allowed me to create this work that I can be proud of. She consistently allowed this thesis to be my own work but steered me in the right direction whenever she thought I needed it. I want to further extend my thanks to Dr. David R. Johnson of the School of Industrial Engineering and Dr. Konstantine Gkritza of the Lyles School of Civil Engineering, whose help was much appreciated. Without their flexibility and guidance, this thesis would not be complete. I would also like to thank Dr. Jie Shan of the Lyles School of Civil Engineering at Purdue University for sharing their data with me and allowing me to make this thesis unique from the existing works of others.

I would also like to acknowledge Hao Luo of Environmental and Ecological Engineering at Purdue University. His role in creating this thesis is immeasurable. He gave me precious comments on this thesis. I also would like to show my appreciation for Kendrick and Kanaan Hardaway for being willing to listen to me talk about my concerns and ideas about this thesis no matter the time.

I want to thank my family for providing me with the financial and emotional support necessary for chasing my dreams. Last, I must express my very profound gratitude to my girlfriend, Sharon Lee, who was my harshest critic but also the most exuberant supporter through researching and writing this thesis. All these people helped make this accomplishment possible. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Transportation access is an important indicator of the quality of life and if it is inequitable, it will limit the work, leisure, and other essential opportunities for people and worsen the access for the disadvantaged groups. In the U.S., increased auto-dependency and the lack of other feasible alternative transportation modes exacerbate the negative impacts of this inequity, especially for the people without automobiles. The transportation equity in terms of the number of feasible transportation mode alternatives to serve a trip (i.e. mobility option equity) has not been extensively evaluated in the literature. Existing studies mainly analyzed the access to transportation infrastructures (e.g., bus stops, bike lanes, shared bike stations) based on the proximity at the zonal level. However, having access to a certain trip mode based on proximity does not necessarily add to the mobility option equity. First, mismatch may exist between the infrastructure and an individual's travel demand. For example, if someone lives closely to a bus station but the bus route that can be accessed does not align with this person's trip destination, they will not be able to use bus as a feasible mode for this trip. Second, existing accessibility-based studies often lack consideration of the trip feasibility (in terms of cost, quality, and safety) of using transportation infrastructures at the route level. For example, if a walking trip route is generated without considering the existence of sidewalks, the individual might have to walk on a unsafe busy road. In this case they will not be able to walk to satisfy their travel demand. Therefore, better transportation equity metrics concerning the feasibility of using transportation infrastructures to serve individuals' travel demands are needed.

To address this gap, this thesis defined the "travel-demand-relevant access" (mobility-need-relevant access) metric to evaluate transportation access in the context of individual travel demands and route-level infrastructure constraints and developed a framework to use GPS data to quantify the proposed metric for transportation equity analysis. Assessing which transportation modes are feasible alternatives to serve a trip, requires trip-level disaggregated travel demand data and detailed transportation infrastructure information. The recent development of information and communication technologies and open data efforts provide unprecedented opportunities for such trip-level analysis. With these developments it is now possible to evaluate the feasibility of a mode both the cost- and quality-based measures. The cost-based method estimates the monetary and time cost of using each mobility option and compares it with prominent trip mode (car) to examine

“forced car use” concerning the travel demand. The quality-based method comprises accessibility and mobility-based performance measures to evaluate the feasibility of a certain trip mode regarding the ease of use and safety with relation to the infrastructure characteristics. The mobility options/alternatives deemed feasible with these two methods were used in the equity analysis, where the travel-demand-relevant access on the spatial and sociodemographic level was evaluated.

The proposed framework was applied to the Indianapolis Metropolitan Planning Area (MPA) as a case study. The key insights of this study can be listed as (1) it is important to consider travel-demand-relevant access to evaluate transportation equity because we found that 40% of the trips that were identified as accessible by public transit are not feasible when travel-demand-relevant access is considered; (2) suburban areas on average have 12% less mobility options available compared with the urban core which forces high car ownership in these areas; amd (3) people with non-college educational attainment, households with more crowded rooms, and larger families are the negatively impacted disadvantaged groups while census block groups with high composition of white middle-class suburban families have the lowest number of options (1.5 on average) available.

The suburban populations with a low number of mobility options (with a vehicle) are not necessarily at a disadvantage in terms of mobility option equity, since suburban areas are by design made to be car dependent. However, the lower number of feasible mobility options in these areas possesses a risk for the future if the consequences are not evaluated carefully. In terms of urban migration, if out-migration from the urban core to suburban areas keeps increasing as the pandemic trend suggests, the forced car ownership in suburban areas could increase and create/worsen transport deserts. This increase in vehicle ownership contradicts equity and environmental goals regarding transportation. If we observe an increase in the suburban to urban core migration trend, it can force disadvantaged groups to move into suburban areas because of gentrification and increasing prices. These disadvantaged groups could suffer from the limited amount of mobility options in suburban areas, since their access to opportunities would decrease.

1. INTRODUCTION

The literature has extensively recognized the transportation quality and mobility option access as one of the key indicators of quality of life in the literature. In the National Cooperative Highway Research Program (NCHRP) report 184, transportation has been described as “being fundamental to achieving national goals, sustaining community values, and promoting personal well-being” (NCHRP, 2011). Until recently, transportation equity was only concerned with driver-related aspects such as fairness of transportation funding, while the impacts on non-drivers were considered with little concern (Litman, 2021). The long-term projects such as construction of highways did not include the external impacts, such as air pollution, reduced access, and livability of a certain area. As Reichmuth (2019) pointed out, the air pollution burden due to transportation (mainly highways) is not equally shared among different sociodemographic groups. Additionally, access to economic opportunities is distributed inequitably both on spatial and socioeconomic levels (Jones & Lucas, 2012; Martens et al., 2012; Pyrialakou et al., 2016). If the spatial and sociodemographic distribution of impacts of transportation (such as air quality and access) is not equitable, unemployment and social exclusion are likely to rise while the quality of life of disadvantaged groups is likely to decline (Jones & Lucas, 2012; Preston & Rajé, 2007; Pyrialakou et al., 2016). The most commonly and severely affected groups by this inequity are the individuals without automobiles, low-income groups, people with disabilities, and people of color and minorities (Golub & Martens, 2014). Furthermore, providing equitable transportation to the individuals and neighborhoods with disadvantages was recently recognized as a current federal issue by the Biden administration with Executive Order (EO) 13985 “Advancing Racial Equity and Support for Underserved Communities through the Federal Government” (Executive Order No. 13985, 2021). In addition, transportation-related capabilities (bodily integrity, affiliation, play, and control over one’s environment) from Nussbaum’s Central Human Functional Capabilities (CHFC) depicted transportation as a basic human right and freedom (Nussbaum, 2001). Lastly, adequate and fair access to transportation was highlighted in the United Nations Sustainable Development Goals (UN SDGs) at the goal 11.2 as “By 2030, provide access to safe, affordable, accessible and sustainable transport systems for all.” (United Nations, 2015).

The most common description of transportation equity comes from Litman (2021): “Equity (also called justice and fairness) refers to the distribution of transportation impacts (benefits and costs) and whether they are appropriate". This definition is generally accepted without careful consideration and evaluation of their appropriateness for the specific case (Lewis et al., 2021). However, it is important to identify underlying social justice philosophies because how the equity is desribed has an important effect on the analysis (Lewis et al., 2021). Another common definition comes from Karel Martens's 2017 book "Transport Justice", as he defended the thesis that "a transportation system is fair if, and only if, it provides sufficient level of accessibility of all under most circumstances", hence implying the sufficientarian (discrete cut-off between sufficient versus insufficient) and prioritarian (continuous curve of need) justice theories (Martens, 2017). Both approaches are advocating for reducing the "misery" by distributing resources favorably toward already suffering groups, rather than aiming for total equality Martha Nussbaum's advancement of Amartya Sen's Capabilities Approach also adopted a similar school of thought with transportation justice and recognized the improvements on the threshold for "misery" is more important than trying to aim for total equality (Nussbaum, 2001). Additionally, it is important to recognize amending the situation for these worse off groups might have unwanted effects on the urban space and further exacerbate the inequities through gentrification (Sheller, 2018, p.35). (*Please refer to* Lewis et al., 2021; Martens, 2017 *for more information about social justice theories and their applications in transportation)*.

Various research papers have studied the implications of equity on transportation systems with the two main isolated perspectives: mobility versus accessibility (Litman, 2003). In general, we can describe mobility as the ability of a person to physically travel from one location to another (Lefrançois, 1998). The mobility based measures, such as cost, in vehicle and out of vehicle travel time, are the most common variables used when developing these mode choice models (Hillel et al., 2021). These types of studies rely heavily on stated preference surveys or making underlying assumptions about people's travel behavior, thus limiting their applicability (Guo et al., 2020). Since mobility can also be described as as "how far you can go in a given amount of time?", it favors faster travel modes. This can limit the benefits of non-motorized travel compared to the motor vehicles if only mobility based measures are used (Litman, 2021). This issue is also exacerbated by the lack of standardized and generalized values for benefits from non-motorized travel that can be adopted in an individual mode choice model. These mode-choice models assign a larger intercept value in utility functions for non-motorized travel to account for the benefits from faster travel (such as in Reck et al. (2022)), however, they use sociodemographic characteristics as input which may limit their transferability to other study areas and population. To the best of our knowledge, there are no travel mode choice models that used disaggregated (per-mile, per-hour) values for benefits from non-motorized travel. This is likely because of the benefits of non-motorized modes (walking, bike, etc.), such as the safety benefits (accidents prevented), health benefits (healthier individuals, less missed workdays), and environmental benefits (reduced greenhouse gas emissions), being considered through total annual vehicle miles rather individual trip level, thus making it difficult to convert per-mile/per-hour benefits for a certain trip (Caltrans, 2022). The disaggregated values can better help evaluate the utility from each individual trip when no sociodemographic characteristics of the individual is known, thus making it easier to compare potential mobility options that can be used to satisfy the travel demand.

On the other hand, accessibility-based studies are concerned with the ability to reach certain opportunities or places in a given amount of time (how many opportunities you can get to in a given amount of time). For example, Páez et al. (2012) developed a two-component approach to evaluate accessibility to day-care centers which considered distribution of opportunities and cost of travel. Other accessibility studies, such as Golub & Martens (2014), Minocha et al. (2008), Pyrialakou et al. (2016) evaluate "potential" opportunities that are reachable from a certain location (how many opportunities are reachable from home) or they evaluate the destination that is reachable in certain time threshold (e.g., can people reach work locations in a given time). Although these studies are very beneficial in identifying general patterns, they lack the spatial resolution to understand individual level impacts. While quality of service metrics help identify if a mobility option is equitable among different groups in ideal situations, they cannot consider if the actual destination can be reached with adequate options because these studies lack the spatial resolution to understand the accessibility in the context of travel demand (Kamruzzaman & Hine, 2011, 2012). In general, accessibility-based studies do not consider multiple trip modes (Golub & Martens, 2014; Karner, 2018; Minocha et al., 2008) and even when they do, the infrastructure conditions are not accounted for (there are some studies that consider quality of service such as Welch & Mishra (2013) but they are rare in the U.S.).

The transportation equity literature agrees upon that an automobile-oriented development can cause unequal benefits and opportunities for some disadvantaged groups who do not have a car or cannot drive because of other reasons (Jones & Lucas, 2012; Litman, 2021). The forced automobile ownership, or willingness to purchase an automobile to avoid consequences, was explored to assess the high vehicle ownership in urban low-income groups and rural areas as a limiting factor rather than a prospering advantage in Melbourne (Currie & Senbergs, 2007) and in the United Kingdom (UK) and Germany (Mattioli, 2017). Since communities in the United States (U.S.) are usually very automobile-dependent, the groups without automobile access face disadvantages and potentially suffer from forced-automobile ownership (Kamruzzaman & Hine, 2012; Pyrialakou et al., 2016). Furthermore, unequal share of air pollution burdens among sociodemographic groups are mainly driven by this automobile oriented transportation development which signals the need for better mobility options (Reichmuth, 2019). Newer and cleaner mobility options can help reduce these unequal environmental and health related impacts while helping disadvantaged communities to improve their economic and social outcomes (Cohen & Cabansagan, 2017). However, the existing literature on transportation equity generally only analyzed one individual travel mode, such as public transit, bike sharing, or walking, limiting their understanding of the system level impacts of the transportation network. In the past, the studies that concerned with the benefits and costs of highway projects did not account for negative impacts on other modes (walking) for residents living along the proposed highway, thus degrading urban neighborhoods' walking access (Litman, 2021). Understanding system level impacts of the transportation network is important because not every option could be feasible for everyone. Simply having access to one mode of transportation does not ensure that an individual will have the means to fulfill their travel needs. Physical conditions (disability) or preferences can make this feasible trip mode infeasible in an individual context. Moreover, travel modes such as walking and biking are considered being better for health and climate, thus highlighting their importance for better quality of life (CARB, 2018; Vale et al., 2015). Therefore, simply having a car to perform a trip versus having an option to walk or bike to satisfy that demand does not show the same level of access. This is because only having a car as an option can be classified as "forced" vehicle ownership, while having options available brings freedom to people to choose their travel mode to satisfy their travel demand.

To the best of our knowledge, transportation equity studies have not used travel demand data with high spatial resolution (with exact locations of trip origin and destination known). Travel surveys used in the transportation equity studies only report trip characteristics, such as trip distance, trip duration, trip mode, trip purpose, and trip start time. National Household of Travel Survey (NHTS) is the most commonly used data source for U.S. based transportation equity studies (Federal Highway Administration, 2017). Other sources such as Longitudinal Employer-Household Dynamics Survey (LEHD) aggregate the travel demand information on lower spatial resolution (origin and destinations at the census blocks level) because of privacy concerns. This type of data has limitations because it may make the issues invisible when certain disadvantaged groups' travel needs were not met if they were grouped with majority advantaged neighbors. Their travel demands could be completely different, but due to aggregated trip data, this difference is not captured. In conclusion, the three gaps identified from the literature are only focusing on single travel mode, the use of isolated measures of either mobility or accessibility, and not evaluating feasibility in the context of disaggregated travel demand.

The goal of this thesis is to assess the transportation equity considering individual travel demand and feasibility of trip mode alternatives. Based on this, in this thesis we define the disadvantage as having a lesser number of feasible mobility options available for an individual to serve their travel demands. The research question that this study is trying to answer is: *"What is the relationship between the number of feasible mobility options available for an individual with the place he/she lives and his/her sociodemographic characteristics?"* To address the three research gaps, and answer the main research question, this study defined the "travel-demand-relevant access" (mobility-need-relevant access) metric to evaluate transportation access in the context of individual travel demands and route-level infrastructure constraints and developed a framework to use GPS data to quantify the proposed metric for transportation equity analysis. Travel-demand-relevant access can be defined as having feasible mobility options to satisfy an individual's observed travel demand concerning infrastructure constraints. For example, having travel-demand-relevant access for public transit can be described as having a bus stop in proximity which also has a transit route that goes to the destination in the time frame that the trip needs to happen. With the recent development of information and communication technologies and open data efforts it is now possible to evaluate the feasibility of a trip mode both the cost and quality-based measures. The cost-based method estimates the monetary and time cost of using each mobility option and compares it with prominent trip mode (car) to examine "forced car use" concerning the travel demand. The quality-based method comprises accessibility and mobility-based performance measures to evaluate the feasibility of a certain trip mode regarding the ease and of use and safety with relation to the infrastructure characteristics. The trip mode alternatives deemed feasible with these two methods were used in the equity analysis, where the travel-demand-relevant access on the spatial and sociodemographic level was evaluated.

The structure of the rest of the thesis is as follows. The detailed literature review is discussed in Chapter 2. The characteristics of the empirical study area of Indianapolis Metropolitan Planning Area (MPA) boundary and the rationale for it selection is discussed in Chapter 3. The data and methods used in this framework are introduced in Chapter 4. The overview of the framework developed is introduced in Chapter 4.1 while the data sources used in this study are explained in Chapter 4.2. The GPS data processing steps are explained in Chapter 4.3 while the methods used in assessing travel-demand-relevant access are discussed in Chapter 4.4. The methods used to evaluate transportation equity based on the feasible trip modes identified in this chapter are discussed in Chapter 4.5. The results are presented in three categories: validation of trip origin and destination identified by GPS data processing steps in Chapter 5.1, feasible mobility options and travel demand relevant access in Chapter 5.2, and equity analysis in spatial and sociodemographic level in Chapter 5.3. Last, Chapter 6 summarized the conclusion and discusses the limitation of this study and future research directions.

.

2. LITERATURE REVIEW

In this chapter, we first look at the existing literature on transportation equity, which differs in three categories. We can list these differences as the number of travel modes considered, the performance measures used, and the data source and data characteristics. To the best of our knowledge, the existing transportation equity studies that evaluate equity in the context of travel demand and regarding the infrastructure constraints are highly limited in the U.S. The existent studies usually focus on one mode of transportation (no consideration for forced use), use either mobility or accessibility-based measures, and lack the necessary disaggregated travel demand data. Below, we discuss all three gaps, while Table 2-1 presents an overview of the existing literature.

The transportation equity literature widely recognizes that spatial discrepancy, unequal levels of accessibility and mobility among different areas, is existent because of uneven distribution of proximity and availability of transportation services (Jones & Lucas, 2012; Martens et al., 2012). Accessibility-based studies are important in assessing access to opportunities (work, grocery store, day care, etc.). However these opportunities are not evaluated in the context of individual travel demands (the feasibility of trip mode is not known) (Kamruzzaman & Hine, 2011). On the other hand, mobility based studies can capture the feasibility (ease of travel) and forced ownership of a certain mode, but they cannot account for availability of the opportunities or mobility option alternatives (Kamruzzaman & Hine, 2011, 2012; Pyrialakou et al., 2016). Since the distinction between accessibility and mobility-based measures is not always very clear, it is hard to find studies that only evaluated mobility-based measures. Some studies also evaluate mobility-based measures in accessibility-based studies without addressing limitations of mobility-based measures. To make this distinction clear, this study categorized studies that evaluated equity only regarding the observed travel demand (aggregated or disaggregated) with no consideration for opportunities into the studies as mobility-based studies, studies that evaluated the equity in terms of opportunities as accessibility-based studies and lastly studies that explicitly evaluated both measures. Some studies define accessibility using the travel demand data, but if this travel demand data is aggregated and feasibility of individual trips is not considered (due to lack of origin and destination data), we still defined those studies as accessibility-based studies.

Accessibility-based studies are generally used to analyze the equity with public transit systems. Minocha et al. (2008) laid the groundwork for equity analysis for public transit using accessibility-based composite measure and availability of transit in census tract level. They simply used four categories to describe each tract level to show census tracts' transit availability and need In their analysis, they found that census tracts with low transit availability and low transit employment accessibility were the areas where automobile dependency is the highest while the areas with low transit availability and high transit employment accessibility to be in need for additional transit investments (Minocha et al., 2008). Griffin & Sener (2016) used accessibility-based measures for evaluating the spatial equity regarding income level. They used nine large auto-oriented cities' public transit system (including Indianapolis) to develop spatial screening tool using the measure of employed population able to access the census block group within a 45-minute transit commute from their home location as a percentage of total regional employed population. They found all cities had a mismatch between low-wage workers and accessibility to work destinations, meaning low-income groups (those who need it most) do not have more access to public transit compared to other groups. Karner (2018) explored the route level accessibility (the ease with destinations can be reached) measures for low and high-wage jobs using a gravity model and found that there is no difference in accessibility level. While the methods used in this thesis addressed some limitations of accessibility-based studies such as consideration of route availability based on the travel demand between census blocks, the only trips that were considered in this study were commuting trips that started during morning peak (7-9 am) and origin locations were transit stops. Additionally, they evaluated access only based on the public transit, comparing no other feasible mobility options. Golub & Martens (2014) looked at the differences between auto trips and public transit trips in terms of access to opportunities reachable within certain time limits from Traffic Analysis Zones (TAZ). They found almost all Bay Area neighborhoods to suffer from significant service gap (accessibility) between modes. Even though their findings provided valuable conclusions, the use of spatially aggregated travel demand data for measuring equity cannot capture the reality of the observed travel behavior. The opportunities are not based on the exact travel demand, which could change the level of access certain individuals have based on having feasible routes to their destinations. Smith evaluated accessibility of station based bike sharing systems (BSS). They evaluated large bike sharing system around the U.S. with more traditional accessibility (proximity base) analysis regarding the station placement and found that

75.4% of stations are in communities with low or lowest economic hardship. Similarly, Aside from BSS (Braun et al., 2019) looked at the bike-lane infrastructure at 23 U.S. cities and found that some disadvantaged groups (lower education attainment, higher proportions of Hispanic residents and lower composite socioeconomic index) have unequal access to bike-lane infrastructure. Interestingly, this disadvantage did not exist in groups with higher black population, lower median income, and higher poverty rate groups. McNeil (2011) and Winters et al. (2016) evaluated the access for active transportation modes: walking and biking using emerging metrics such as walk score (mainly destination based opportunities), bike score, and bikeway quality index (BQI). Walk score measure the comfort and feasibility of walking using population density and road characteristics, but the exact method is proprietary. Bike score on the other is based on four characteristics: bike lane availability, hilliness, destinations & connectivity (opportunities), and bicycle mode share (more people that bike the better the score) (Winters et al., 2016). McNeil (2011) used land topography and land use factor along with BQI which includes, motor vehicle speeds and volumes, number (width) of travel lanes, bicycle lane availability, quality of infrastructure and number of stops to assess feasibility of biking “bikeability”. McNeil (2011) found that there are spatial access differences between inner Portland and East Portland. They also showed lack of access to grocery stores as the main factor for low bikeability scores. Winters et al. (2016) found that topography as the main factor for certain cities’ low bike scores, while also noting that it does not create as big of a barrier for walking. Chen & Wang (2020) studied the green transportation modes, transit, and cycling, in Cincinnati and Fresno which are both very similar to size of Indianapolis. Their results showed that, for these cities, cycling provides a larger access than transit, which shows the need for improvement in the quality of transit. They also found that disadvantaged groups at the urban core do not suffer from low accessibility, depending on the trip duration threshold. Meng & Brown (2021) expanded upon the other green transportation modes such as BSS and e-scooter and evaluated their equity in terms of availability for 32 U.S. cities. Their results showed docked BSS are less equitable compared with the dockless services. They also highlighted that spatial distribution is the primary obstacle for accessing equitable services and the cities where policies aimed to ensure equity (Chicago) achieved lower Gini scores (higher means more inequitable in terms of formal equality-equal for all) compared with others in terms of spatial distribution.

Studies that use only mobility-based measures to assess equity are very limited compared with the studies that only focus on accessibility-based measures. This is likely due to the distinction between accessibility and mobility-based measures not being defined clearly. Generally these studies that focus only on mobility based measures analyze the overview of travel survey's such as NHTS trip characteristics (distance, mode share) regarding socio-economic characteristics and urban-rural differences as Pucher & Renne (2005) did. In their paper, they observed high car ownership and dependence independent of age, income, or race, once again showing the high automobile dependence in U.S. Their findings support the forced car ownership in U.S. that was discussed previously. Similarly, Shirmohammadli et al. (2016) looked at the number of trips by each trip mode regionally in Aachen Germany, and among 5 different groups. Their analysis based on formal equality showed that automobile is the most equitable trip mode on spatial level among all, meaning that there are not differences between regions in terms of car access. This is in line with the Pucher & Renne (2005) study which found similar results on national level in U.S.. Shirmohammadli et al. (2016) also reported that inner core achieved more mobility through bus and non-motorized modes compared with the outer core (urban core vs suburban). It is important to note that they concluded improving the access for public transit and non-motorized modes may lead to the increase in number of trips, but it would not shift the car trips to these modes because of no significant correlation between trip modes. Additionally, they found students use the public transit as their first choice of mode trip mode, which is likely because of the lower fare structures for them. Age was also found to be not a significant barrier to car mobility, meaning older people are still capable of driving (Shirmohammadli et al., 2016). Brown & Taylor (2018) looked at the ride sharing and ride hailing services using mobility-based measures such as the number of trips on spatial and sociodemographic level. They found that ride-hailing services can improve car access and do so equitability (no differences between different sociodemographic groups). Brown & Taylor (2018) also evaluated at the physical accessibility which differs from the accessibility that has been described so far. Physical accessibility mainly refers to the feasibility of a trip mode regarding physical conditions, such as disability or age, rather than having transport access in terms of proximity-based measures or cumulative opportunities. We consider this evaluation as part of the mobility-based measures rather than accessibility-based measures, since it is regarding the ease of access to ride hailing and ride sharing for certain populations rather than opportunities.

There are several studies that evaluated access in terms of feasibility of trip modes, which puts them into the last category of studies which evaluated equity in terms of accessibility and mobility-based measures. Welch & Mishra (2013) developed an integrated index of accessibility and mobility to create a quality index for assessing public transit equity. This quality index is based on connectivity, which includes frequency, speed, distance, capacity, required transfers, and activity density of land use served by a transit node. Their analysis showed differences between transit equity levels in Washington and Baltimore region. However, the methods used in this thesis are not asserting if the current system is fair or not, but providing a way for assessing equity before and after a transit project. This was mainly due to not having travel demand data incorporated in the model (Welch & Mishra, 2013). Pyrialakou et al. (2016) evaluated transportation equity regarding accessibility, mobility, and outcome-based measures for car, public transit, and non-motorized trip modes. Their study based on Indiana found that rural areas and small urban areas suffer from lack of opportunities and lack of transit supply. Mooney et al. (2019) looked at the availability of dockless bikes in different neighborhoods and found that dockless bikes are not distributed equitably on a spatial level. They also did not find any disparities in racial and ethical composition of neighborhoods with more or less access to dockless bikes. It is worth noting, their study evaluated the "realized access" rather than "potential access" so equity was evaluated in the context of travel demand. However, the travel demand was on aggregate level, which they showed on their paper as sensitive to results when different units of spatial aggregation were used (Mooney et al., 2019). Chen et al. (2019) evaluated the reachability for bike sharing stations using walking thresholds to incorporate willingness to walk for bike sharing trips for high resolution simulated trip origin and destinations. While their study incorporated necessary elements (such as exact origin and destination), they did not evaluate alternative travel modes or the infrastructure constraints. They found that distribution of bike sharing accessibility was unevenly distributed spatially and between sociodemographic groups. Their results showed downtown areas have higher accessibility with bike sharing, while they found white non-middle class (low and upper) males to have the most accessibility.

Although the existent research is extremely beneficial for identifying equity implications of the specific mode, they lack painting a complete picture. Using only a single mode of transportation in transportation equity analysis limits the comprehensiveness of the findings and makes them less applicable to policy discussions. In addition, it is agreed upon in the literature there is an unequal distribution of benefits of a transportation system in highly automobile dependent communities of U.S. (Jones & Lucas, 2012; Martens et al., 2012). The negative health and environmental impacts of this automobile dependency also impact people of color disproportionally (Reichmuth, 2019). Therefore, we need to focus on alternative mobility options and their feasibility to understand the forced automobile ownership and the negative impacts of this forced automobile ownership. Griffin & Sener (2016), Minocha et al. (2008), Welch & Mishra (2013) used public transit as the only mode of interest while Chen & Wang (2020), Golub & Martens (2014), Karner (2018) integrated bike, auto, and walking to public transit respectively in their analysis. While Chen & Wang (2020) and Golub & Martens (2014) compared different modes, bike-public transit & auto-public transit, specifically, Karner (2018) incorporated pedestrian network into transit trips thus indirectly analyzed if walking is equitable or not. While others have also used walking time as first-mile, last mile connection criteria for transit as well, only Karner (2018) explicitly pointed out the use of a pedestrian network for assessing walking time. This is important because using walking versus biking as first and last mile travel mode can drastically change the accessibility of public transit since bike can allow individuals to travel longer distances compared with walking. We assumed for all other papers used walking as the first & last mile mode if it wasn't explicitly stated. Meng & Brown (2021) looked at emerging mobility options (station-based BSS, dockless BSS, and e-scooters) together and were able to understand what type of systems are more equitable. Last, Pyrialakou et al. (2016) and Shirmohammadli et al. (2016) evaluated the transportation system equity as a whole by comparing auto, transit, and non-motorized modes. Their results showed automobile as the most equal (not equitable since they adopted formal equality as equity) mode compared with the others. Studies that evaluated multiple modes could identify equity problems more frequently than studies that used single mode could. This is because of the ability to compare travel modes with each other and understanding the forced use of certain trip modes. Evaluating multiple modes is also helpful in identifying what type of trip modes can be beneficial in improving certain regions' travel-demand-relevant access.

As highlighted above, many studies rely on aggregated data in evaluating equity. The spatial resolution lacking from aggregate data prevents evaluating feasibility of mobility options in the context of the travel demand. Without evaluating the quality of service and infrastructure constraints, travel-demand-relevant access can't be analyzed in individual level. The limitation with aggregated data is, they overlook detailed geolocation and the actual travel demand. For example, as pointed out earlier, simply living near a bus stop does not guarantee access unless a transit route can be taken to reach the exact destination. Additionally, having a travel demand that can be satisfied by walking based on the trip distance does not directly show access to the route characteristics might make that walking trip infeasible. From the reviewed literature, only Chen et al. (2019) used disaggregated data (from a simulation) while a majority of the literature that was reviewed used some sort of aggregated travel data. Some studies try to eliminate these problems by using travel survey data with higher spatial resolution (census block level) such as Pyrialakou et al. (2016) and Karner (2018). LEHD dataset both papers used only includes commuting trips, which is not ideal. Compared with the existent literature, knowing the exact location of the trip origin and destination allows for having individual level trip characteristics and evaluates everyone independently of others. Recently, with the emergence of tech-based mobility options created a unique opportunity for determining the disaggregated travel demand. Using the Global Positioning System (GPS) data, some studies could expand the level of detail with their analysis by eliminating the limitation associated with the aggregated travel survey data by using origin and destination pairs. The GPS studies can provide trip origin and destination information for all modes without the burden on participants for travel surveys and the need for specific data collection methods (Schuessler & Axhausen, 2009). Other advantages of using GPS data are related to the sample size for analysis and the continuity component of the data. Since data is collected in relative frequency, it is possible to generate trip chain information, which is not possible with any other methods on a large scale. As for our knowledge, there is no literature that generated origin and destination pairs from raw GPS data and used this information in the transportation equity field. A multi-modal (mixed modes and single modes are both included) study based on the real observed trip behavior derived from mobile phone sensors in U.S. can address the spatial resolution need however sociodemographic characteristics of these individuals can't be reported because of privacy concerns. Having the exact location of origin and destination can enable assessment of feasibility of each trip mode based on the ease of use and with respect to the infrastructure constraints. In addition, evaluating the feasibility of mobility options with mobility and accessibility-based measures in a multi-modal transportation system can aid more comprehensive understanding of forced use of certain trip modes and can help explain the specific limitations of access.

Table 2-1. Literature review for transportation equity (organized based on transport measures used, mobility, accessibility, or both)

Study	**Transport Measures Used** (Mobility, accessibility, etc.)	**Travel Modes Considered**	**Data Sources / Contents**
(Shirmohammadli et al., 2016)	Mobility—number of trips	Car–Driver Car–Passenger Bus Non-motorized (Walking and cycling)	Travel survey with trip mode and spatially aggregated travel demand (on regional level), along with socio-economic data to identify 5 groups (students, seniors, people without cars, unemployed, and employed people)
(Pucher & Renne, 2005)	Mobility—mode share, average trip distance	Car–Single/High occupancy Transit Non-motorized (walking and cycling) School bus	Travel survey (NHTS)—mode share and average trip distance regarding urban-rural spatial differences. Sociodemographic characteristics and mode share relationships are also used.
(Brown & Taylor, 2018)	Mobility—total trips, trips per capita (residents + job), ease of access for physically disadvantaged populations	Ride hailing Ride sharing	Lyft trip-level records: origins and destinations are aggregated on the census tract level. Trip start time, trip price, and trip distance are all aggregated in intervals
(Griffin & Sener, 2016)	Accessibility—comparison between transit access (being able to access work within 45-minute transit commute–walking, waiting, and transit time included) for low-wage workers and all workers	Public Transit (Bus and Rail for all cities besides Indianapolis where only bus was evaluated)	EPA’s Access to Jobs and Workers via Transit database (spatial aggregation on census block group level) This paper used% of low-wage workers with transit access and% of all workers with transit access
(Minocha et al., 2008)	Accessibility—infrastructure level transit quality (frequency, hours of service, service coverage of 0.25 mile for bus and 0.5 for rail)	Public Transit (Bus and Rail)	Individual bus route and rail line schedule charts were evaluated to create a transit availability index. Compared census tracts with high vehicle ownership and low vehicle ownership for socioeconomic characteristics.
(Karner, 2018)	Accessibility—stop (bus)-level gravity measures for low and high-wage jobs derived from travel times (only during morning peak) to service areas (0.25 mile for bus and 0.5 for rail)	Public Transit (Bus and Rail)	Travel survey (LEHD): Database with estimates for number of resident workers, jobs, and flows of resident workers to jobs in census block level. Data is on the aggregate level where origins are transit stops (with service area) and destinations are work locations.

Table 2.1 (continued): Literature review for transportation equity (organized based on transport measures used, mobility, accessibility, or both)

Study	**Transport Measures Used** (Mobility, accessibility, etc.)	**Travel Modes Considered**	**Data Sources / Contents**
(Smith et al., 2015)	Accessibility—bike station availability	Bike sharing (station based)	Bike station data for 42 bike sharing systems
(Braun et al., 2019)	Accessibility—bike lane access	Biking (infrastructure)	GIS shapefiles for unprotected, buffered, and protected bike lanes from local and regional data sources
(McNeil, 2011)	Accessibility—bikeability based on adjacent motor vehicle speeds, number of lanes, bike lane existence, quality of pavement and number of stops. Land use and land topography were also incorporated.	Biking (personal)	Infrastructure information for measuring bike quality index and travel surveys for determining trip characteristics such as trip distances that are feasible for biking
(Winters et al., 2016)	Accessibility—bike score which includes bike lane availability, hilliness, destinations & connectivity (opportunities), and bicycle mode share	Biking (personal)	Infrastructure and topographical values for bike score, travel surveys for mode share
(Meng & Brown, 2021)	Accessibility—mode availability	Station-based bike sharing Dockless bike sharing E-scooter	Aggregated General Bike Feed Specification (GBFS) data for parking space for dockless systems and station locations for docked systems
(Golub & Martens, 2014)	Accessibility—cumulative opportunity, destinations reachable in certain time limit (manufacturing and service jobs)	Automobile Public Transit	Generated travel demand
(Chen & Wang, 2020)	Accessibility—number of opportunities (dining, jobs, social activities, and schools) reachable in certain time thresholds	Public transit Cycling	Location of jobs, dining, physical activities, social activities, and schools. LEHD for aggregated block level trips for commuting travel demand

Table 2.1 (continued): Literature review for transportation equity (organized based on transport measures used, mobility, accessibility, or both)

Study	**Transport Measures Used** (Mobility, accessibility, etc.)	**Travel Modes Considered**	**Data Sources / Contents**
(Chen et al., 2019)	Accessibility—travel distance for walking and cycling (single trip and tour-based approach)	Bike sharing (station based)	Disaggregated activity (trip origin and destination) and sociodemographic data (simulated)
(Welch & Mishra, 2013)	Accessibility and mobility interestedly considered- quality index based on connectivity, which includes frequency, speed, distance, capacity, required transfers, and activity density of land use served by a transit node	Public Transit	Transit network data (stations, stops, lines, operational capacity etc.), and urban form characteristics
(Pyrialakou et al., 2016)	Accessibility—opportunities reachable in certain time thresholds Mobility—need based measures from travel behavior	Walking Automobile Public Transit	National Household Travel Survey and Bureau of Transportation Statistics average household person and vehicle trips/trip-miles
(Mooney et al., 2019)	Accessibility—(availability and idle times/rebalanced bikes)	Bike sharing (dockless bike)	Spatially aggregated bike availability data for neighborhoods

This thesis developed an integrated quality index of mobility and accessibility-based measures to eliminate problems arising from using them separately. This is because an integrated index can provide a more comprehensive perspective on evaluating the quality of each mobility option alternative in the context of travel demand and with respect to the infrastructure constraints. In Chapter 4.4, these methods with these integrated measures are explained in more detail. Naturally, this method will require a multi-modal approach rather than focusing on a single mode of transportation. Last, by using disaggregated mobile phone based sensory data, trip-chain information and detailed travel behavior can be extracted to identify the feasible mobility options available for a specific user. This individual information will be used to draw a holistic picture of the transportation equity. Since privacy concerns limit access to disaggregated sociodemographic data, this part will be aggregated. However, by using a large dataset, a bottom-up approach can be adapted, thus reduce the issues with using aggregated data.

3. EMPIRICAL SETTING

This study used Indiana as a case study based on data availability and local knowledge. The state of Indiana spans an area of 36,418 sq mi (94,321 km^2), has 6,785,528 residents, and is in the Midwest Region of the United States of America (Census, 2020a). The empirical setting that was used in this study is the greater Indianapolis Metropolitan Planning Area (MPA) boundary area. The Indianapolis Metropolitan Planning Organization (MPO) indicates:

> "The MPA is the urbanized area of Central Indiana (the areas that are already mostly developed, identified on the map by the Urbanized Area Boundary - UAB) plus the areas that are expected to urbanize over the next 20 years. The MPO guides the development of a multi-modal transportation system within the MPA. (MPO, 2012)"

This area spans outside of Indianapolis and includes the surrounding suburbs' block groups and its population according to the 2020 Census is estimated to be 1,769,964 (Census, 2020b). The MPA boundary used in this study was not drawn with respect to census block group boundaries, which is why there were some census block groups that were cut off by the boundary. To account for this, this study used census block groups that have their centroid inside the boundary (please see Appendix A.1 for details) for equity analysis in Chapter 4.5. Most of the area inside this boundary are well connected with interstates, state highways, and urban arterials. While in this area only 6.53% of both owner and renter occupied households do not own private vehicles, 34.53% of households that have at least one private vehicle and remaining 59% of households own multiple vehicles (ACS, 2018). This shows that there is a high level of car dependency, which could exacerbate the negative impacts on households without vehicles.

This study selected Indianapolis MPA as the empirical study area due to its high auto dependency and similarity with other mid-size cities around the world. Additionally, the empirical setting for this thesis, Indianapolis, had one of the greatest discrepancies in terms of transit access when Griffin & Sener (2016) looked at nine mid-size U.S. cities. Since the MPA area is based on the urbanized areas (or expected to be) rather than using urban/rural classification, urban core /suburban classification was adopted when discussing the results. This classification of urban-core/suburban is based on the classification scheme adopted by National Center for Health Statistics (NCHS) derived from Office of Management and Budget's (OMB) February 2013 delineation of metropolitan and micropolitan statistical areas (Ingram & Franco, 2012). Based on the method used by NCHS, this thesis adopted an urban core–suburban definition based on the classification of counties as large "central" metro (inner cities, urban core) and large "fringe" metro (suburban). Consequently, Marion County was denoted as large central metro, Boone, Hamilton, Hancock, Hendricks, Johnson, Morgan, and Shelby Counties were denoted as large fringe metro. For the rest of this thesis, any census block group inside the Marion County is denoted as urban core, while any other area in the empirical setting is denoted as suburban.

In August 2017, inside the MPA boundary, there were only three fixed route public transit systems (IndyGo, CIRTA, and Johnson County) and four appointment-based public transit systems (Boone, Hamilton, Hancock, and Hendricks County). The four appointment-based systems are all tailored for senior citizens and passenger must request these services. Indianapolis has a large-scale station-based bike sharing system (50 stations) while surrounding towns such as Carmel (six stations), Noblesville (two stations), Lawrance (two stations), Fishers (two stations) have small scale bike sharing systems. The two stations in Fishers only post the membership costs on the bike sharing app and, unfortunately, we could not find the membership costs for these two stations. This is why we removed those two stations from the study. In Figure 3-1 the fixed route public transit systems and the bike sharing station location inside the study area are shown. We assumed the two separate bikes sharing systems were not used during the same trip, thus making each trip mutually exclusive. There are also three e-scooter companies (Bird, Lime, Spin) that have operations inside the MPA boundary. In addition, there are several bike lanes and walking trails in Indianapolis that can be used for active travel modes. Lyft and Uber serve as ride hailing services inside the boundary. Bike sharing systems outside of Indianapolis did not exist in August 2017, but the assumption is that the same travel demand will hold true on the current day in 2022. For the Chapter 4.4, even though those stations might not exist in August 2017, the current infrastructure (in 2022) is considered as the infrastructure criteria.

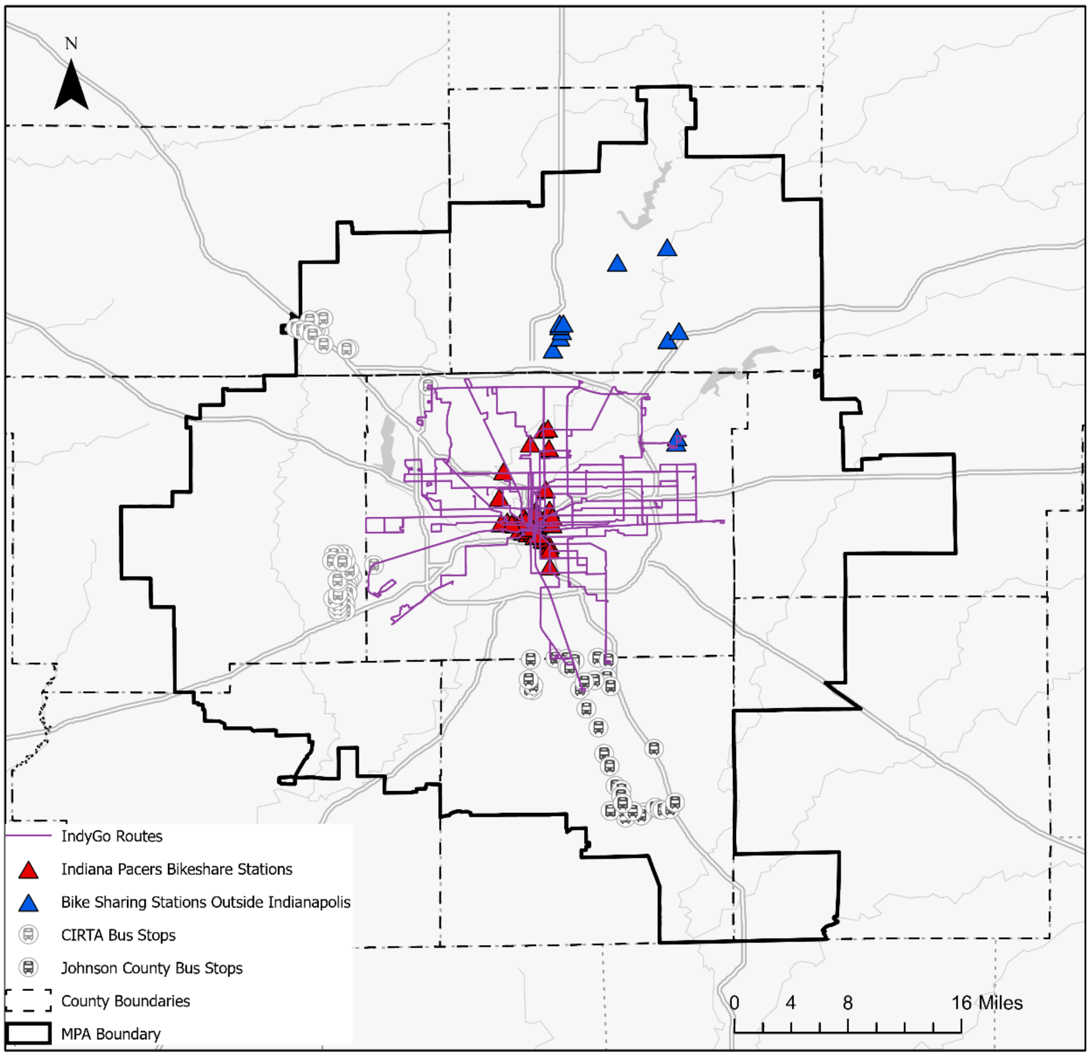

Figure 3-1. Public transportation networks and bike sharing stations in the study area.

4. METHODS

4.1 Overview

In this thesis, we developed a three-step method to evaluate travel-demand-relevant access. This three-part framework includes the GPS data processing, mobility option comparability and feasibility evaluation, and lastly the equity analysis in Indianapolis MPA. The overview of data and methods used for each part can be seen in Figure 4-1. Although the main output of GPS data processing step is to find the trip origin and destinations, this step also included any necessary pre-processing steps. These pre-processing steps are discussing in detail in Chapter 4.3.1. The origin and destination pairs that were identified were then evaluated for feasibility of different mobility options using two separate methods; cost based, and quality based. This step is the main contribution of this study, where we highlighted a framework for how to do travel-demand-relevant access analysis using multiple trip modes. For this step, we used publicly available Python and R packages and data to evaluate routing (trip planning) using different trip modes regarding infrastructure constraints. These infrastructure constraints were integrated based on several publicly available datasets, such as OpenStreetMap (OSM). These data sources are discussed in detail in Chapter 4.2. After identifying the feasibility of each mobility option, we compared spatial and sociodemographic characteristics at the census block level to assess spatial and sociodemographic equity inside MPA. GPS data processing step is discussed in more detail in Chapter 4.3, mobility option comparability and feasibility evaluation is discussed in Chapter 4.4 and lastly equity analysis is discussed in more detail in Chapter 4.5. In summary, this study's objective is to analyze spatial and sociodemographic equity based on travel-demand-relevant access using GPS data.

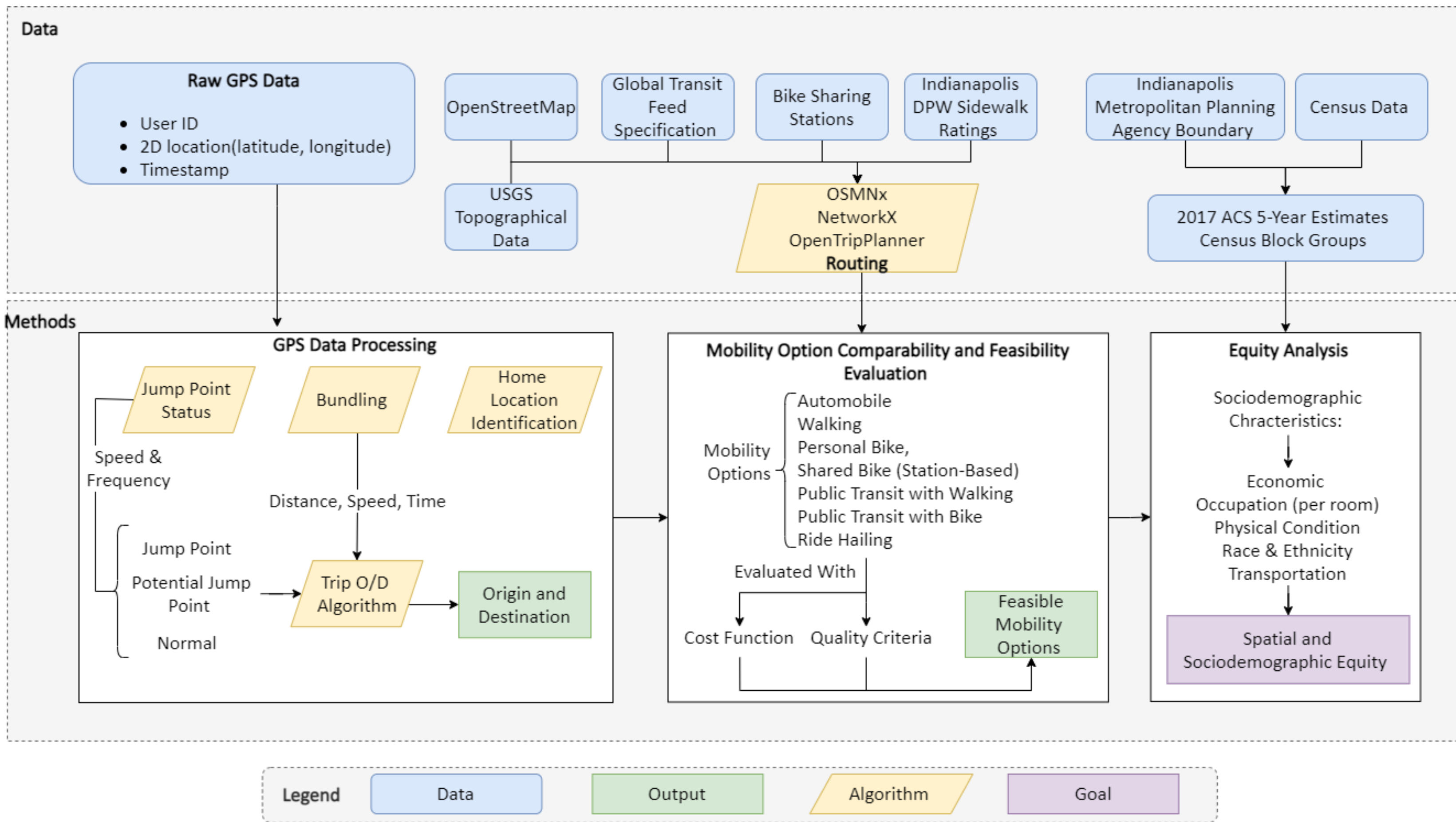

Figure 4-1. Overview of data and methods used in this study

4.2 Data

4.2.1 GPS Data

Modern smart phones are usually equipped with assisted GPS (A-GPS) to determine GPS location of the user. These systems prevent the warm-up period required for traditional GPS devices (Vallina-Rodriguez et al., 2013). Although this warm-up period is the main source of errors in traditional GPS devices, A-GPS systems are less accurate than traditional GPS receivers (Zandbergen & Barbeau, 2021). The accuracy of A-GPS sensors varies given the environment they are used in, the time of the day, and the model of the mobile phone. In general, accuracy can be placed between 8 and 13 meters when compared to the actual path for each point in a controlled experiment in an urban environment (Merryid & Bettinger, 2019). The GPS data used in this research is provided by SafeGraph. SafeGraph is a private company that collects GPS location information from several mobile applications with permission from the users. Data attributes used in this study are unique user ID, geographical latitude, and longitude coordinates in the WGS 84 reference system, UNIX timestamp in Universal Time Coordinated (UTC), and 9-character geohash code based on the public domain geocode system (Niemeyer, 2008), which separates the world into rectangular grids of 4.77 m x 4.77 m. By only using the first 7 characters of this 9-character geohash variable, a larger grid of (153 m x 153 m) was created using the same geocode system. The users analyzed in this study were selected based on having their estimated home locations within the borders of Indianapolis MPA. These home locations were determined by the most frequent 7-character geohash during nighttime (7 PM–7 AM) for each unique user. The assumption is that the GPS location where most data points were collected is either in the exact location of the house or very close. Please see Appendix A.2 for an example of this method. This filtering resulted in with a dataset which had a little over 60 million GPS points from 171,739 unique users. The homes that were inside the census block groups that were not included in the empirical setting (e.g., orange block group from Fig A.1 in Appendix A.1) were also removed from the final dataset used for equity analysis during the trip aggregation stage. This operation reduced the number of users to 153,391 and 3,807,003 GPS data points. The UNIX timestamp was converted into local time based on the geolocation with respect to the daylight-saving time changes. The 10-day time-period used in this study was between “2017-08-21 00:00:00” and “2017-08-31 23:59:59” at local time. The time interval between points is irregular and varies substantially by the user. Aside from these parameters, the haversine distance, time interval, and average speed between two consecutive points were calculated to create the final input matrix of $\boldsymbol{trip\ matrix}$**.** It is also worth noting that we assumed that even though the time difference between two consecutive GPS points might be longer than a day, there are no missing trips in between those two points. This assumption is based on the data characteristics and might not hold true if app stopped collecting location information from users while they were still traveling.

4.2.2 Transportation System and Geographical Data

OpenStreetMap (OSM)

OpenStreetMap (OSM) data is used to provide the transportation network information such as road types, posted speed limits, road segment lengths (OpenStreetMap, 2022). It is possible to create the networks from OSM for a given geographic area using a very useful Python package called “OSMNx”. OSMNx was developed by Geoff Boeing to analyze street networks by utilizing publicly available OSM data (Boeing, 2017). Besides the OSMNx, the network analysis package “NetworkX”“ (Hagberg et al., 2008) and its tools were heavily used with routing algorithms. These trip planning (routing) algorithms were used to find if optimal conditions are assumed how many options would be feasible for user for a specific O/D-pair. Both packages were used without modifications for shortest path routing with driving, walking, biking (personal and bike sharing), and ride hailing. Specific details and tags used for each travel mode are discussed in their respective chapter below. In addition, the shortest path algorithm from NetworkX and OSMNx both use the Dijkstra method for finding the shortest network path given the edge weights (Dijkstra, 1959).

Topographical Data

Topographical data is downloaded as 1/3 arc-second (10 meter) Lidar data from U.S. Geological Survey for Indiana was used (USGS, 2021). This data was used to build a bike network with elevation so that gradient (slope) of each segment can be calculated and used in routing decision. Please see Chapter 4.4 for the details.

Sidewalk Rating

Not all streets have sidewalks to support safe walking. To assess the feasibility and quality of a walking trip, this theis used the 2018 sidewalk rating data compiled by the Indianapolis Department of Public Works (DPW, 2018). This dataset rates sidewalks with scores ranging from zero to three, with zero representing no concrete sidewalk or curb (or in very bad ondition) on that section. Sidewalks with good rating (1,2,3) showed that some sort of sidewalk infrastructure is present, but there might be some vegetation growth for sidewalks with rating one or two. The sections with sidewalk rating of zero were eliminated from walk network because of this reason. Interestingly, out of 69,767 sidewalks listed in the dataset, 39,742 (56.96%) were rated 0; 167 (0.24%) were rated as 1; 3,438 (4.93%) were rated 2; and 26,420 (37.87%) were rated 3. There has been 20 years of sidewalk moratorium in Indianapolis (K. Dwyer & J. Ryan, 2021), which is likely the reason the majority of sidewalks are non-existent or in a bad condition. One limitation of this dataset is that it only has the sidewalks inside the Marion County, which meant that the remaining area outside of MPA boundaries were assumed to have a proper sidewalk for all network edges. Figure 4-2 shows the sidewalks with a bad rating (0) and a good rating (1,2,3) in Marion County.

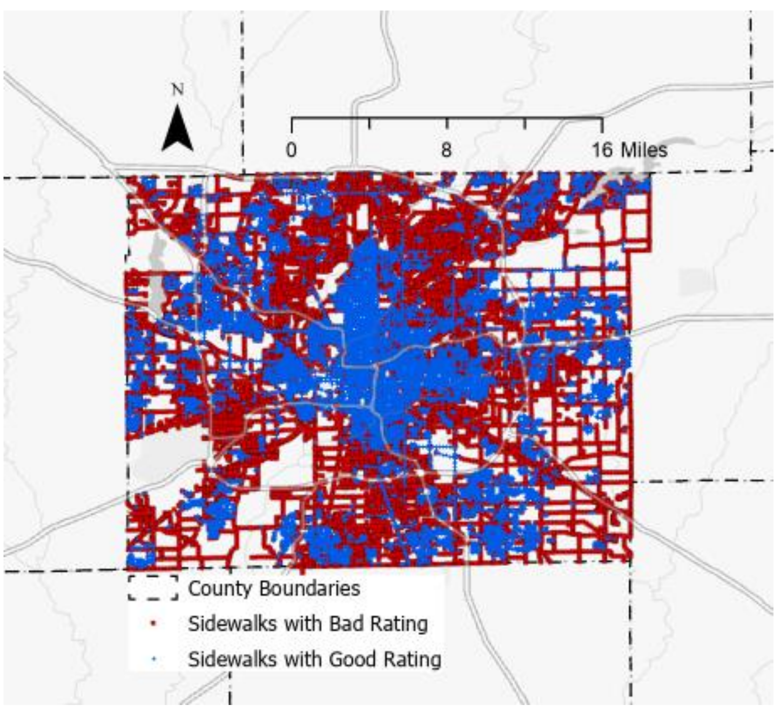

Figure 4-2. Sidewalk quality in Marion County (bad rating: no sidewalk or curb or sidewalk with a terrible condition, good rating: sidewalk with at most some vegetation growth)

Bike Sharing Stations

As described in the empirical setting, there are 61 bike sharing stations inside the MPA boundary. The geographic location of these stations is publicly available. Indiana Pacers Bikeshare station data was pulled from the General Bikeshare Feed Specification (GBFS) data, while Carmel, Noblesville, Fishers, and Lawrance bike share stations were manually listed based on the pin locations on the official app for all systems “Movatic”. Indiana Pacers Bikesharing, which is the primary bike sharing system (BSS) in Indianapolis, indicates on their website that they have 50 stations while the GBFS data has only 49 stations. Similarly, the Carmel Bike Share shows that they have 7 stations while the official app “Movatic” only shows six stations. In total, we consider 59 total stations that location information was found. All stations that had longitude and latitude data that were used for assessing bike sharing as a mode option.

Public Transit Data

For public transit information, Global Transit Feed Specification (GTFS) data for Indianapolis in June 2017 was used. This data was the closest schedule that we could find to the GPS data (August 2017). As described in Chapter 3, some other public transit systems exist inside the MPA boundary. Out of seven systems, only three have fixed route and schedule while four others are appointment based public transit systems tailored for senior citizens. Unfortunately, out of three fixed route systems, only IndyGo GTFS data was publicly available, thus neither Johnson County nor CIRTA systems were considered in this analysis. The GTFS data for IndyGo had 32 routes with 3641 stops in total and included the schedules, timetables for each route.

4.2.3 Sociodemographic Data

For this study, census block groups of Indiana determined by the 2010 Census were used as the spatial areas of interest. Census block groups are the second smallest geographic area after census block, and they are the smallest area for which the decennial census tabulates and publishes sample data (Census, 2021). Because of the trip observation data that was used in this study is from August 2017, all sociodemographic data has been collected from 2017 ACS 5-year estimates to ensure consistency in our analysis. As described in the empirical setting (Chapter 3), only the block groups that were identified to be within the MPA boundary were considered in this study.

This theiss identified race, vehicle ownership, educational attainment, age, economic characteristics (median income, poverty, median property value), and physical disability as the most important sociodemographic variables. Please see Appendix B for which variable is used in which studies in existing literature. The descriptive statistics of all sociodemographic variables used in this study are represented in Table 4-1 below.

Table 4-1. Descriptive statistics of sociodemographic data used in this study regarding the study area (MPA boundary)

Variable	Description	Mean	1ST Quartile	Median	3rd Quartile
Economic					
Income	Median household income	$57,461	$34,398	$50,257	$71,956
Property Value	Median property value	$146,892	$86,175	$129,600	$175,800
Poverty	% Of population 20 to 64 years under poverty line	16.04%	4.61%	11.51%	24.3%
Education					
Degree	% Of population over 25 with no schooling	1.15%	0	0	1.6%
	% Of people over 25 with highest level of education high school and below	36.98%	21.65%	37.12%	51%
Social Conditions					
Age	% Of under 20 and over 65 population	39.54%	34.54%	40.21%	44.92%
Ethnicity	% Of Hispanic or Latino population	7.85%	0.91%	3.85%	10.58%
Race	% Of non-white population	28.87%	6.11%	18.20%	45.92%
Occupancy					
Room Occupancy	% Of households with 1+ average room occupancy	1.73%	0	0	2.57%
Transportation					
Vehicle Ownership	% Of households with zero vehicle	8.48%	1%	4.88%	12.83%
Physical Condition					
Disability	% Of population with disability under poverty line	3.9%	0	1.66%	5.55%
	% Of population with disability over poverty line	8.78%	4.55%	8.11%	11.8%
Number of observations (census block groups)			912		

4.3 GPS Trajectory Data Processing

There are several steps required to generate trip origin and destination pairs (O/D pairs) from raw GPS data. This is because GPS data has likely to have errors / jump points that need to be filtered out before any operation for identifying trip origins and destinations. Below, we explained the methods used for processes used for data pre-processing (Chapter 4.3.1) and illustrated the rule-based methods used for extracting O/D pairs (Chapter 4.3.2).

The variable notations used in trip origin and destination algorithm is shown below and each variable will be defined as: v_{ij} where i refers to user and j refers to the index value of the point j where calculated values (distance, time difference, and average speed) are from the transition from point $j-1\ to\ j$. Orange text in algorithms throughout the text shows comments.

Input Data:

// Point j could refer to either an individual point or cluster of points (bundle) for below variables

$User\ i = \{1, \dots, I\}, point\ j = \{1, \dots., J\}, trip\ k = \{1, \dots, K\} for\ each\ user$

$\boldsymbol{t_{ij}} = unix\ time\ in\ second\ of\ point\ j\ of\ User\ i$

$\boldsymbol{lon_{ij}} = longitude\ of\ the\ point\ j\ of\ User\ i$ //updated as centroid longitude after bundling

$\boldsymbol{lat_{ij}} = latitude\ of\ the\ point\ j\ of\ User\ i$ //updated as centroid latitude after bundling

$\boldsymbol{sg_{ij}} = small\ grid\ (4.77m \times 4.77m)\ geographic\ location\ the\ world\ of\ point\ j\ of\ User\ i, geohash\ string$

$\boldsymbol{lg_{ij}} = large\ grid\ (153m \times 153m)\ geographic\ location\ on\ the\ world\ of\ point\ j\ of\ User\ i, geohash\ string$

$\boldsymbol{\Delta d_{ij}} = haversine\ distance\ in$ **miles** $for\ point\ j\ of\ User\ i\ (between\ j\ and\ j-1)$

$\boldsymbol{\Delta t_{ij}} = elasped\ time\ in$ **seconds** $for\ point\ j\ of\ User\ i (between\ j\ and\ j-1)$

$\boldsymbol{v_{ij}} = average\ speed\ in$ **mph** $for\ point\ j\ of\ User\ i\ (between\ j\ and\ j-1)\ (\boldsymbol{\Delta d_{ij}} / (\boldsymbol{\Delta t_{ij}/3600}))$

Output Data from pre-processing steps and trip O/D algorithm:

$\boldsymbol{\beta_{ij}} = gps\ jump\ point\ indicator, faulty\ GPS\ observation, \{normal, potential\ \ jump\ point, jump\ point\}$

$\boldsymbol{n_{ij}} = number\ of\ points\ in\ the\ bundle\ j\ for\ User\ i$

$\boldsymbol{\varphi_{ij}} = arrival\ time\ of\ bundle\ j\ for\ User\ i$

$\boldsymbol{\phi_{ij}} = departure\ time\ of\ bundle\ j\ for\ User\ i$

$\boldsymbol{s_{ij}} = status\ of\ oint\ j\ of\ User\ i, \{end\ point, not\ moving, moving\}$

$\boldsymbol{o_{ij}} = origin\ status\ of\ point\ j\ of\ User\ i\ \{k\ or\ 0)$

$\boldsymbol{d_{ij}} = destination\ status\ of\ point\ j\ of\ User\ i\ \{k\ or\ 0)$

$\mathbf{O}_{\boldsymbol{ik}} = origin\ for\ trip\ k\ for\ User\ i\ (lat, lon)$

$\mathbf{E}_{\boldsymbol{ik}} = destination\ for\ trip\ k\ for\ User\ i\ (lat, lon)$

$\mathbf{T}_{\boldsymbol{ik}} = start\ time\ of\ the\ trip\ k\ for\ User\ i$

$\boldsymbol{\Delta D_{ik}} = haversine\ distance\ between\ \boldsymbol{O_{ik}} and\ \boldsymbol{D_{ik}}$

$\boldsymbol{\Delta T_{ik}} = time\ between\ between\ \boldsymbol{O_{ik}} and\ \boldsymbol{D_{ik}}$

$\boldsymbol{\omega_{ik}} = speed\ (v_{ij})\ of\ the\ origin\ for\ trip\ k\ for\ User\ i$

$\mathbf{Z}_{\boldsymbol{ik}} = jump\ point\ status\ of\ the\ origin\ for\ trip\ k\ for\ User\ i$

$\boldsymbol{\zeta_y} = dummy\ variable\ for\ distance$

$\boldsymbol{\eta_y} = dummy\ variable\ for\ time$

$\boldsymbol{\alpha_{ki}} = index\ of\ the\ origin\ of\ trip\ k\ for\ User\ i\ (when\ all\ data\ points\ are\ sorted\ based\ on\ time)$

$\boldsymbol{\sigma_{ki}} = index\ of\ the\ destination\ of\ trip\ k\ for\ User\ i\ (when\ all\ data\ points\ are\ sorted\ based\ on\ time)$

4.3.1 Data Cleaning and Pre-processing

GPS Jump Points

General method for identifying the jump points is to use a maximum speed threshold value. There is one limitation of this method where the high speeds are caused by the previous point being a jump point, so points that are part of the trip trajectory are excluded if speed is used independently. Schuessler & Axhausen (2009) used 50 m/s (112 mph) as the maximum speed threshold and in this thesis speed limit was rounded up to the nearest whole number of 120 mph to have a more relaxed speed threshold. To ensure real trip trajectory points were not excluded, a method was identified based on the characteristics of our GPS dataset. When looking at the different users of the dataset, the GPS jump points were found to be clustered around the same geolocation. (Please see Appendix A.3 for an example user and how jump points look like on a map.) At most, they were concentrated into one large cluster or two large clusters for each user. This is likely because of having multiple accounts using different electronic devices at the same time. For example, as a user is moving along with a trip trajectory, their GPS data point suddenly jumps back to their home location and then jumps back to the location they were supposed to be. Since these clusters were matching with the frequently traveled locations, a process was developed to identify jump points based on the frequency and speed. By using the existent grid variable geohash (9-character), we extracted the two most frequent small grids for each user. These points are later described as potential jump points or jump points, the distinction being made if the maximum speed threshold of 120 mph is exceeded. If a point was neither a jump point nor a potential jump point, it was denoted as normal. Potential jump points were further examined when later the trip origin and destinations were identified. If an origin or destination is found to be a potential jump point and this trip created an impossible trip chain; these trips were trusted less than another trip that might not have an origin or destination denoted as potential jump point. With that, the jump point identification process can be listed as a three-step process. In this process, the model used both small and large grids. Small grids $(4.77m \times 4.77m)$ almost exactly match with standard home sizes when considered with their neighboring grids, which is why we choose this to capture all GPS points that might spread around the house. Large grids $(153m \times 153m)$ allow for clustering all the points around the areas where jump points most frequently happen. The jump point identification can be described as three step process as it is explained below.

1. Identify the two most frequent small grids and the neighboring grids. All these points that are inside these $14.31\ m \times 14.31\ m$ areas will be labeled as “potential jump point”. All others are labeled as “normal”.
2. Identify the two most frequent large grids and for the points inside this grid
    a. If a point has high speed, change the status as “jump point”
    b. If it doesn’t satisfy speed criteria, do not change the jump point status
3. For points that did not move, copy jump point status

The method we identified as fitting to our dataset most likely will not work any other data source that might have different attributes so when replicating this research better methodologies for filtering out the jump points such as the Kalman filter and Gauss kernel smoothing approach should be used. These methods were not used because the end goal of this project focused on trip origin and destination rather than on the trip trajectory. For further reference on these, please refer to Zheng & Zhou (2011). With this jump point identification process, a new variable represents the jump point status of each point as one of the following: “normal”, “potential jump point”, and “jump point”. The points with jump point status were eliminated from the GPS data and distance, time, and speed variables were recalculated to eliminate the effects of jump points (such as having high speed because of the jump points).

Clustering / Bundling

After eliminating jump points from the dataset, a clustering / bundling operation was performed to streamline finding the trip origin and destination. By creating these bundles, we aimed to understand how long the user stayed at one location for over one consecutive GPS point collections. The assumption for bundling these points together is that if a user stayed at a location for an extended amount of time without changing locations (less than 0.25 miles location change), they are not likely to be in the middle of a trip. For clustering, the existent geohash variable was once again used. The python library ‘geolib’ was used to extract the neighborhood for each grid (4.77 m × 4.77m), surrounding eight grids and the grid itself (Veness, 2014). If a point is located inside this 14.31m × 14.31m square of the previous point, it was placed in the same bundle as the previous point. The following point is then checked with this point and placed in the same bundle if it is located inside this point’s neighborhood square. With every additional point this method add, the centroid of all points in a bundle was calculated and the next point was evaluated based on its relationship to this center. The arrival time (φ_{ik}) is the timestamp of this first point while the departure time (ϕ_{ik}) is assigned as the timestamp of the last point of the bundle. The number of points in each bundle is also recorded as: n_{ik}. After this operation, all closely located points in a likely stay are combined into one "point" to be used in the Trip O/D extraction algorithm. From this point forward, these bundles were referred to as a point for the rest of the Chapter 4.3. This means that a point could comprise multiple GPS data points or a single point.

4.3.2 Trip Origin / Destination (O/D) Extraction

After removing jump points and creating bundles, status of each point can be identified. This status is different from the jump point status identified in Chapter 4.3.1. Since the end goal is to identify the trip O/Ds, the status can be one of the "end point", "not moving", and "moving". "End point" will indicate the point of interest is the destination of a trip started at an earlier point. "Not moving" categorization is for the points that are closer to the previous point than the minimum distance threshold. The minimum distance threshold of 0.25 miles which was derived from the 2017 NHTS average person trip length data for under 20 minutes trips (Federal Highway Administration, 2017). "Moving" on the hand other indicates all other points that are part of an individual trip trajectory. The general idea with the Trip O/D algorithm is to use rule-based algorithms to detect these statuses. This is consistent with the existing literature where the dwell time threshold between two consecutive points is evaluated for identifying trip O/D's (Shen & Stopher, 2014; Wang et al., 2018). This study adapts a variation of the existing methodologies by checking both transitions, arrival to point and departure from it. This means that if the point of interest is defined as $point\ j$, the transition from $j-1\ to\ j$ and the transition from $j\ to\ j+1$ will be used to decide the status of $point\ j$. In some cases, transition from $j-2\ to\ j-1$ was also used.

To identify all trips precisely, three-time thresholds were used for these transitions' dwell times. The most used time threshold in literature for dwell time is 120 seconds (Feng & Timmermans, 2014; Schuessler & Axhausen, 2009; Wang et al., 2018) while some studies used a higher time 900-second threshold (Schuessler & Axhausen, 2009). Values between 900 seconds and 1200 (20 minutes) seconds with 60 seconds increments was tried on previously mentioned 8 manually created user trip trajectory files and 1200 seconds was the best fitting for these users in identifying the correct trip origin and destinations. The max commuting threshold is determined as one hour, since this was the approximate maximum time that would be required to travel from one end of the MPA to another end. Higher time threshold value was needed because of irregular time intervals between pairs of bundled points and among users. By keeping this threshold large, it can also be ensured that fewer trips are likely to end abruptly. The three dwell time thresholds are identified as 120, 1200 and 3600 seconds accordingly. These three-time thresholds allowed identification of trips with different purposes by differentiating the stay time at a certain location. Shorter stays at a location (2 minutes threshold) identified trips with very low time stayed at the destination (e.g., drive through) while slightly longer time threshold (20 minutes) helped to identify most of other trip destinations (e.g., grocery shopping). Last, the longest time threshold (1 hour) helped with identifying trip destinations with longer activities (e.g., work). To validate the results from the trip O/D algorithm, this study compared the trip origin and destination locations identified by the algorithm with the manually evaluated 8 user files. These user's trips were evaluated by projecting their GPS point onto the map and identifying the origin and destination by analyzing these points and the distance, time, and speed values of each point. Aside from time-based classification, speed was a secondary criterion used to identify the status of a point. The minimum speed threshold which represents the average walking speed was calculated using the average gait speed of people between ages 20-79 for all genders (3 mph) from the meta-analysis of walking speeds (Bohannon & Andrews, 2011). Based on the speed of the point $j-1, j, and\ j+1$ it is possible to understand if the point j was part of an individual trip trajectory (moving) or an endpoint by evaluating trip continuation. Meaning the speed of these three points is like each other and above the speed threshold, it is more likely they will be part of the same trip trajectory. All thresholds used in this chapter are represented in the Table 4.2 below.

Table 4-2. Threshold values used in the Trip OD algorithm.

Threshold Value Name	**Represented by**	**Value**
Short Time Threshold	$time_{short}$	120 seconds (2 minutes)
Medium Time Threshold	$time_{medium}$	1200 seconds (20 minutes)
Long Time Threshold	$time_{long}$	3600 seconds (1 hour)
Distance Threshold	$distance_{threshold}$	0.25 miles
Speed Threshold	$speed_{threshold}$	3 mph

In Algorithm 1, we describe the rule-based algorithm used to identify the status of a point. First, bundle information was evaluated to check if the user stayed at the same location for over 120 seconds and while multiple GPS points were recorded. If the distance between this bundle and the previous one was larger than the threshold, we assigned this point as “end point” since the user stayed at this location for more 2 minutes. The status will be set to not moving if they didn't move more than the threshold. Check branch 1 on Algorithm 1 for this group. If either of the transition times is longer than an hour, these could only indicate end point or not moving status because the assumption is that points that have longer than one hour time gap between them will not be part of the same trip trajectory (trips where other points are part of the same trip). Branches 2 and 6 show the other criteria used the determine these points’ status. However, when the time between point $j-1$ and j is between 20 minutes and one hour, this assumption regarding points being in same trip trajectory is relaxed, because when evaluating manually created user files, it was noted that some points in an individual trip trajectory can be apart from each other for more than 20 minutes. Additionally, it is also more plausible for points to be part of the same trip trajectory if the time between them are less than an hour because of the size of the study area. Branch 3, 4, and 5 in Algorithm 1 depicts the points that fall into this category. The difference between these branches is the different values for Δt_{ij+1}(the time gap between point j and $j+1$). We checked the speed of point $j-1, j, and\ j+1$ to distinguish moving points from end point or not moving statuses. The assumption with these branches is that if the speed is consistently larger than the threshold, it would indicate continuation rather than an ending. For branches 7, 8, and 9, since the time is less than 20 minutes, all transitions need to be carefully examined. Since some points that does not satisfy the distance threshold criteria were also included as end point to not miss any trips.

Algorithm 1: Initial Trip O/D Algorithm

Input: $\boldsymbol{Trip\ matrix}$: $\{v_{ij}, \Delta d_{ij}, \Delta t_{ij}, \beta_{ij}, \phi_{ij}, \varphi_{ij}, n_{ij}\}$
Output: $\boldsymbol{trajectory}$: $\{v_{ij}, \beta_{ij}, \phi_{ij}, s_{ij}\}$

For each user, $j = 1$ **to** N **do**

1) If $n_{ij} \geq 2\ and\ \phi_{ij} - \varphi_{ij} \geq time_{short}$ **do**
  If $\Delta d_{ij} \geq distance_{threshold}$ **do**
    Assign: s_{ij} = end point //the user moved more than distance threshold to stay in a location for more than 120 seconds with 2+ recordings
  Else
    Assign: s_{ij} = not moving //the user stayed in a location for more than 120 seconds but this location is very close to the previous one
  End

2) Else if $\Delta t_{ij} \geq time_{long}$ **or (** $\Delta t_{ij} \in [time_{medium}, time_{long})$ **and** $\Delta t_{ij+1} \geq time_{long}$) **do**
  If $\Delta d_{ij} \geq distance_{threshold}$ **do**
    Assign: s_{ij} = end point //if either one of the times is more than an hour, we assume that they can't be part of same trip trajectory since the point j is farther than $j-1$, it is assigned end point status
  Else
    Assign: s_{ij} = not moving
  End

3) Else if $\Delta t_{ij} \in [time_{medium}, time_{long})$ **and** $\Delta t_{ij+1} \in [time_{medium}, time_{long})$ **do**
  If $\Delta d_{ij} \geq distance_{threshold}$ **do**
    If $v_{ij-1} \geq speed_{threshold}$ **and** $v_{ij} \geq speed_{threshold}$ **and** $v_{ij+1} \geq speed_{threshold}$ **do**
      Assign: s_{ij} = moving //consecutive points are more than 20 minutes but less than one hour away from each other can be part of same trajectory if the speed is consistently above the speed threshold
    Else
      Assign: s_{ij} = end point //due to Δt_{ij} being longer than 20 minutes if speed was not consistent this will indicate trip between j-1&j
    End
  Else Assign: s_{ij} = not moving
  End

4) Else if $\Delta t_{ij} \in [time_{medium}, time_{long})$ **and** $\Delta t_{ij+1} \in [time_{short}, time_{medium})$ **do**
  If $v_{ij} \geq speed_{threshold}$ **and** $v_{ij+1} \geq speed_{threshold}$ **do**
    Assign: s_{ij} = moving //only these two speeds needs to be checked because $\Delta t_{ij+1} < 20$ minutes away, likely to be part of the same trip
  Else if $\Delta d_{ij} \geq distance_{threshold}$ **do**
    Assign: s_{ij} = end point //if speed condition not met it is likely because of the v_{ij+1} thus only check distance now
  Else
    Assign: s_{ij} = not moving
  End

5) Else if $\Delta t_{ij} \in [time_{medium}, time_{long})$ **and** $\Delta t_{ij+1} < time_{short}$ **do**

If $\Delta d_{ij} \geq distance_{threshold}$ **do**

If $v_{ij+1} \geq speed_{threshold}$ **do**

Assign: s_{ij} = moving //since the next point is less than 2 minutes away if speed criteria is satisfied it means that trip is continuing

Else

Assign: s_{ij} = end point //if speed criteria is not satisfied the point j is an end point

End

Else

Assign: s_{ij} = not moving

End

6) Else if $\Delta t_{ij} < time_{medium}$ **and** $\Delta t_{ij+1} \geq time_{long}$ **do**

If $\Delta d_{ij} < distance_{threshold}$ **do** //since the next point is more than one hour away this point can only be end point or not moving0

If $v_{ij} \geq speed_{threshold}$ **do**

Assign: s_{ij} = end point //the trip ended at point j and a there is a trip between j and j+1

Else

Assign: s_{ij} = not moving //if the speed criteria not satisfied this means that distance is very low. Even though this point is identified as not moving the destination for the previous trip is in very low distance.

End

Else

Assign: s_{ij} = end point

End

7) Else if $\Delta t_{ij} < time_{medium}$ **and** $\Delta t_{ij+1} \in [time_{medium}, time_{long})$ **do**

If $\Delta d_{ij} \geq distance_{threshold}$ **do**

If $v_{ij} \geq speed_{threshold}$ **and** $v_{ij+1} \geq speed_{threshold}$ **do**

Assign: s_{ij} = moving //in this case consecutive speed threshold is necessary because time threshold for point j is less than 20 minutes

Else

Assign: s_{ij} = end point

End

Else

If $v_{ij} \geq speed_{threshold}$ **and** $v_{ij+1} \geq speed_{threshold}$ **do**

Assign: s_{ij} = moving //in this case consecutive speed threshold is necessary because time threshold for point j is less than 20 minutes

Else if $v_{ij-1} \geq speed_{threshold}$ **and** $v_{ij} \geq speed_{threshold}$ **and** $v_{ij+1} < speed_{threshold}$ **do**

Assign: s_{ij} = end point //this will indicate a sudden change in the speed which assumed to

Else

Assign: s_{ij} = not moving

End

End

8) Else if $\Delta t_{ij} < time_{medium}$ **and** Δt_{ij+1} in $[time_{short}, time_{medium})$ **do**

If $\Delta d_{ij} \geq distance_{threshold}$ **do**

If $v_{ij} \geq speed_{threshold}$ **and** $v_{ij+1} \geq speed_{threshold}$ **do**

Assign: s_{ij} = moving //in this case consecutive speed threshold is necessary because time threshold for point j is less than 20 minutes

Else if $v_{ij} < speed_{threshold}$ **and** $v_{ij+1} \geq speed_{threshold}$ **do**

Assign: s_{ij} = end point //since the distance criteria is already satisfied

Else

Assign: s_{ij} = end point //since the distance criteria is already satisfied

End

Else

If $v_{ij} \geq speed_{threshold}$ **and** $v_{ij+1} \geq speed_{threshold}$ **do**

Assign: s_{ij} = moving //in this case consecutive speed threshold is necessary because time threshold for point j is less than 20 minutes

Else if $v_{ij} \geq speed_{threshold}$ **do**

Assign: s_{ij} = end point //this condition might create extra trips but without trips will be missed,

Else

Assign: s_{ij} = not moving

End

End

9) Else if $\Delta t_{ij} < time_{medium}$ **and** $\Delta t_{ij+1} < time_{short}$ **do**

If $\Delta d_{ij} \geq distance_{threshold}$ **do**

If $v_{ij+1} \geq speed_{threshold}$ **do**

Assign: s_{ij} = moving //next point is less than 2 minutes away so if the speed criteria is satisfied it will be part of the trip

Else

Assign: s_{ij} = end point //distance criteria satisfies so it can't be not moving point

End

Else

If $v_{ij} \geq speed_{threshold}$ **and** $v_{ij+1} \geq speed_{threshold}$ **do**

Assign: s_{ij} = moving

Else if $v_{ij} \geq speed_{threshold}$ **and** $v_{ij+1} < speed_{threshold}$ **do**

Assign: s_{ij} = end point //this condition might create extra trips but without trips will be missed, but short trips are deleted later

Else

Assign: s_{ij} = not moving

End

End

End

End

Since "end point" status represents the destination points, only the points where the trip originated from needed to be found. To find origins, statues that were determined from Algorithm 1 were used. If a point had status of end point, the next trip was assumed to start from the same point. For example, if the point j was assigned to the trip k as the destination, it will also be assigned as the origin for trip $k + 1$. This is because of the trip chain assumption with GPS data used in this study. If a bundle was found to be destination or origin, we used respective arrival and departure times to find trip characteristics. The speed and the jump point status of origin and destinations were also recorded to be used in trip O/D filtering algorithm. Aside from "end point" status, "not moving" status was used to indicate the next point as the start of the next trip. "Moving" status was only used to differentiate between the trip trajectory points and the points that do not belong to any trips. The details of this algorithm can be seen below in Algorithm 2.

Algorithm 2: Trip O/D Assignment Algorithm

Input: $\boldsymbol{trajectory}$**:**$\{v_{ij}, \beta_{ij}, \varphi_{ij}, s_{ij}\}$
Output: $\boldsymbol{tripOD}$: $\{\mathbf{O}_{ik}, \mathbf{E}_{ik}, \mathbf{T}_{ik}, \boldsymbol{\Delta D}_{ik}, \boldsymbol{\omega}_{ik}, \mathbf{Z}_{ik}\}$

Initialize $k = 1$ for each user
Initialize $\mathrm{O}_{ik} = [lat_{i1}, long_{i1}]$ and $\mathrm{T}_{ik} = \varphi_{i1}$ and $\omega_{ik} = v_{i1}$
For all u_i, $j = 1$ **to** N **do**
    If s_{ij} = "end_point" **do**
        Assign $d_{ij} = k$, $o_{ij} = k + 1$
        Assign $\mathrm{E}_{ik} = [lat_{ij}, long_{ij}]$ $\mathrm{O}_{ik+1} = [lat_{ij}, long_{ij}]$, $\mathrm{T}_{ik+1} = \varphi_{ij}$, $\omega_{ik+1} = v_{ij}$, and $\mathrm{Z}_{ik+1} = s_{ij}$
        Calculate ΔD_{ik} = haversine distance between the O_{ik} & E_{ik}
        Calculate $\Delta T_{ik} = time\ between\ between\ \boldsymbol{O_{ik}} and\ \boldsymbol{D_{ik}}$
        Increment k
    Elif s_{ij} = "not_moving" **do**
        Assign $o_{ij-1} = 0$, $o_{ij} = k$ //to ensure if status is not moving there is no trip between $j\ and\ j - 1$
        Assign $\mathrm{O}_{ik} = [lat_{ij}, long_{ij}]$ and $\mathrm{T}_{ik} = \varphi_{ij}$ and $\omega_{ik} = v_{ij}$
    End
End

After identifying the trip origin and destination and necessary characteristics (total trip distance, total trip duration, jump point status of the origin and destination, arrival time to bundle, departure time from bundle, and the speed of the origin and destination) were recorded. Since it is impossible to account for all combination of distance, time interval, and speed there were a good number of trips that were identified wrongfully, such as extremely short trips (less than 0.25 miles) or some trips that were identified because of unidentified jump points (e.g. 5 mile trip in 5 seconds). To make sure these trips do not change the results of thiis study, a secondary filtering algorithm was developed. This algorithm evaluated the total haversine distance between origin and destination, and the total trip time and checked if any modification on destination points would make the trip possible if they did not satisfy time or distance criteria. This modification included testing and updating eligible points for assigning new destination and origin locations to ensure trip distance criteria was satisfied. The eligible points here were defined as the "free" points between two consecutive trips. For example, after ordering all points by time, if the trip k ended at point 85 and the trip $k + 1$ started at point 90. Here, free points would be points 86 to 90. By checking the distance between the origin and the eligible destination points, we identified some trips that would have been deleted if the modification was not made. Same method was also applied to find new origins if there were no eligible destination points. If we didn't find any points that make this trip possible based on the trip distance criteria, those trips will be deleted. In addition, the trips that had "potential jump point" tag for their jump point status as an origin or destination were checked one more time to ensure that there are no inconsistencies with previous and later trips. One example of these inconsistencies was the trips that were identified because of a large distance between two points but relatively small-time interval. Since in our algorithm, distance was the final and most important criteria for checking trip status (meaning distance only could be an indicator for assigning an end point status even if nothing else is meets the conditions), these sudden jumps between two distant points could be defined as trips. As described in Chapter 4.3.1, the trips that have potential jump points as their origin or destination were trusted less than another trip that might not have an origin or destination at a point described as potential jump point. If the origin of the trip k was tagged as "potential jump point", and the algorithm said there was trip between $point\ j$ and $j + 1$ with high-speed value between them, the trip k was deleted based on proximity criteria (less than 0.25 miles) between the destination of trip $k - 1$ and point $j + 1$. Please see Algorithm 3 below for details of this method. The total number of trips from the 153,391 users whose home is located inside the MPA boundary was 4,248,320 in a ten-day span. These trips were then filtered, as both trip origin and destination would be inside the MPA boundary. With this filtering operation, the total number of trips that were used in the Chapters 4.4 and 4.5 resulted in 3,807,003. This is a relatively large sample of trip origin and destination pairs compared with the number of observations used in the literature.

Algorithm 3: Trip O/D Filtering Algorithm

For all u_i, $k = 1$ **to** K **do**

  If $\omega_{ik} > 120\ mph$ **do**

    //origin having high speed value is a problem because trip chain between trip $k-1$ and trip k is noted as impossible

    Calculate ξ = haversine distance between the E_{ik-2} & O_{ik}

    //check the distance between the destination of trip k-2 and origin of the trip k

    If Z_{ik-1} = "potential jump point" **do** //since we trust potential jump points less

      If $\xi < distance_{threshold}$ **do**

        Delete trip k-1 from trip chain (delete O_{ik-1}, E_{ik-1}, T_{ik+1}, ω_{ik-1}, and Z_{ik-1})

        Connect trip $k-2$ and k //reforge the trip chain as trip k following trip k-2, change trip numbers respectively

      End

    Else //Z_{ik-1} = "normal" since jump points were removed

      If $\Delta D_{ik} < distance_{threshold}$

        For $y = \sigma_{ik}$ **to** $\alpha_{ik+''}$ **do**

          ζ_y = haversine distance between the O_{ik} & $[lat_{iy}, long_{iy}]$

          If $\zeta_y > distance_{threshold}$ **do**

            Assign $\mathrm{E}_{ik} = [lat_{iy}, long_{iy}]$ //assign a new destination, remove the old one

          Else

            If $y = \alpha_{ik+1}$ **do**

              Delete O_{ik}, E_{ik}, T_{ik}, ω_{ik}, and Z_{ik} //loop trips are deleted

              Connect trip $k-1$ and $k+1$

              //reforge the trip chain as $k-2 \rightarrow k-1 \rightarrow k+1 \rightarrow k+2$

            End if

          End

        End for

      End

    End

  Else

    If $\Delta D_{ik} < distance_{threshold}$ **and** $\Delta T_{ik} < time_{short}$ **do**

      For $y = \sigma_{ik}$ **to** α_{ik+1} **do**

        ζ_y = haversine distance between the O_{ik} & $[lat_{iy}, long_{iy}]$

        η_y = time between O_{ik} & $[lat_{iy}, long_{iy}]$

        If $\zeta_y > distance_{threshold}$ **and** $\eta_y > time_{short}$ **do**

          Assign $\mathrm{E}_{ik} = [lat_{iy}, long_{iy}]$ //assign a new destination, remove the old one

        Else

          If $y = \alpha_{ik+1}$ **do**

            Delete O_{ik}, E_{ik}, T_{ik}, ω_{ik}, and Z_{ik} //loop trips are deleted

            Connect trip $k-1$ and $k+1$

            //reforge the trip chain as $k-2 \rightarrow k-1 \rightarrow k+1 \rightarrow k+2$

          End

        End

      End for

    End

  End

End for

4.4 Mobility Option Comparability and Feasibility Evaluation

Generally, comparing and evaluating different trip modes has been done using mode choice models. These models use different trip values and sociodemographic characteristics to determine how users choose the mode they use for their trip. To apply this general method for comparing trip modes, this thesis defined general cost functions for all modes considered in this study. Since the focus of this study is not to determine these functions from preference surveys, this study used generalized costs from literature and transportation agencies around the nation. The cost function method enables straightforward comparisons between trip modes, thus aiding in our understanding of actions of a rational person. However, the cost-based methods evaluate trip modes independent of the build environment and quality characteristics. This means that a walking on a relaxing and safe walking trail is evaluated with same criteria as walking alongside an unsafe high-speed road with no proper sidewalk infrastructure. This limitation arises from using only mobility-based measures with these naïve cost functions, since mobility is concerned about how far you can in a given amount of time, but not how comfortable that trip is. Additionally, this creates another limitation, as pointed out in Litman (2003) where faster travel modes such as cars are favored highly compared with non-motorized modes (biking and walking). This is a problem because the cost function method could identify non-motorized (active) travel modes as not feasible (comparable to the baseline trip mode which is car) for especially for leisure and activity focused trips might be costlier but also preferable. Consequently, we also developed an alternative decision criterion for which travel-demand-relevant access based measures were prioritized in understanding the feasible mobility options. The idea with this method is to develop a binary index to assess the quality of each trip mode and evaluate each mode within its own context. For this quality criteria method, the condition of the infrastructure, topographical characteristics of the routes, and quality of the mobility option were evaluated for each trip based on the relevance of each criterion. Since we do not have any information regarding the trip-purpose, all trips are assumed to be all-purpose trips as defined in (DOT, 2016).

This chapter used OSMNx package for Python (Boeing, 2017) and OpenTripPlanner (OTP) package for R (Morgan et al., 2019) to build network graphs for different trip modes (driving network, biking network, walking network, and combination of all with time dependent public transit network). We found the routes for each 3,807,003 trips within their respective networks and relevant trip characteristics were used in cost function and quality criteria methods. OSMNx was used for walking, biking, and bike sharing on their respective built networks (walking network, biking network, and combination of both) because of the ease of modifications on the network nodes and edges thus allowing for incorporating build environment condition for these trip modes. Please see details of the modifications for these three (walking, biking, and bike sharing) modes below in Chapter 4.4.1 and 4.4.2. OTP package was used for the trip modes: car, ride hailing, and public transit with walking and as first & last mile and public transit with biking as first & last mile trip modes. Since OTP routing algorithm accounts for time dependent nodes, meaning public transit trips can only be satisfied if the trip start time and additional time required to reach the station were matching with the route that passed by the location of the destination and its stop time at that station. The default routing algorithm for OSMNx networks is Dijkstra's algorithm (Dijkstra, 1959) for finding routing with minimum travel time. OTP uses A* algorithm (Hart et al., 1968) for path searching with minimum travel time with time dependent nodes (for the public transit portion of the routing) and allows for configuration on routing parameters. Please see Appendix A.4 for details about drivable, bikeable, and walkable road types in their respective networks created with OSMNx and OPT. Also see Appendix A.4 for more details about the tools and routing configurations used with these packages. Below, we discuss the cost function and quality criteria methods for each mobility option considered in this study.

4.4.1 Mobility Option Comparability with Cost Function

The list of variables and notations used in this chapter

Notations:

$C_{A_k} = Cost\ of\ using\ automobile\ for\ trip\ k$
$C_{W_k} = Cost\ of\ walking\ for\ trip\ k$
$C_{PB_k} = Cost\ of\ using\ personal\ bike\ for\ trip\ k$
$C_{SB_k} = Cost\ of\ using\ shared\ bike\ for\ trip\ k$
$C_{PTW_k} = Cost\ of\ using\ public\ transit\ with\ walking\ as\ first\ \&\ last\ mile\ mode\ for\ trip\ k$
$C_{PTW_k} = Cost\ of\ using\ public\ transit\ with\ biking\ as\ first\ \&\ last\ mile\ mode\ for\ trip\ k$
$C_{RH_k} = Cost\ of\ using\ ride\ hailing\ for\ trip\ k$

Variables:
$t_k = total\ trip\ duration\ in\ hours\ for\ trip\ k$
$x_k = total\ trip\ distance\ in\ hours\ for\ trip\ k$
${t_{IV}}_k = in\ vehicle\ trip\ duration\ in\ hours\ for\ trip\ k$
${t_{OV}}_k = out\ of\ vehicle\ trip\ duration\ in\ hours\ for\ trip\ k$
${x_{IV}}_k = in\ vehicle\ trip\ distance\ in\ hours\ for\ trip\ k$
${x_{OV}}_k = out\ of\ vehicle\ trip\ distance\ in\ hours\ for\ trip\ k$

For routing with an automobile, this study used OTP as the primary tool with assistance from the OSMNx driving network map when we found no routes because of the mismatch in terms of timestamp between the travel demand data and timestamp of the networks used. This study generated the routes using the drivable road types listed in Appendix A.4. Based on these drivable network edges; we found the minimum travel time path (with respect to the speed limits of each road edges) between the origin and destination and noted total driving distance in miles and total driving duration in hours. It is important to note that the driving path between some origin and destination pairs were not found with OTP which is why OSMNx driving network was used to identify the paths for those OD pairs. The most plausible explanation for these unidentified routes is the time difference between the travel demand and the OSM based networks. While trip origin and destination information are from August 2017, OSM map that was used to build the OTP graphs was from July 2021. This time gap could mean some trips that was possible in 2017 are not possible anymore. This idea of mismatch between times can also explain the extra trips identified by using OSMNx which used OSM data from March 2021.

In general, utility/cost functions for automobiles include a combination of trip duration and its relative value of time cost and per mile costs, such as fuel and vehicle operating costs. Equation 4.1 shows the generalized cost function for a car. The coefficients represent the value of time (time cost) and driving costs (fuel, ownership, maintenance, etc.) respectively. The value of time is from the U.S. Department of Transportation (DOT) based on the all-purpose in-vehicle travel for intercity travel in 2016 from their Valuation of Travel Time guideline, which was noted as $20.40 (DOT, 2016). This price is then adjusted by the Consumer Price Index from September 2016 to August 2017 (1.02) to ensure consistency with travel demand data. The driving costs per mile are from American Automobile Association (AAA) for driving 10,000 miles per year. The value is listed as 73.54 cents and include ownership (insurance license, depreciation etc.), fuel, and maintenance (repair and tires) costs. This driving costs do not cover parking costs but Indianapolis in general does not have a large parking problem thus parking costs can be negligible. Because more than 92% of trips in Indianapolis are made with car (Census, 2020a), driving is also considered as the baseline value to compare with other modes to determine their comparability based on the cost function method.

$$C_{A_k} = -\,(\$20.75\ \boldsymbol{t_k} + \$0.7354\ \boldsymbol{x_k}) \qquad (4.1)$$

This thesis did the routing with walking on a new network build from OSMNx using walkable roads (please see Appendix A.4). These roads include all roads besides the ones pedestrians are not allowed to get on such as highways. The closest network nodes of origin and destination were found and the path with the minimum travel time between these nodes was calculated using this walk network. The cost function for walking trips is only based on the DOT guidelines which has the same value as car ($20.75) when walking is the main mode of transportation for all-purpose trips. Equation 4.2 shows the cost function for walking trips. Since cost function is based on per hour, the speed of 3 mph (Bohannon & Andrews, 2011) for walking is used to calculate trip duration from trip distance when finding the minimum travel time path (since speed is constant this is same as finding the shortest length path).

$$C_{W_k} = -\,(\$20.75\ \boldsymbol{t_k}) \qquad (4.2)$$

The routing with biking (personal bikes) was done on the biking network map created with OSMNx. The bikeable road segments from OSM are noted Appendix A.4. The elevation data discussed in Chapter 4.2.2, is incorporated into the biking network to enable impedance-based routing. Impedance based routing for biking allows for finding the most suitable paths for cyclists. The impedance function is based on findings from (Broach et al., 2012) where they find that average non-cyclists are willing to travel 1.72 miles if the alternative is 1 mile 2-4% up-slope for non-commuting trips and 1.37 miles for commuting trips. Equation 4.3 represents the relative weight based on the slope of an edge assigned to it. These weights are calculated based on the relative attribute values from (Broach et al., 2012) for commuting and non-commuting trips and taking their average based on DOT methodology for determining all-purpose value of travel time (21.4% commuting, 78.6% non-commuting for intercity travel). This means that an edge that has a length of 1 mile with 3% slope will have 1.65 impedance while another one with 2% slope will have an impedance of 1. These impedance values were multiplied with the actual length of each segment to create a new weighted length attribute for these. The route for personal bike was generated by minimizing the sum of these weighted lengths rather direct edge lengths. We assume constant speed on all types of roads even if the speed might differ in high impedance routes due to topographical characteristics of study area. This assumption is based on the very low likelihood of long sections of climbs. It is more likely that high impedance roads are relatively short and unlikely to cause noticeable speed differences along the path. Consequently, the total travel time and total travel distance are calculated based on the actual edge lengths not the weighted length attributes.

$$Impedance\ weight\colon \begin{cases} slope \leq 2\% \rightarrow weight = 1 \\ 2\% < slope \leq 4\% \rightarrow weight = 1.65 \\ 4\% < slope \leq 6\% \rightarrow weight = 3.54 \\ slope > 6\% \rightarrow weight = 10.39 \end{cases} \qquad (4.3)$$

Similar to the walking and driving trips personal bike use is assumed to be used as the main mode of transportation for all-purpose trips which is why the cost function for value of time for the trip distance is same as walking and driving, which is \$20.75. Outside of the travel time cost, personal bike use has associated cost with bike ownership and the energy required to operate the bike (food as fuel). Ownership cost per mile was calculated from the tax credit for travel by bike in UK which is calculated to be \$0.26 per mile (eBikesHQ, 2020). This value is derived from UK but considering the similar prices of bike and food, it can be applicable to U.S.. The average speed of 10 mph was used for both personal and shared bikes which was derived from (Jensen et al., 2010) where they found that average speed for intercity travel as 15 km/h for biking which roughly translates to 10 mph.

$$C_{PB_k} = -\,(\$20.75\ \boldsymbol{t_k} + \$0.26\ \boldsymbol{x_k}) \qquad (4.4)$$

Routing for shared bikes used the same bike network as personal bike use but with the routing criteria being minimizing the total travel time rather than minimizing the impedance. This is because the assumption with bike sharing users is, they would not check elevation data and would like to minimize travel time since they are less likely to be aware of the high impedance sections. The cost function for station-based shared bikes is very similar to the personal bikes. The only exception is the per hour cost for renting the bike. Indiana Pacers Bikeshare lists the cost as \$1/ride + 0.15/min (\$9/hour, adjusted from January 2022 to August 2017: \$7.86) while other bike sharing systems list their cost as \$1.5 / 30 minutes (in March 2022). Based on the specific bike sharing systems used during the trip, two cost functions were created. Equation 4.5 is for bike sharing trips where Indiana Pacers Bikeshare was used, and equation 4.6 is for the bike sharing trips where other bike sharing systems. In vehicle time for this equation refers to biking portion while out of vehicle time refers to the walking portion of the trip. This is because DOT lists walking access value of travel time (\$27.20 in 2016 dollars) higher than all-purpose travel. Both travel time values were converted into August 2017 time using CPI index

$$C_{BS_{1_k}} = -(\$20.75 + \$7.86)\, \boldsymbol{t_{IV_k}} - 1 - (\$27.66\, \boldsymbol{t_{OV_k}}) \qquad (4.5)$$

$$C_{BS_{2_k}} = -(\$20.75)\, \boldsymbol{t_{IV_k}} - (\$27.66\, \boldsymbol{t_{OV_k}}) \quad + \$1.5\ (every\ 30\ minutes)\ (4.6)$$

Public transportation system in Indianapolis has the lowest service area (1.026 sq km) and the lowest annual passenger km per capita (130) among all cities of similar size (Griffin & Sener, 2016). This might indicate the public transportation system in Indianapolis area do not have spatial major omissions across the region, meaning there are no areas that lack access compared with their neighbors (Griffin & Sener, 2016). The OTP package checks origin and destination geolocation as well as trip start time to see if this trip can be satisfied with existing infrastructure (Please see Appendix A.4 for default routing configurations and the modified routing attributes used in this study). In general, the routing with OTP considers the trip start time along with the origin and destination and assigns a route based on the time and distance thresholds. This is important because it allows for travel-demand-relevant access.

OTP has built-in tool where first-last mile trip mode can be defined for the routing algorithm. For this study walking and the personal bike was the only travel modes considered due to lack of consistent location data for emerging options such as e-scooters. The cost function for public transit is similar to the bike sharing option where in-vehicle and out of vehicle time are evaluated separately. The time value of in-vehicle time is \$20.75 while the value of out of vehicle time is \$27.66 in August 2017. IndyGo has several options for passes that can be used, but for consistency among users, the base ticket cost of \$1.75 was used for each trip. Different fare structures (monthly fare) were put in sensitivity analysis in Chapter 4.6. Since personal bikes has ownership and fuel costs to account for, cost function for walking as first & last mile and biking as first & last mile are different. Cost function for walking as first & last mile option is shown on equation 4.7 while cost function for biking as first & last mile option is shown on equation 4.8.

$$C_{PTW_i} = -(\$20.75)\,\boldsymbol{t_{IV_k}} - (\$27.66)\,\boldsymbol{t_{OV_k}} - \$1.75 \qquad (4.7)$$

$$C_{PTB_i} = -(\$20.75)\,\boldsymbol{t_{IV_k}} - (\$27.66)\,\boldsymbol{t_{OV_k}} - \$1.75 - \$0.26\,\boldsymbol{x_{OV_k}} \quad (4.8)$$

Ride hailing is an emerging transportation mode provided by Uber and Lyft inside the MPA boundary. Both companies offer a variety of price options for their ride hailing services but for this analysis the focus was on the cheapest alternatives which are the low-cost Lyft and UberX. The path was found by the minimum trip duration using the same driving network as the car. Ride hailing trips are similar in the sense of shared bike and public transit trips where in vehicle and out of vehicle travel time costs are evaluated differently. The average wait time is assumed to be out of vehicle travel time. According to the Uphail this value is 5 minutes for Lyft and 2 minutes for UberX ride (UPHAIL, 2022). Average of 3.5 minutes was then multiplied by the $27.66 to find the cost for out of vehicle travel time per hour ($1.61). The other costs of each system are listed as $1.25 base fee and $2.55 safe ride fee, 81 cents per mile, and 15 cents per minute for both options. Since the exact date of these values are not indicated, they were assumed to be same in August 2017. The cost function for ride hailing is shown below in equation 4.9.

$$C_{RH} = -(\$20.75 + \$0.0025)\boldsymbol{t_{IV_i}} - \$0.81\boldsymbol{x_i} - \$1.61 - \$3.8 \quad (4.9)$$

Cost of using every mobility option was compared to the cost of using a car as the trip mode and the trips where cost of using a certain was lower than using a car. Additionally, trip cost that are within the 1.3 times of the cost of car were also assumed to be comparable to account non-commuting trips such as leisure and activity focused. The assumption is these trips can be made with costlier alternative if reaching the destinations faster was not the main goal. This percentage is derived from the time value differences between business focused travel ($25.40) and personal travel ($19) (DOT, 2016). Additionally, this ratio was put in sensitivity analysis to see its impact on the final results in Chapter 5.2.

4.4.2 Feasibility of Mobility Options with Quality Criteria

To account for the constraints with the cost function method, a new method to evaluate the feasibility of mobility options were developed with several quality criteria. The quality criterion for automobiles is dependent upon if the trip started during the morning peak (7 am–8:59 am) or any other time. The morning peak was based on the study done by Karner (2018). If a trip started during off-peak time regardless of the trip duration, all trips are considered as feasible. The assumption here is that trips made during off-peak times represents trips for the user that was intentionally taken without much concern for trip duration or length. However, if a trip started during morning peak period during weekdays that likely meant that the user had to commute which is why a maximum threshold for trip duration was set. The 45 minute commuting trip threshold was based on the time threshold used in a similar accessibility study done by (Golub & Martens, 2014; Karner, 2018). If a trip is deemed to be infeasible, that doesn't mean the car trip would not happen. It is more representative of the limited access these users might have to work opportunities.

For walking trips, the walking network edges were modified to represent the status of the infrastructure when deciding on the route. While the cost function method utilized the entire walkable network, the quality criterion for walking modified the network to only have the nodes and edges that are walkable with safety and comfort in mind. These safety and comfort criteria were created based on the findings and suggestion from the WalkWays initiative (Walkways, 2016). This initiative is a partnership between non-profit organizations and Indianapolis Metropolitan Development Commission developed in 2016. The three criteria decided on by this study are listed below:

1. Sidewalk rating of sidewalks inside the Marion County
2. Posted speed limit of the adjacent road to all sidewalks in study area
3. Number of roads at all the intersections that are accessible by pedestrians in the study area

First, as described in Chapter 4.2.2 sidewalk rating dataset from DPW was used to quality of sidewalks inside the Marion County. Even though the dataset only had the edges inside Marion County, this is not a big limitation because walking is not likely to be used frequently for suburb to downtown travel. However, eliminating dangerous edges inside city limits enable capturing downtown trips that might not be feasible by walking. The posted speed limits, number of travel lanes, traffic volumes, and number of intersections were some of the characteristics that were highlighted in WalkWays report as important for pedestrian safety. We decided to use the number of intersections and posted speed limit of adjacent road to sidewalk as the two criteria to check since roads with large number of lanes typically also have high posted speed limits. Traffic count based on specific road segments data was not publicly available which is why it was not implemented in quality criterion. Based on the graphs depicted in the Walkways report (Figure 4-3), most roads in Indianapolis have a speed limit of 35 or 40, expect the heavily used segments which have posted speed limits above 40 mph. In addition, in the same report getting hit by a vehicle 40 mph was noted as high risk (1/10 chance of surviving), therefore we decided to eliminate edges where posted speed limit was higher than 40 mph (Walkways, 2016).

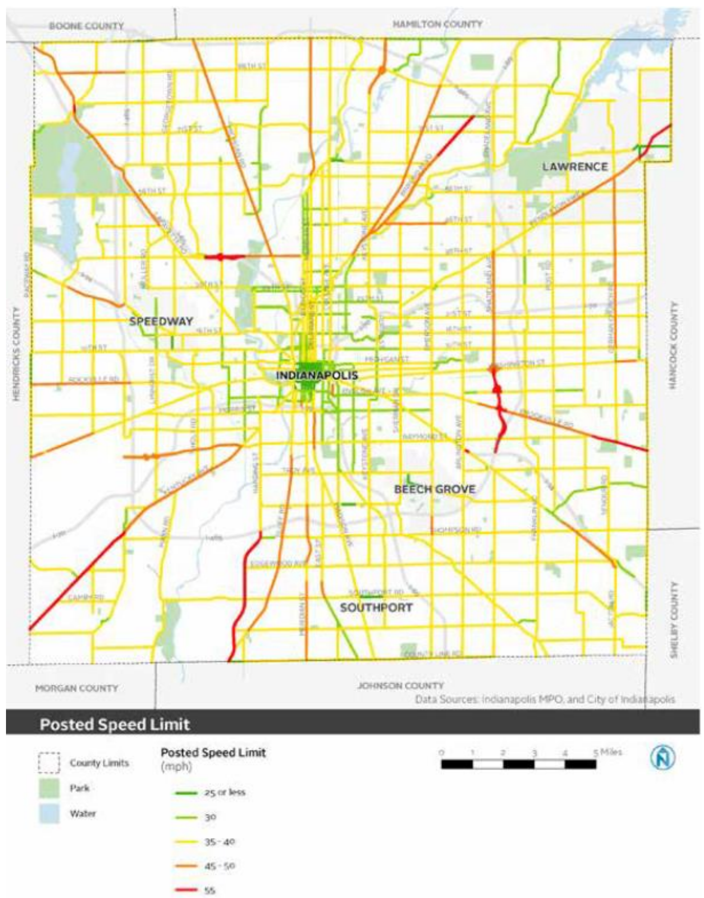

Figure 4-3. Posted speed limit of the roads in Marion County from WalkWays report (Walkways, 2016)

This thesis also made sure no intersection with more than 4 roads were part of the route generated, since a higher number of roads to cross would make the crossing more unsafe. If a route were to be generated with all these criteria, the feasibility of the trip was determined by the total trip distance traveled. For this, 2.4 miles was selected as the threshold which was the largest distance threshold in the review paper for active accessibility methods (Vale et al., 2015). The reason for selecting the largest threshold is to be able to capture leisure trips as well as the commuting trips.

When deciding on the quality criterion for bikes this thesis used a similar approach as (Winters et al., 2016) took. The methodology is explained in (Bike Score, n.d.) and it is based on infrastructure and topographical values. Bike paths and bike lanes are regarded as better options than shared structures. In order to check bikes lanes and bike paths, modification on OSMNx network was made to create cycleway network built in the OSM using the Python code developed by (Thyer, 2021). These cycleways include dedicated bike paths with separation, bike lanes on the road network, shared lanes with the bus, and dedicated bike tracks and trails. The quality criterion regarding cycleways was, if there are no road segments of the trip that utilize the cycleways and if the total trip distance is more than 2.5 miles, the bike trip would not be considered feasible when all other criteria (grade related) are satisfied. 2.5 miles was the maximum trip distance found as a threshold in active accessibility review paper (Vale et al., 2015) which was from (McNeil, 2011), that we decided as the maximum distance a user will be willing to travel for day to day use. If a bike route used bike lanes and bike paths for the trips over 2.5 miles, we assumed these trips to be activity or leisure related which is why they are left as feasible options. Bike Score also calculates the hilliness of an area where 10% would get a score of 0 and 2% would get a score of 100. Based on this hilliness quality criterion was defined as, any route with more than 6% mean grade would not be a feasible trip. Max grade threshold was set to be 15% to avoid very steep, almost non-bikeable section. This is to ensure everyone regardless of their athletic status can bike comfortably to meet their travel demand. Destinations + connectivity and bicycle mode share criteria were the other two criteria in Bike Score which were decided as unrelated for this thesis. Destinations + connectivity criteria were mainly concerned with the potential opportunities while in this thesis we define the opportunities through the observed travel demand. Bike share criteria was found to be not significant since most block groups in study area had 0% commuting trips done by bicycle on ACS.

The quality criterion for station-based bike sharing is mainly based on the maximum walking distance to reach a station and final location from the station. One mile distance threshold was used for both reaching to station from origin and to the final location. Any trip that requires longer walk time was denoted as unfeasible. This distance is significantly lower than the walking distance threshold used in walking as the primary mode, but the assumption is that using bike will require some energy thus it is less of a quality option if walking for long distances is required. One mile is commonly used in accessibility analysis as reviewed in (Vale et al., 2015) and is almost equal to the default value for OTP (1500 meters) for walking to access public transit. The quality criterion for public transit is mainly based on the out of vehicle portion of the trip. Maximum walking distance was set to be for 1 mile in total, a slightly larger distance than the default value of 1500 m defined in OTP package. While maximum biking distance was noted as 2.5 miles since this was the threshold value for personal bike use. Besides maximum walking distance number of transfers were limited to 2 for comfort. Quality criterion for ride hailing is exact same as the driving option. The only exception is the average wait time (3.5 minutes) is added into the total trip duration which was then compared with max time threshold if the trip started during morning peak time.

After feasibility of each mobility option was determined with the cost function and quality criteria method, we combined these methods into a stricter feasibility metric called "combined criteria". The trips were deemed feasible if and only if they were indicated as feasible with the quality criteria and the cost function methods.

4.5 Equity Analysis

After we identified the number of feasible mobility options for each trip origin and destination pair with cost function, quality criteria, and combined criteria methods, we categorized each trip based on their distance (short, medium, and long distance) and aggregated the results based on the users' home locations on census block group level. The average number of options for each user represent the feasible supply available regarding their observed travel demand. This information was then merged with sociodemographic data to create a dataset in which each census block group had an average number of mobility options available for their users with the cost function method, quality criteria method, and combined criteria. The sociodemographic data used in this portion was related to the disadvantaged groups which were defined in Chapter 4.2.3. Regression trees and multivariate multiple linear regression models were built to analyze if any certain sociodemographic groups are having fewer mobility options than the others (*Please see*

James et al., 2013; Johnson & Wichern, 2014 *for details of these methods*). By using a tree based and a parametric model, a more comprehensive understanding of the relationship between number of mobility options and sociodemographic characteristics can be achieved. For the multivariate multiple linear regression model, we modeled average number of feasible mobility options with three methods (quality criteria, cost function, and combined criteria) as the dependent variable. After eliminating the highly correlated sociodemographic variables we tested multivariate analysis of variance of the model and eliminated the insignificant sociodemographic variables among all three models. For the regression tree method we used 10-fold cross validation was used to improve the robustness of the model and eliminate multicollinearity (James et al., 2013). With the regression tree model, census block groups with a lowest number of mobility options are identified and the characteristics of these census block groups are recorded. Furthermore, as part of the social justice perspectives that were adopted in this thesis, disadvantaged census block groups were identified based on, above average percentage of households with no vehicles, above average percentage of population in poverty, above average of population of people of color, above average of households with more than one average room occupancy, above average percentage of people with disability, above average percentage of non-college educated population, and above average percentage of old and young population. These variables are selected as collection of metrics used in literature explained in Appendix B and especially with the studies that evaluated equity in terms of multiple modes such as (N. Chen & Wang, 2020; Golub & Martens, 2014; Meng & Brown, 2021). If a census block group satisfied at least three of the seven “disadvantaged” sociodemographic criteria, we defined them as the most “disadvantaged” areas. The total number of census block groups in this category was 348 (38%). In Appendix A.5, a map of census block groups that meet three and above criteria is shown for comparison. These census block groups were mainly located in urban core compared with the suburban areas. The number of mobility options available for these census block groups were compared with the others to see, if the people who need the greatest number of travel modes, have the options available for them. For this test Wilcox Test also known as Mann-Whitney test was used (Wilcox et al., 1986). The spatial equity was evaluated on county level averages and regarding urban core-suburban differences using Anselin Local Moran’s I analysis in ArcGIS (*Please see* Anselin, 2010 *for details of this method*) by identifying clusters and outlier census block groups.

4.6 Sensitivity Analysis

In this study a lot of assumptions regarding travel behavior needed to be made to capture different purpose trips with trip O/D algorithm and to compare feasible mobility options in terms of mobility and accessibility related measures. In this chapter, the sensitivity analysis methods for these decisions is explained. First, we tested the different fare structures used for the bike sharing systems and the public transit system evaluated in Chapter 4.4.1. The reason for this is to identify what percentage of trips could have been done with the cheaper fare alternatives. To test this, we compared the percentage of public transit and bike sharing trips that were identified as feasible for all price structure. For testing the monthly and annual fees, the fare cost was converted into daily cost. For bike sharing this was tested with the $5 monthly pass for those who receive social services and annual pass of $133.75 which allowed access to unlimited 60 minutes for every day and the per minute cost of 15 cents were added after every minute (Indiana Pacers Bike Share, 2022). For public transit 31 days pass with full cost ($60) and half cost was evaluated in addition to the base fare of 2-hour transfer ticket. Full fare is for everyone while youth (18 and younger), old (65+), people with disability can be eligible for half fare ($30) (IndyGo, 2021).

This thesis also test the sensitivity of the ratio used for comparing cost of alternative mobility options with cost of car values between 1-2 with 0.1 increments. The resulting graph is discussed in Chapter 5.2. Additionally, as discussed, the medium time threshold used in Chapter 4.3.2 was tested with values varying from 900-1200 seconds with 60 seconds increments.

5. RESULTS

This chapter represents the results in three main categories: trip validation, mobility options, and equity analysis. Chapter 5.1 represents the comparison between the trip O/D algorithm discussed in Chapter 4.3 and NHTS data to validate the trips used in this study and ensure the trips identified by the trip O/D algorithm represent the actual known travel patterns. In addition, overall trip characteristics such as average trip distance and average number of trips per day per user for the empirical study area on spatial level are presented. We represent the feasibility of mobility options with all three methods used in Chapter 4.4 (cost function, quality criteria, and combined criteria) on three different trip distance categories (short, medium, and long) and regarding two different spatial levels (census block group and county level) in Chapter 5.2. In addition, comparison between travel-demand-relevant access with public transit and traditional accessibility-based methods are shown to further aid in highlighting the importance of the framework identified for this study. Last, we discuss the equity implications of these mobility options on spatial and sociodemographic level in Chapter 5.3. For Chapter 5.3, the spatial equity is analyzed by identifying clusters of census block groups with high and low number of mobility options and the outliers in these clusters, as well as the county level averages. Sociodemographic equity is analyzed with multivariate multiple linear regression and regression trees.

5.1 Trip Validation

Trips identified with the proposed algorithm in Chapter 4.3 were found to be representative of the trip behavior based on the NHTS. The output from Chapter 4.3, trip origin and destination pairs, enabled better understanding of travel demand from each individual user and provided the input for the mobility option comparability and feasibility evaluation in Chapter 4.4. In Figure 5-1, the distance distribution of these trips is shown. As seen from the figure, the trips can be roughly categorized into equal sized bins, which was described as short distance trips (less than 1 miles), medium distance trips (between 1-3.5 miles), and long distance trips (longer than 3.5 miles).

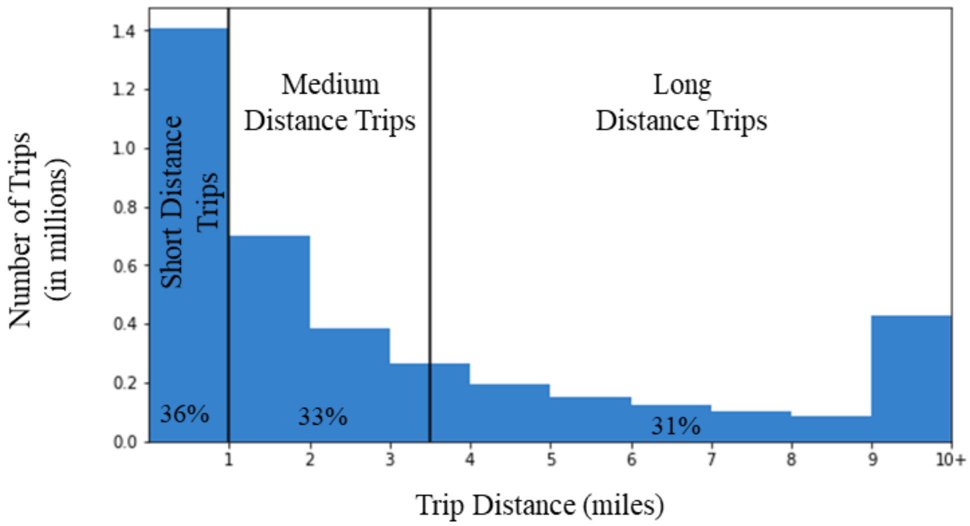

Figure 5-1. Distribution of trip distances and the categories determined

For the rest of this chapter when NHTS data is mentioned, it is referring to the weighted NHTS Indiana only data. This filtration was done to ensure representatives of NHTS data for the empirical study area. If we plot the trip distances to the same categories as weighted (trip based) Indiana NHTS as shown in Figure 5-2, many of the bins are consistent among the two, except for trips less than a mile and longer than 10 miles. This trip distance differences can also be seen if we group trips from NHTS into the same distance categories using the same cutoffs (1 mile and 3.5 miles), where the short trips (36% in this study and 20% in NHTS) and medium trips (33% in this study and 30% in NHTS) would be less common than the long trips (31% in this study and 50% in NHTS). For short trips, the difference is because our algorithm can separately identify sub-tips based on trajectories (e.g., walking from garage to destination), however as self-reported travel diaries used in NHTS only includes the main trips in travel diaries. Additionally, since in this we are constricting the trip origin and destinations to be inside the MPA boundary, it is likely that longer statewide trips are not recorded as part of our algorithm, thus the differences between longer trips. Furthermore, these longer statewide trips would not affect the equity analysis since only feasible options would be private vehicle (there are regional bus and rail systems but the feasibility methods we developed does not cover those trip modes). Therefore, the small differences between NHTS, and our algorithm are negligible, and we can say that trips this thesis identified are validated in terms trip distance.

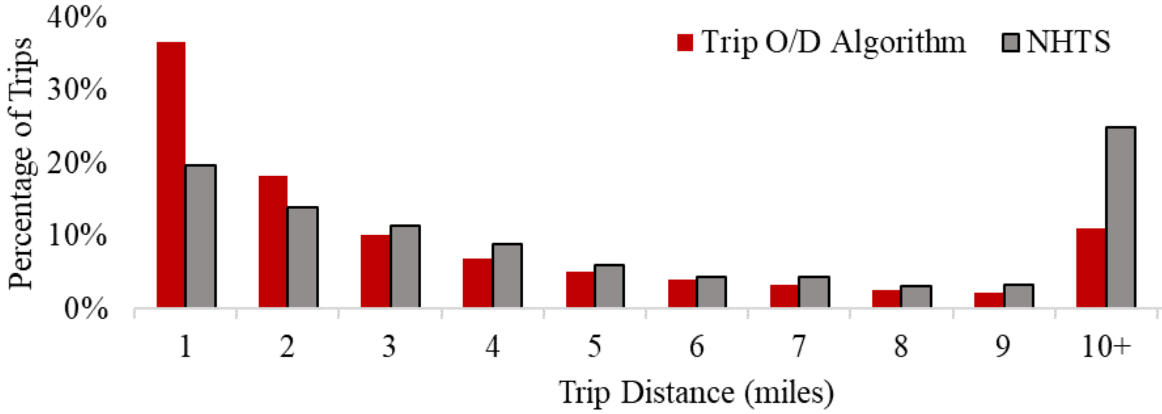

Figure 5-2. Distribution of trip distances with our algorithm compared to the NHTS.

We also plotted the spatial differences in terms of trip distances in MPA in Figure 5-3 and found that people in suburban areas travel longer distances on average compared with people in the urban core. Anselin Local Moran's I cluster, and outlier analysis showed that people in suburban areas are indeed traveling more distances compared with the people living in the urban core areas. As it can be seen from Figure 5-4, suburban areas comprise the entire high clusters while downtown Indianapolis, Fishers, and Carmel areas make up the low clusters. This also indicates that these areas with low trip distance averages have the necessary opportunities in proximity.

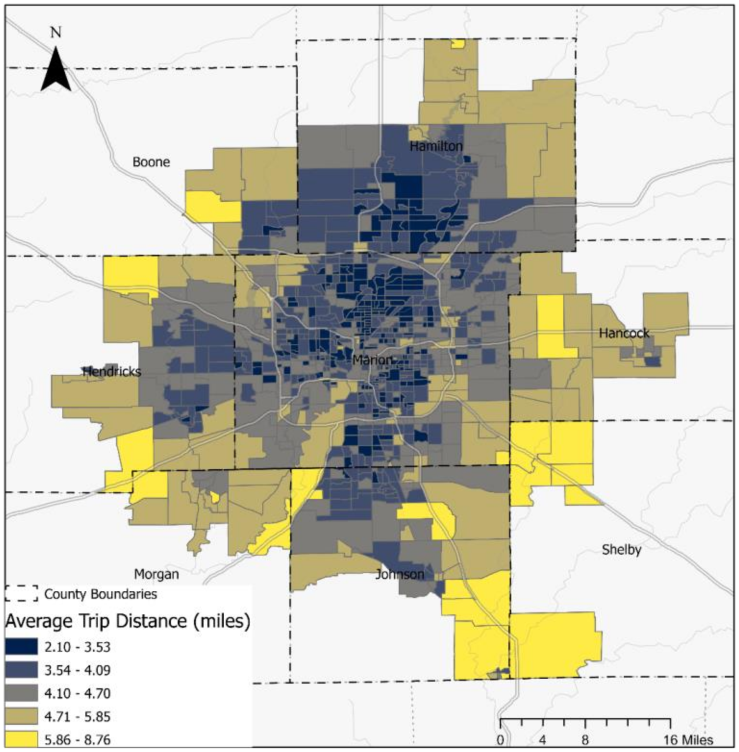

Figure 5-3. Average trip distance per user in miles in census block group level.

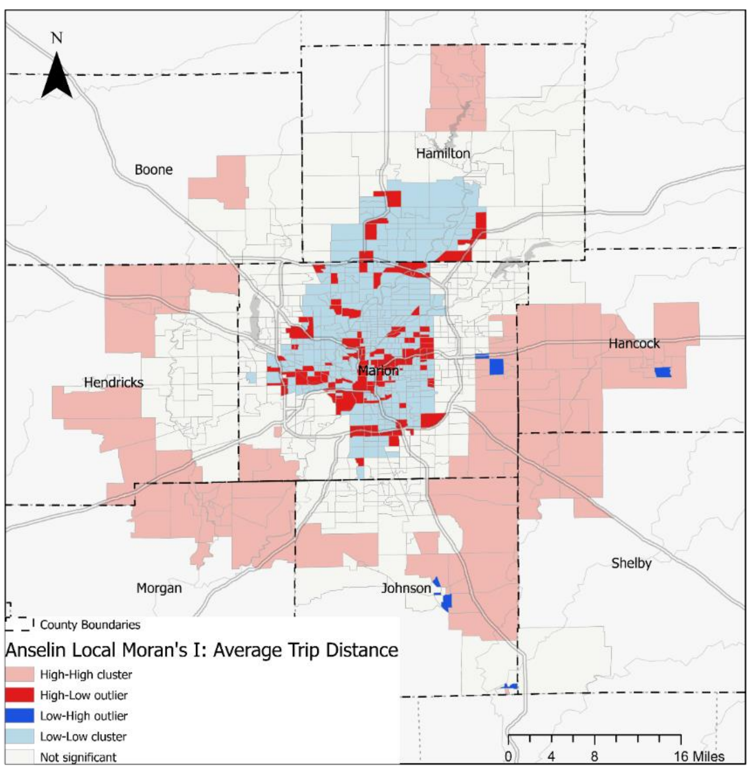

Figure 5-4. Anselin Local Moran's I analysis for trip average trip distance on census block group level.

In addition to the trip distance, start time and end time of the trips were compared with NHTS and temporal pattern was also validated. In Figure 5-5 we plotted the trip start times with our algorithm and NHTS data to validate our trips in terms of trip start times. Comparing these trips start times with NHTS trip start times reveals almost identical match except during early morning hours. This could also indicate some trips that happen at night were captured with our method such as drive through trips or trips for socializing. Afternoon peak time (between 4-6 pm) is the most common trip time which could be explained by the large number of trips done at the end of work hours. Morning peak (between 7-8 am) is also very common for people to start traveling but not as much for ending their trips. Interestingly, as seen in Figure 5-6 there are also many trips that end around 3-4 am using our algorithm. Comparing the trip end times with the number of points in an individual trip trajectory showed that the trips with less than 4 points in their trajectory (54% of all trips) have a larger share of trips that end around this compared to all others. Furthermore, short trips are two times more likely to have trips that end in this timeframe.

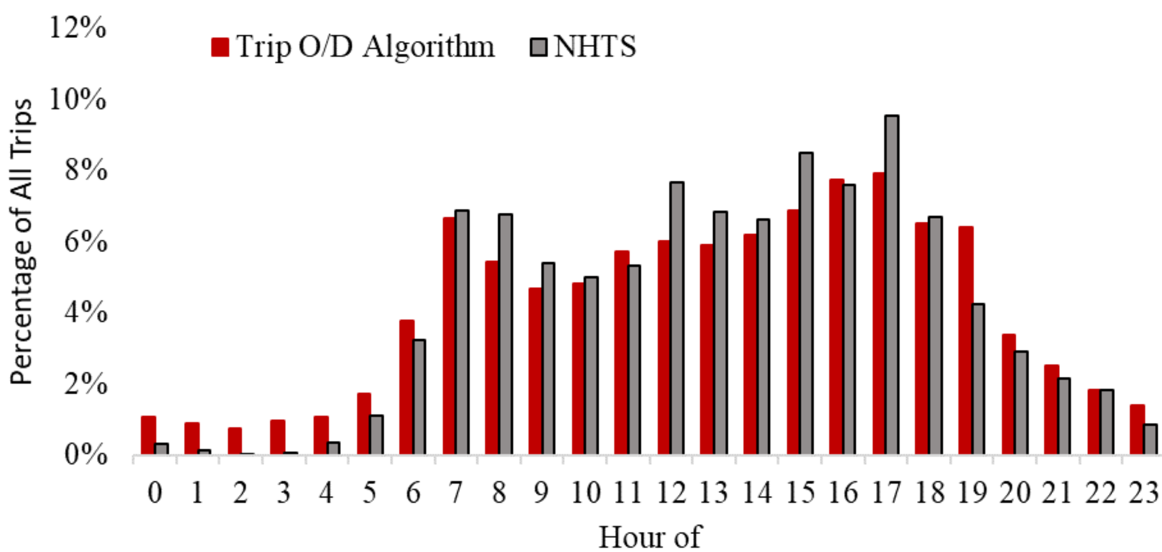

Figure 5-5. Distribution of trip start times with trip O/D algorithm and NHTS data.

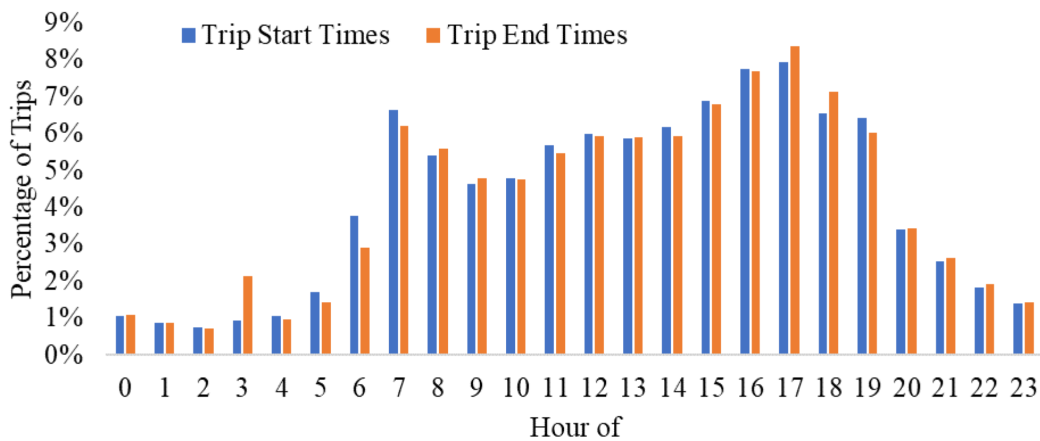

Figure 5-6. Distribution of trip start and end time with trip O/D algorithm.

Lastly, this study also evaluated the ratio of the number of users whose home locations were identified inside a certain census block group and the adult (18-64) population from ACS 2017 to ensure each area is represented somewhat equally. While we see some census block groups with very low ratios compared with other census block groups (dark blue areas in Figure 5-7) generally other areas have similar number of trips per day per user. The yellow census block groups represent the block groups where this study identified more users as having their home location than the actual estimate value from Census. The two block groups where the number of users estimates were higher than the Census value are where the Lucas Oil Stadium, Indianapolis Zoo and Indianapolis Riverwalk are located. This could indicate the method we used for identifying home locations falsely identify some user's home location due to some users having only a small number of GPS points at these locations at the nighttime. In addition, most of the areas with low percentage coverage around the downtown area are non-residential areas thus the lower spatial representation of these areas is negligible. With that we observe spatial representation similar to the real world thus ensuring the results gathered in the following chapters will be applicable to the study area.

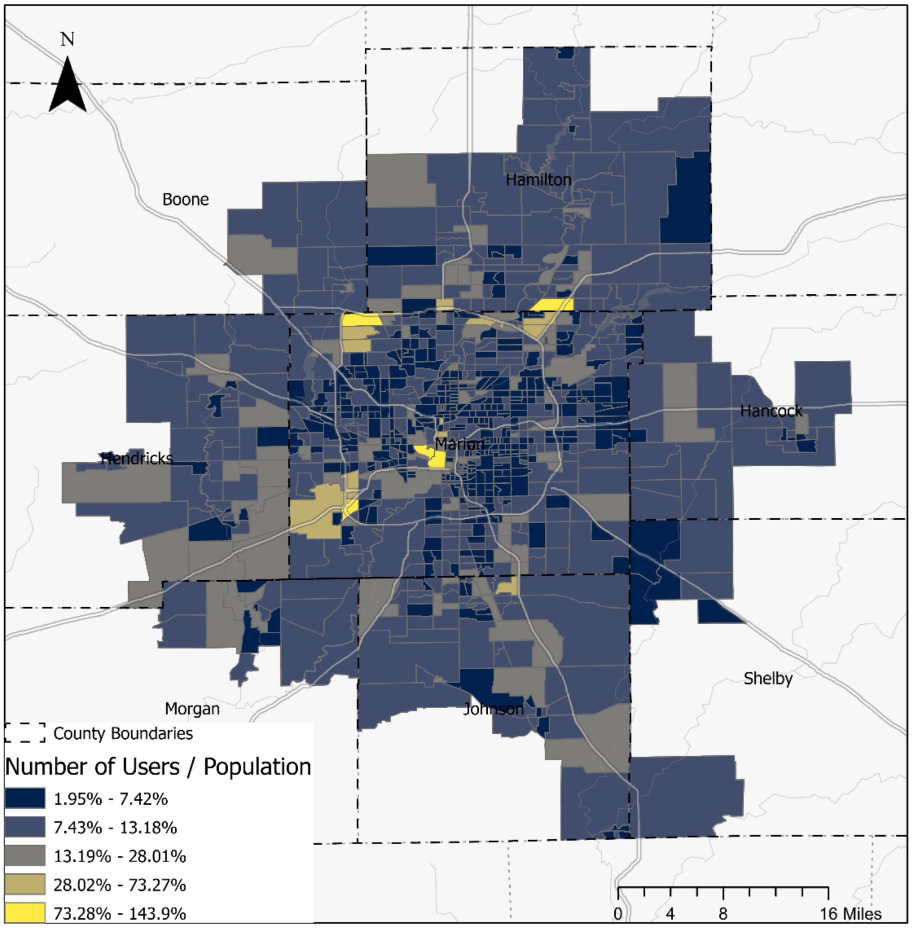

Figure 5-7. The ratio of number of user identified with our method and population of 18-64 in ACS 2017

5.2 Mobility Options

5.2.1 Feasible Mobility Options with Trip Distance and Different Methods

This thesis found that the number of feasible mobility options changes regarding the method (quality criteria, cost function, and combined criteria) used and is spatially heterogenous. First, the number of mobility options available is evaluated with comparison between cost function, quality criteria, and combined criteria methods with respect to the three trip distance categories. Furthermore, individual trip mode characteristics were evaluated in terms of their feasibility and which criteria were the reason for assigning trip as unfeasible. Last, this thesis compared the number of mobility options in different counties with all three feasibility methods and census blocks inside the MPA boundary with combined criteria method to highlight the spatial differences. We found that using a combined criteria of quality and cost-based metrics to assess travel-demand-relevant access is vital because of the role infrastructure constraints and the individual travel demand affect the feasibility substantially. Additionally, number of options were found to be spatially heterogenous.

First, we draw the percentage of feasible trips with combined criteria method when the cost ratio we used for comparing all trip modes (1.3) with values between 1 and 2. Cost ratio of 1 only identifies trip modes as feasible if and only if the cost of that certain trip mode is less than the cost of car. Cost ratio of 2 means that the cost has to be less than double the cost of car. In Figure 5.8, the percentage change in feasible trips is represented for all six of the trip modes that are compared with car. From this figure we see that all trip modes plateau even if the cost ratio is increasing. This is because the feasibility by quality criteria is the limiting factor in this evaluation. This is most clearly seen with bike where the cost of bike being comparable to the car does not affect the feasibility with combined criteria method since there are only a little less than 50% of trips that are feasible with quality criteria. We also see that the cost plays an important role for public transit and ride hailing trips. Last, only a very small percentage of walking trips are feasible showing that doubling the price of using the car still results in car being the cheaper option compared with walking. This proves the limitation of using only cost-based methods.

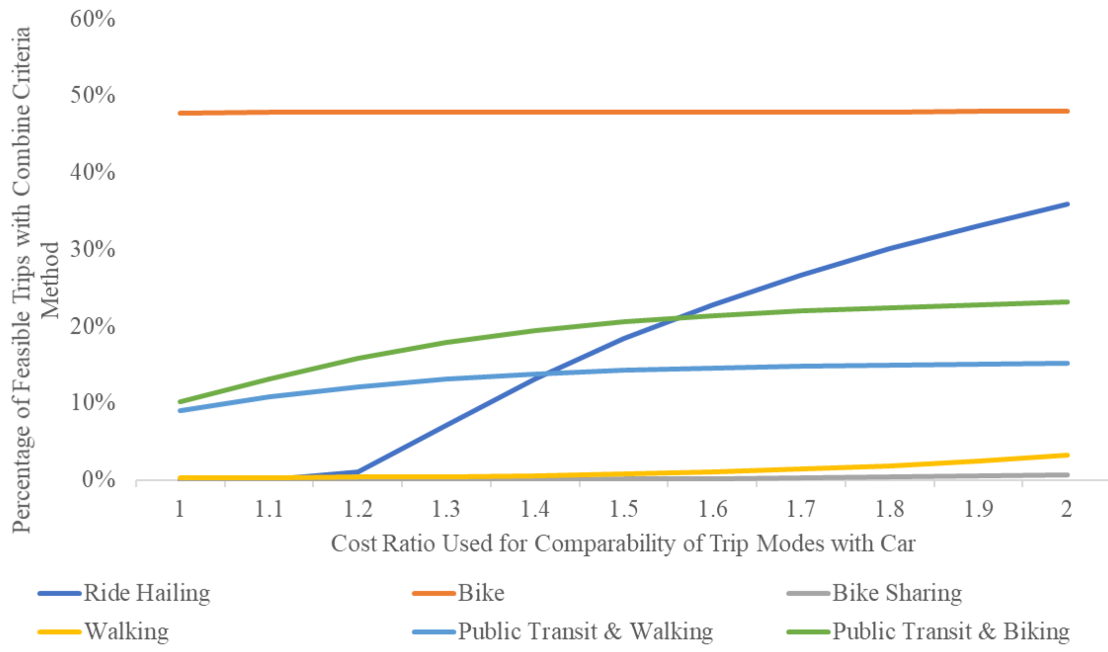

Figure 5-8 Percentage of feasible trips with combined criteria method when different ratio values are used for comparing cost of mobility options with car (1.3 will mean the options with cost below 1.3 times of the cost of car will be feasible)

Comparing the number of mobility options determined as feasible with cost function and quality criteria shows the importance of using a combined criteria method. First, quality criteria method has a higher average number of feasible mobility options than cost function based feasible options as it can be seen from the Figure 5-9. This shows that the infrastructure is there, but the option is not as desirable comparing with driving with the cost function that this study has used. In addition, as the sensitivity analysis on public transit fare shows, high initial costs and higher out of vehicle costs associated with multi-modal trip modes such as bike sharing and both public transit options and with ride hailing is the main reason for these trip modes to be denoted as infeasible. These modes compared with the use of automobile (baseline) likely to be costlier in shorter trips; but for longer trips, ride hailing, and public transit are two of the best alternatives in terms of cost feasibility. The only trips that were noted with zero options with the cost function method were the trips where no car trip route was found based on the available infrastructure in OSM for drivable roads. This is either because of location not being accessible by driving or as because of the mismatch between trip origin and destinations and the timestamp of the OSM nodes and edges (some roads might not be accessible because of construction that was not existing back in 2017).

These trips made up only 0.05% of all trips which meant that almost all trips generated by our trip O/D algorithm were doable with at least one trip mode with cost function method. On the other hand, 2.85% of all trips didn't have a single feasible option with quality criteria. Most of these trips are in the long distance category which is expected because of the time threshold and distance thresholds used as part of the quality criteria method. Shorter trips had larger number of mobility options identified as feasible with the quality criteria compared with the cost function in the same category (short distance and medium distance). This is interesting because some trip modes that might be feasible to use might not be preferred due to the higher cost associated with them compared with the use of cars.

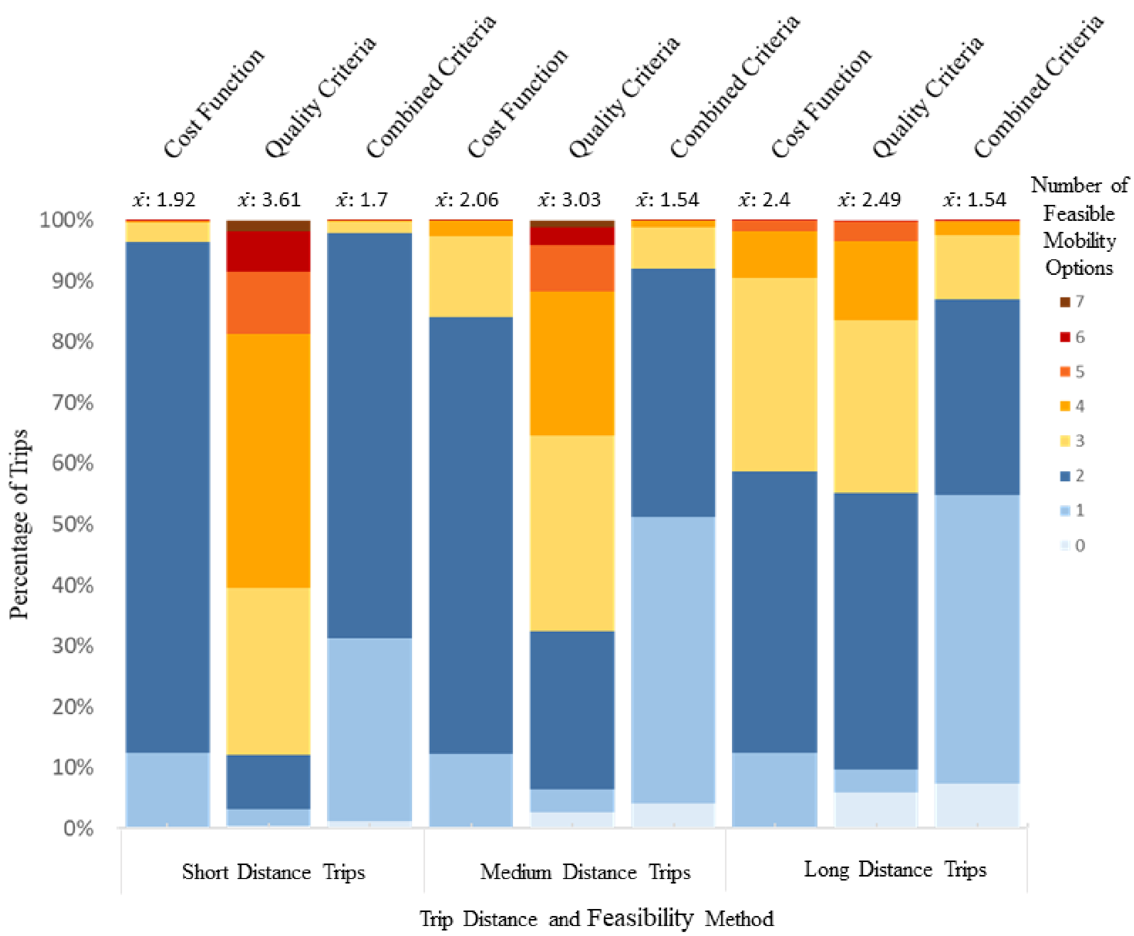

Figure 5-9. Number of mobility options available with quality criteria, cost function, and combined criteria; grouped by trip distance ($\bar{x}$ represents the average)

In addition to the cost function and quality criteria methods, the number of mobility options based on a combined criteria were evaluated to see if mobility options are heterogenous with respect to the distance. As it can be seen from the Figure 5-9 and 5-10, the average number of mobility options with combined criteria method goes down as the distance increases. This is potentially due to driving being substantially faster (thus cheaper) compared with all other alternative mobility options especially in short and medium distance trips and due to infrastructure limitations since no bus or bike share cover the larger area served with long distance trips .When looking at the individual characteristics of each trip mode and their feasibility in Figure 5-10, we see that bike and walking trips are less likely to be feasible as the trip distance increases, while public transit trips become more feasible as the trip distance increases. Since time cost is the biggest cost component with cost function method and the trip level benefits of other modes (especially walking) are not accounted for almost none of the trips being feasible by walking. Because of this this study evaluated the feasibility of walking based on the quality criteria only. Cost of using bikes is comparable to the automobiles however with quality criteria, percentage of feasible bike trips drops substantially by the trip distance. For biking trips with quality criteria, the cycleway criteria introduced accounts for 27% of all unfeasible long distance trips and 49% of the unfeasible medium distance trips while most trips were noted as infeasible due to max grade for a section of the route passed 15% gradient. This further consolidates the idea proposed by (Vale et al., 2015) regarding the importance of using road gradient in bike accessibility studies.

Similar to walking, bike sharing and ride hailing have very small share of feasible trips using the cost function method. This is mainly due to high out of vehicle time cost and fare costs for using the trip mode. Bike sharing also has a very small share of trips as feasible with quality criteria which can be explained by the low number of bike stations located in the area. The percentage of trips that can be done with bike sharing system was only 4.89% of the trips with quality criteria. To understand if the coverage is the main reason for this, this thesis also looked at the census blocks in the downtown area and found these census block groups in close proximity to these bike sharing stations had an average of 41% of their trips deemed as feasible with bike sharing. This shows us that bike sharing can satisfy the travel demand and can be used to increase the alternative options if the spatial coverage is increased. Wait time associated with ride hailing did not impact feasibility much more for trips compared to the same trips with car as it can be seen from the almost identical quality criteria graph for both car and ride hailing in

Figure 5-10. In total only 108 more ride hailing trips were found infeasible compared to the car trips with quality criteria. However, when we look at the cost function only almost no ride hailing trips for short or medium distances were comparable to cost of car indicating high initial fare costs. Those costs become somewhat comparable with long distance trips as the ownership cost of a car is dependent on the trip distance.

Public transit is one of the best alternatives for replacing automobile trips when quality of service is the concern since the differences between cost function and quality criteria is minimal. As discussed in Chapter 4.6, there different fare structures for public transit and bike sharing systems were also tested. For public transit, base fare method identified 13.1% of all trips as feasible for public transit with walking and 11.6% of trips as feasible for public transit with biking using the cost function method. Full fare monthly pass identified 33.3% of all trips as feasible for public transit with walking and 25.5% of all trips as feasible for public transit with biking while half fare monthly pass identified 40.1% of all trips as feasible for public transit with walking and 30.9% of all trips as feasible for public transit with biking. Bike sharing trips did not have significant differences in percentage of trips that were identified as feasible with the cost function method. This shows that the value of time is the primary factor for determining the feasibility of these modes however the fare cost play a considerable role when determining feasibility using the cost function method. Since the differences were minimal for bike sharing systems and the monthly cost structure for public transit trips assumed that every trip is doable with public transit when calculating the cost (entire travel diary for users is also not know), we decided to use the base fare for both options.

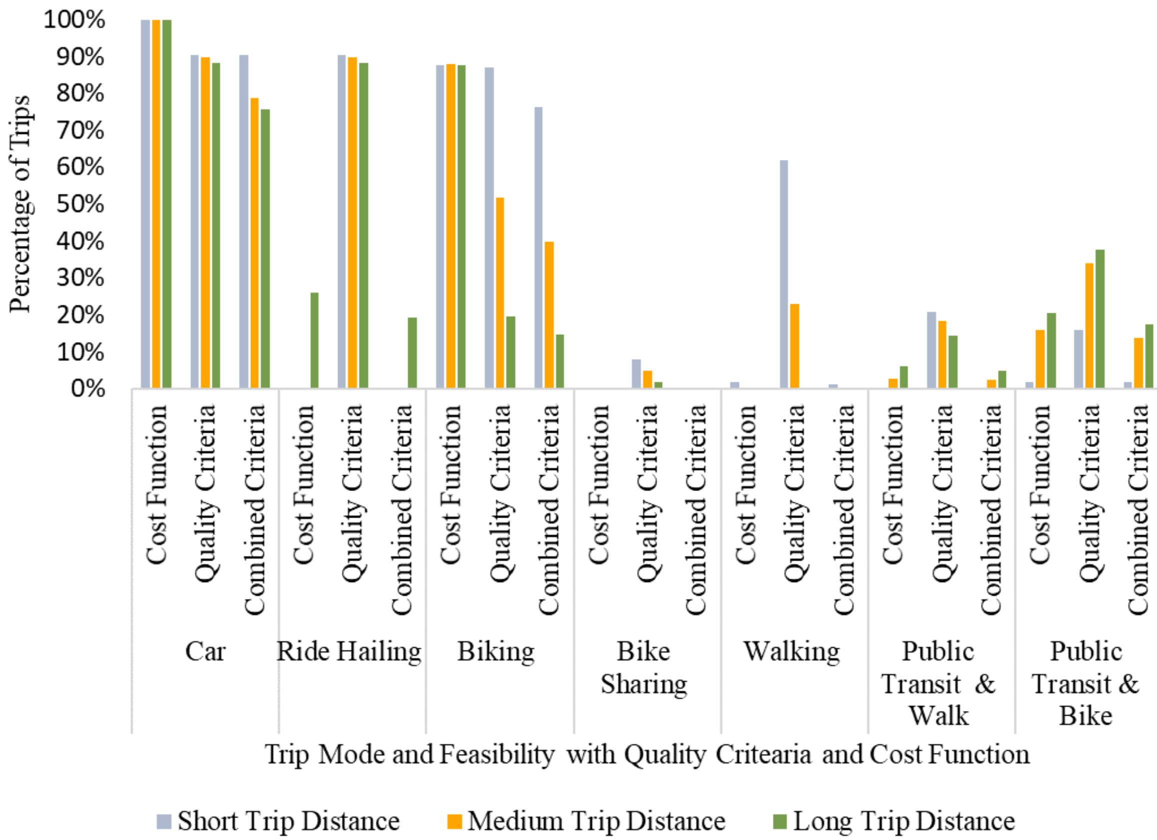

Figure 5-10. Individual mobility options feasibility from quality criteria based on the trip distance.

Figure 5-10 also shows the downward trend with relation to distance with biking and walking; however, the distance is not the only reason for this downward trend. First, out of all short distance trips, 62% of them were determined as feasible with walking while only 23% of medium distance walking trips were feasible. For both categories, around 5-6% of the trips were deemed infeasible only due to the walking alongside high-speed roads. Interestingly, short trips, where the haversine distance between origin and destination is less than one mile, had 2% of the trips noted as infeasible due to network walking distance being higher than the threshold. This shows the alternative routes had to be taken likely due to the lack of proper infrastructure for walking. This is because as described in Chapter 4.4.2, network edges with bad sidewalk rating were removed from the walking network for identifying routes with walking using quality criteria. In addition, this metric to account for the quality of the existing infrastructure is the reason for the biggest proportion of infeasible trips. 29% of the short trips and 37.33% of medium distance trips were identified infeasible due to the bad sidewalks and existence of large intersection along the path. The rest of the canceled trips are due to the combination of distance and high-speed criteria (0.5% for short, 9% for medium). These percentages could indicate the most important criteria for determining feasibility for walking trips is the infrastructure (existence of sidewalk) and pedestrian friendly design (easier to cross intersections).

Table 5-1. Percentage of walking trips that were denoted as infeasible with certain quality criteria.

	Short Distance Trips	**Medium Distance Trips**
Every criterion met	61.8%	23.01%
Only high-speed criterion not met	6.38%	5.08%
Only distance criterion not met	2.22%	25.11%
Distance and high-speed criteria not met	0.52%	9.47%
Infrastructure constraints	29%	37.33%

5.2.2 Travel-Demand-Relevant Access

To understand the differences between proximity-based accessibility and travel-demand-relevant access this thesis also performed a proximity-based accessibility study. For this we evaluated the public transit access based on proximity to transit stops (within ¼ mile) from their home location for every user. Then the feasibility of trips that start 400 m (~ ¼ mile) of home location of the user were evaluated to find if there are differences between accessibility and feasibility. Each trip distance (short, medium, long) and method (cost function, quality criteria, combined criteria) pair showed that the incorporating travel demand in to access evaluation identifies access with public transit much less compared with the traditional accessibility-based studies. This is also shown in Figure 5-11 where the differences between traditional proximity-based accessibility and travel-demand-relevant access are shown for different trip distance categories with combined criteria method. On y-axis in Figure 5-11, the cumulative distribution of the ratio between the trips that were identified as feasible with combined criteria method and trips that were assumed to be feasible with traditional accessibility method is shown (If a user had 5 short distance trips and only 4 of them are feasible with either public transit with walking or public transit with biking using the combined criteria method, the ratio will be 80% on y axis since accessibility-based study assumed that this user would have 5 accessible trips). Another important takeaway from this graph is more than 15,000 users that were identified as having public transit access, do not have any feasible public transit (the 0% on y-axis).

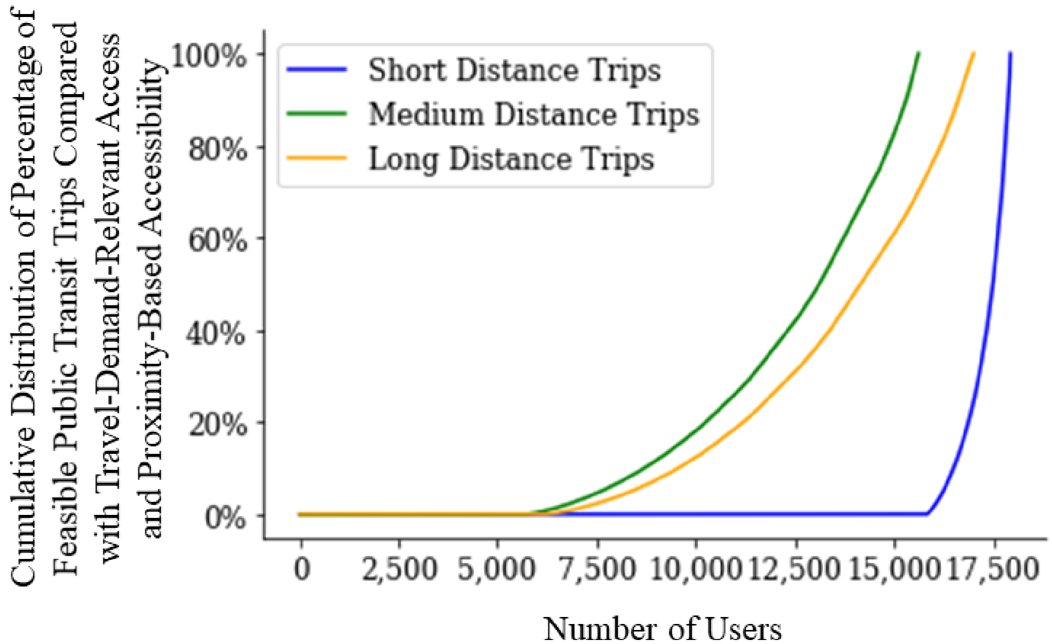

Figure 5-11 Cumulative distribution of users with public transit access based on proximity to transit stops (within ¼ mile) from their home that also have multiple trips starting from within ¼ mile of their home location compared with their feasible public transit with walking or biking as first and last mile mobility option. Larger area below the lines shows traditional accessibility-based method matches with our combined criteria feasibility method. Lower area below lines indicates many misidentified public transit access (you live close to the public transit but your travel demand can't be satisfied with the existing bus infrastructure)

If we look at all the trips that were assumed to be accessible, we see that 40% of trips were infeasible with quality criteria and 76% with cost function, and 76% with combined criteria method. This is an important finding of this study since it indicates that living close to a bus station does not mean you have access if the travel-demand-relevant access is not evaluated.

5.2.3 Feasible Mobility Options on Spatial Level

Last, this study evaluated the feasible mobility options from the spatial perspective. We found that number of mobility options are spatially heterogenous independent of the method used for assessing feasibility. The county level averages, and number of available option percentages are shown in Figure 5-12. County level average number of options were evaluated to understand if there are differences among different counties inside the MPA boundary. The counties inside the MPA boundary and the number of users home locations that located inside each are as follows: Shelby (631 users), Morgan (2,746 users), Boone (3,283 users), Hancock (5,685 users), Johnson (13,876 users), Hendricks (15,246), Hamilton (32,049), and Marion (79,875 users). From Figure 5-12, it is clear that Marion County has a greater number of average mobility options compared with the other counties. While in this study Marion County was the only one whose public transit system was evaluated, as shown in the Figure 3-1 in Chapter 3 Boone, Hendricks, and Johnson counties have their respective public transit systems. These systems could increase the number of available mobility options for these areas but for the Hamilton, Hancock, and Shelby County no unaccounted mobility options exists. The counties with the lowest average mobility options are Morgan, Shelby, Hamilton, and Hendricks County in order. Hendricks (2^{nd}), and Hamilton (4^{th}), are both in the top 10 in terms of percentage of households with at least one vehicle which likely is a result of the forced automobile ownership. Shelby and Morgan counties also have over 95% vehicle ownership thus the forced vehicle ownership trend can also be observed from these counties as well. These results are similar to the equity study by (Pyrialakou et al., 2016) which evaluated accessibility in Indiana and found the limited accessibility in suburban census tracts in these counties compared with the Marion county. Same paper also identified census tracts in Hancock County and Shelby County as very high transport need but did not identify as much transport need in Hamilton or Hendricks County. One other important thing to note is regarding the Marion County. Since Marion County is the only county with public transit, we expect the average number of mobility options to be substantially higher than the other counties. Even though the average is higher, the median value (2) is same as all other counties expect Shelby and Morgan (1). This is interesting because Marion County also has the highest standard deviation, indicating that some areas in Marion County might have less options compared with others. Therefore, we decided to also look at the number mobility options on census block group level. To understand which modes are less equitable spatially we looked at percentage of feasible mobility options on census block group level with combined criteria method.

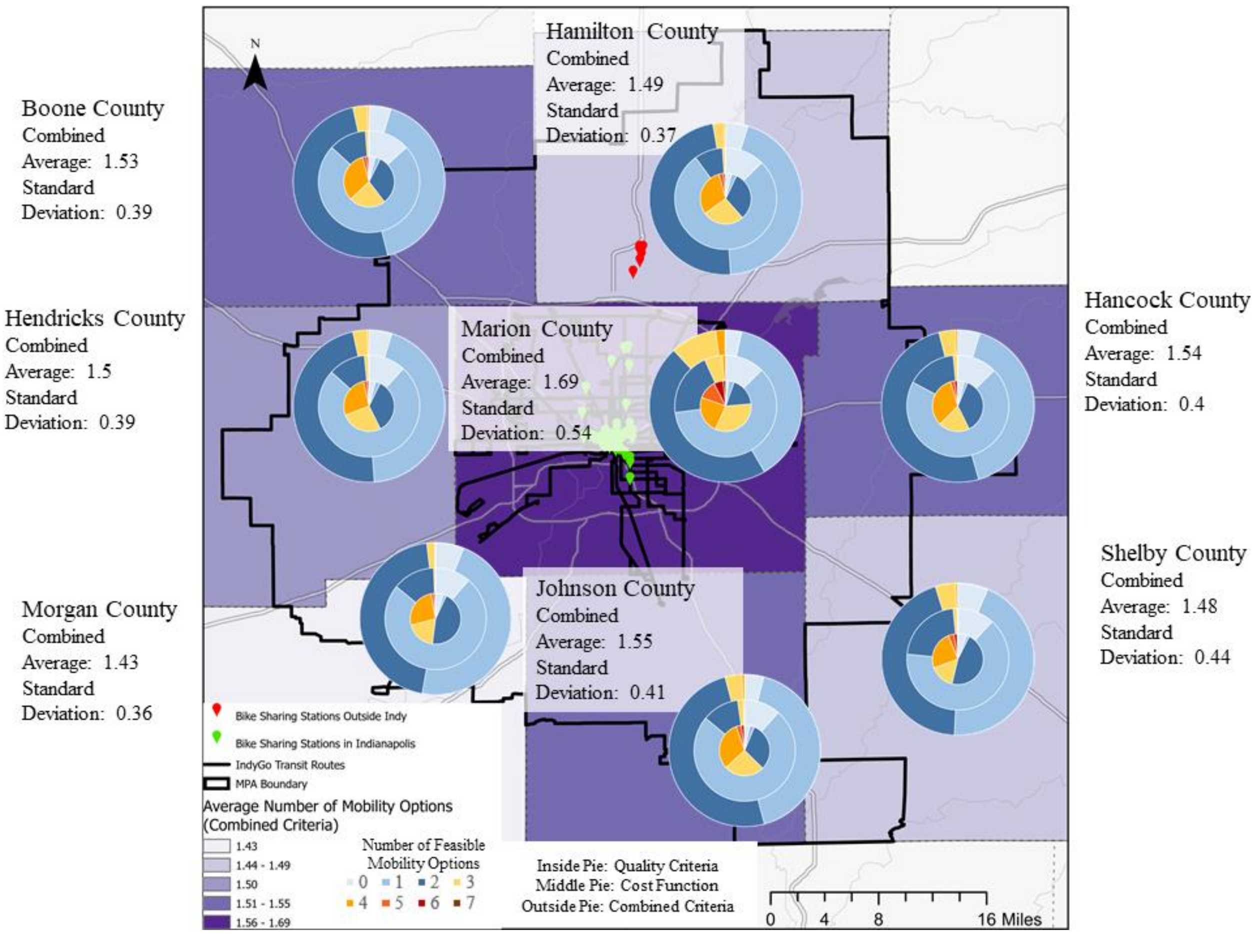

Figure 5-12 Number of mobility options available based on home location county with quality criteria (inner circle), cost function (middle circle), and combined metric (outer circle) overlay on MPA boundary with public transit lines and bike sharing stations. Shelby and Morgan County were the only two counties with median of 1, the other counties had median of 2

Looking at the spatial characteristics of each trip mode using combined criteria method we can see spatial differences in all trip modes. In Figure 5-13, we see that car trips from the edges of the MPA boundary had the largest number of trips deemed as infeasible. This indicates that people living in these areas need to more than 45 minutes to reach work opportunities.

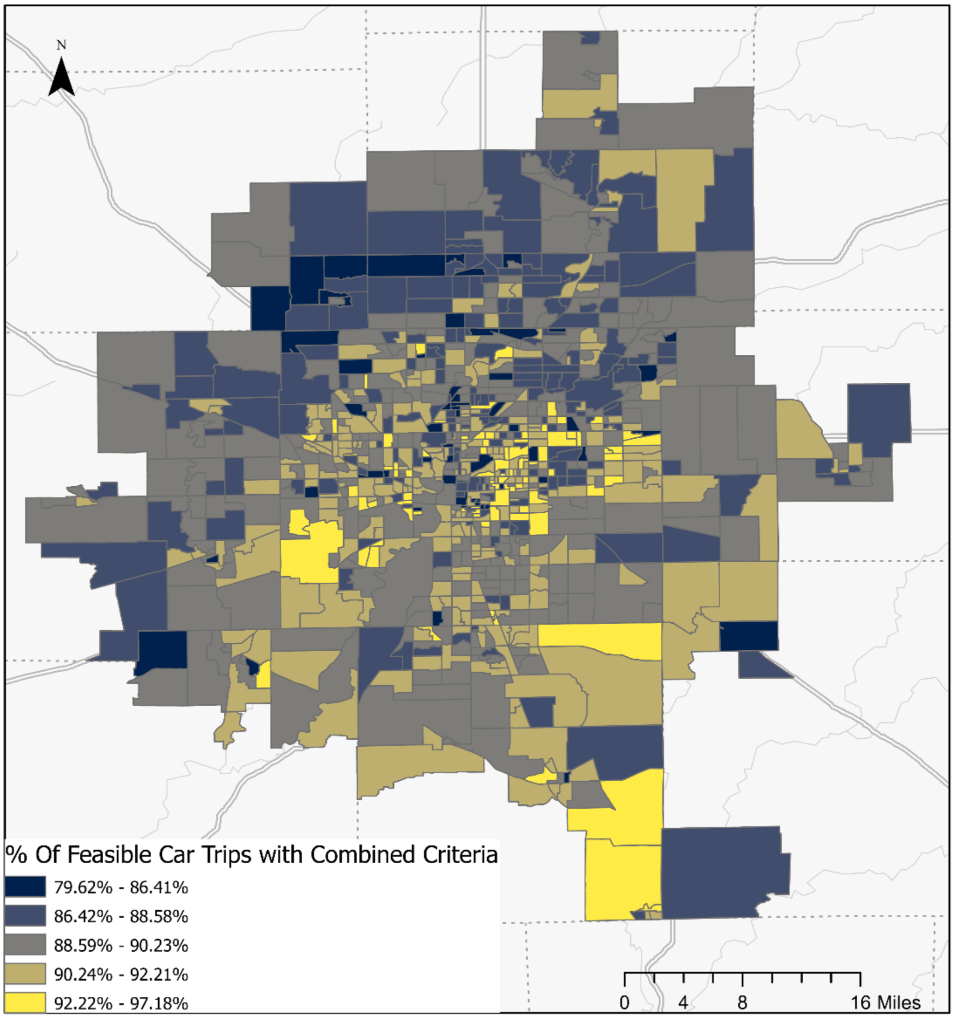

Figure 5-13. Percentage of feasible trips with car based on the combined criteria in census block group level.

For ride hailing trips there are census block groups around the suburban areas north of Indianapolis with very low percentage of trips as feasible as shown on Figure 5-14. This indicates that the travel demand from these census block groups is mainly for shorter and medium distance trips since we found that ride hailing trips become more feasible as the trips get longer with cost criteria (Figure 5-10). It is worth noting that if we vary the wait time for ride hailing trips depending on the location the areas with high percentage of feasible ride hailing trips in suburban regions likely to see a decrease in these percentages since the wait time will be substantially larger for these areas compared with the urban core. Therefore, we can't say for certain that these areas have higher ride hailing access compared with the areas in the urban core.

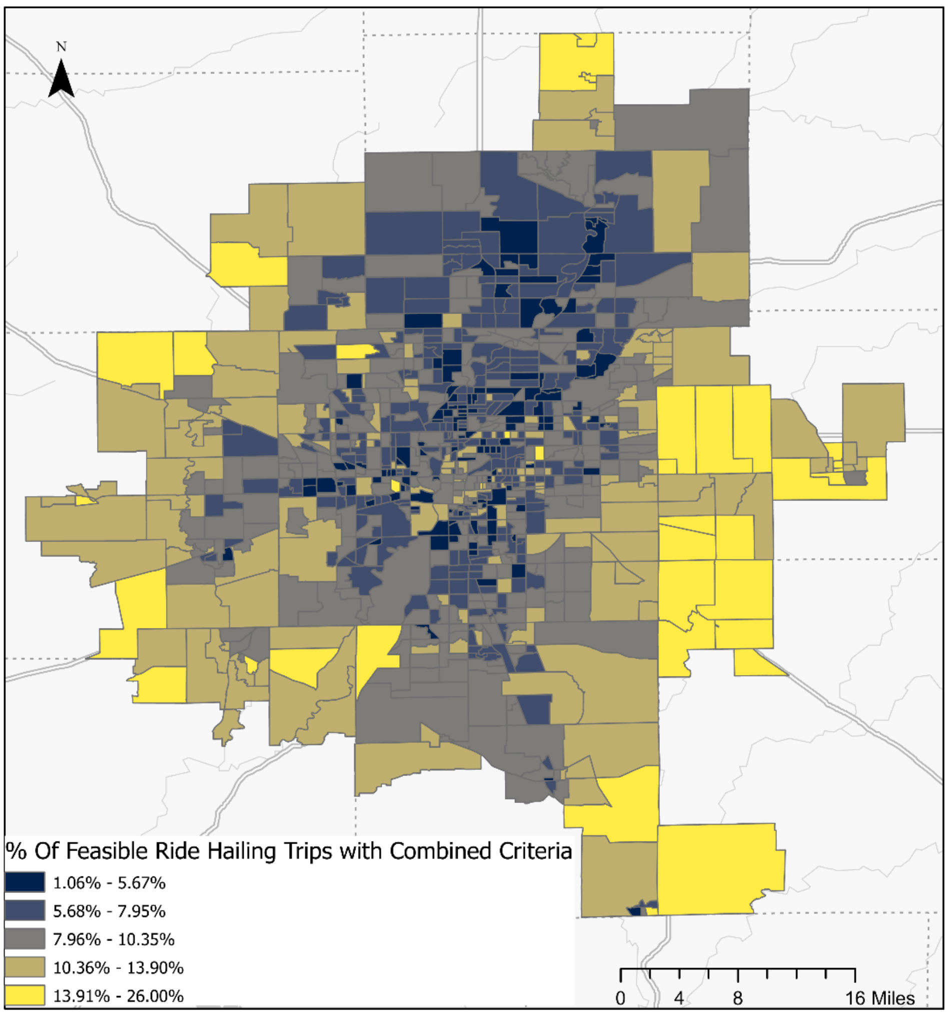

Figure 5-14. Percentage of feasible trips with ride hailing based on the combined criteria in census block group level

When we look at the percentage of feasible biking trips with combined criteria, we see stark differences between Hamilton and Marion County and the other counties. Since bike trips are feasible with cost function compared with quality criteria more (see Figure 5-10), the areas with a high percentage of feasible bike trips are more likely to have bike friendly infrastructure. Affluent suburban areas such as Carmel and Fishers which are marked with a red square on the Figure 5-15 is a good example of affluent areas with bike friendly infrastructure.

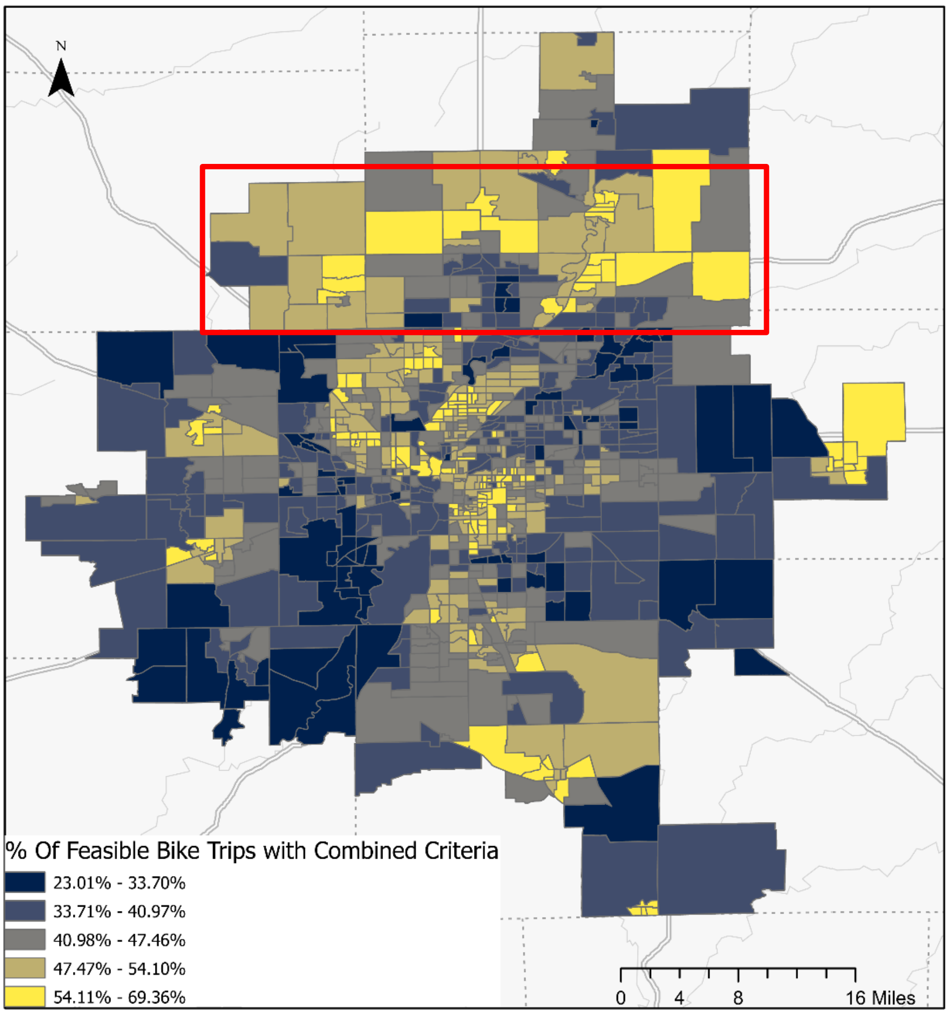

Figure 5-15. Percentage of feasible trips with bike based on the combined criteria in census block group level (red square shows the affluent suburban areas: Carmel and Fishers)

The bike sharing is not surprisingly is only feasible where the stations are located but interestingly, more than 40% of the trips were identified as feasible with bike sharing in downtown Indianapolis in Figure 5-16. In this figure percentage of feasible options with quality criteria is shown because as seen in Figure 5-10 there are almost no feasible bike sharing trips with combined criteria method due to high cost of bike sharing (0.04%). This is also not only because the average trip distance being low since we can see that same census blocks with high percentage of feasible bike sharing trips (yellow areas) in Figure 5-16 are also shown as high-low outlier in Figure 5-4. This means that these census block groups have higher trip distances compared with the neighboring census block groups. This indicates that bike sharing can be a feasible alternative for more people if the spatial coverage is expanded and the cost is adjusted to be more comparable with car similar to public transit.

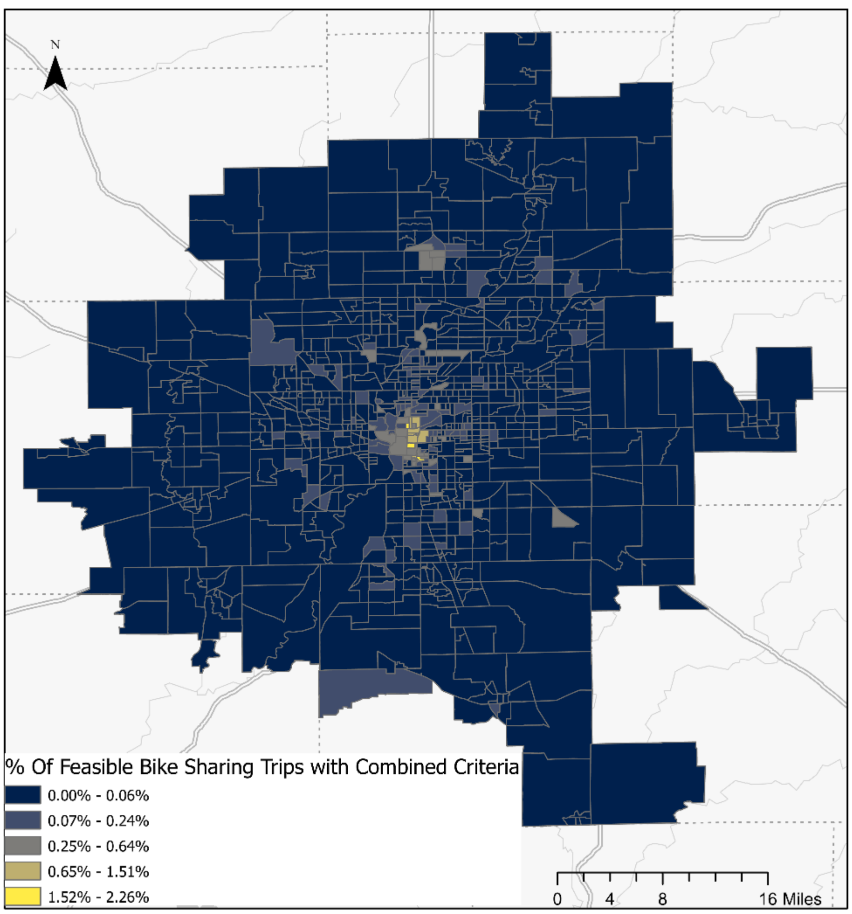

Figure 5-16. Percentage of feasible trips with bike sharing based on the combined criteria in census block group level.

For walking trips, it is important to note that sidewalks outside of the Marion country were all assumed to be in a good condition which could explain the higher percentage of trips being feasible outside of the Marion County in Figure 5-17. Even with this assumption since only a small percentage of trips are feasible with cost function the spatial differences are harder to distinguish. One inference that can be made is the downtown area in urban core has better walking infrastructure compared with the outer areas (not the suburban areas). This difference is another indication of the importance of the infrastructure characteristics when determining the travel-demand-relevant access.

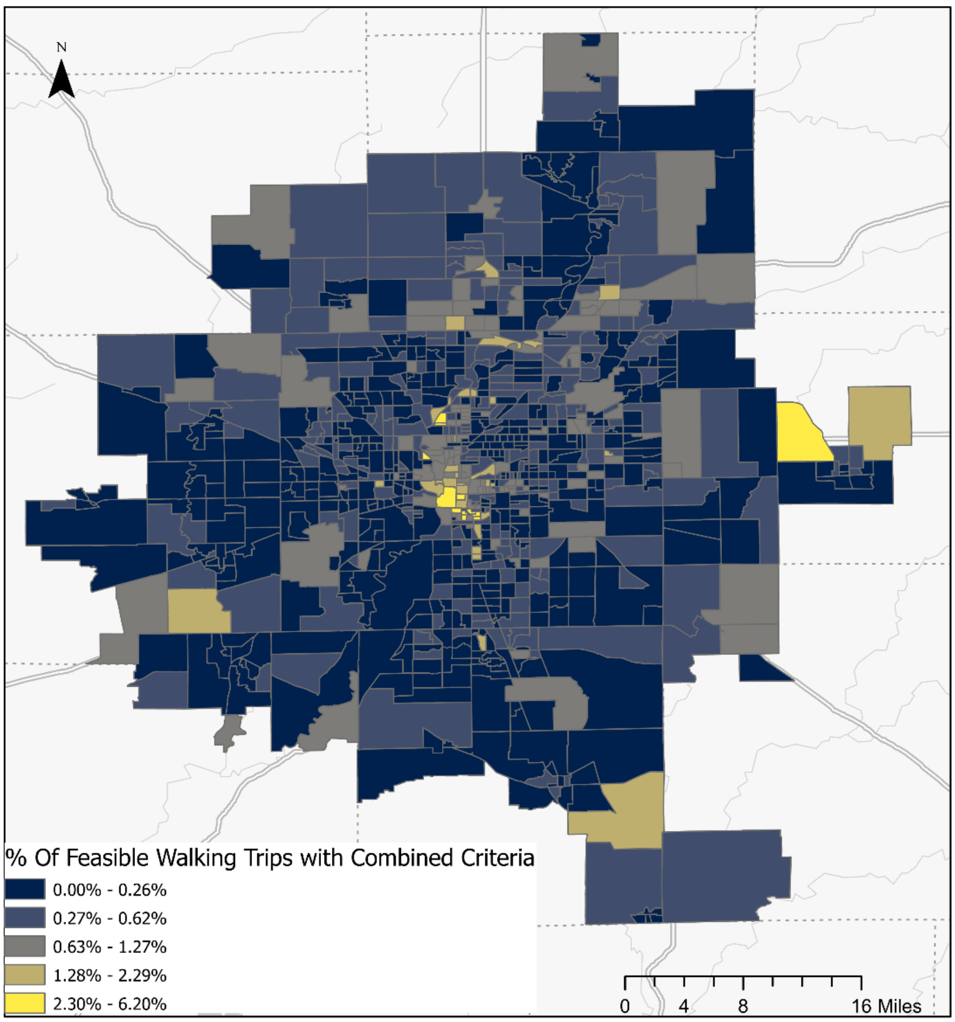

Figure 5-17. Percentage of feasible trips with walking based on the combined criteria in census block group level.

For public transit with walking and public transit with biking, we see that the Southwest Indianapolis have less feasible trips compared with other places with access. The main reason for this is not having a proper public transit route to satisfy the travel demand because we do not see any stark differences in terms of feasibility of other mobility options. In addition, Figure 5-18 and 5-19 shows that using bike with public transit increases the spatial coverage but travel-demand-relevant access in the Southwest Indianapolis is not satisfied once again showing the importance of assessing access regarding travel demand.

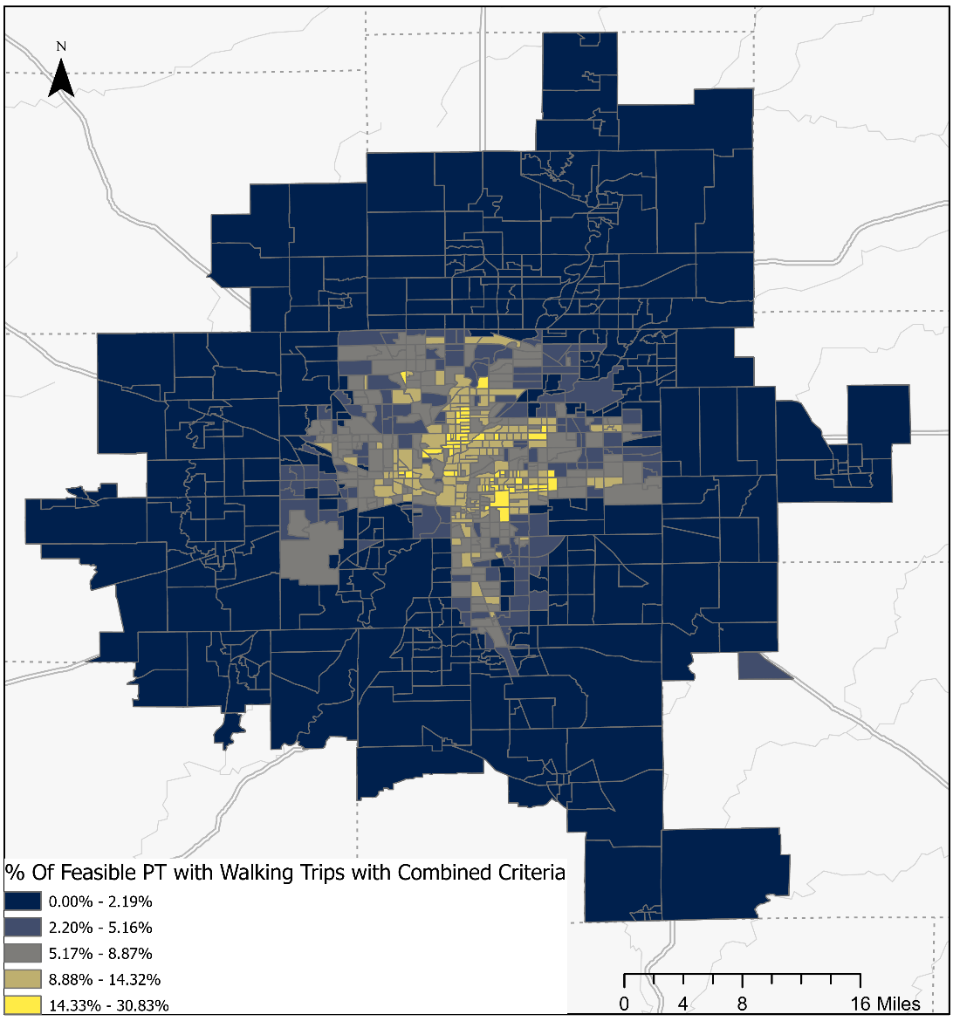

Figure 5-18. Percentage of feasible trips with public transit and walking as first & last mile based on the combined criteria in census block group level.

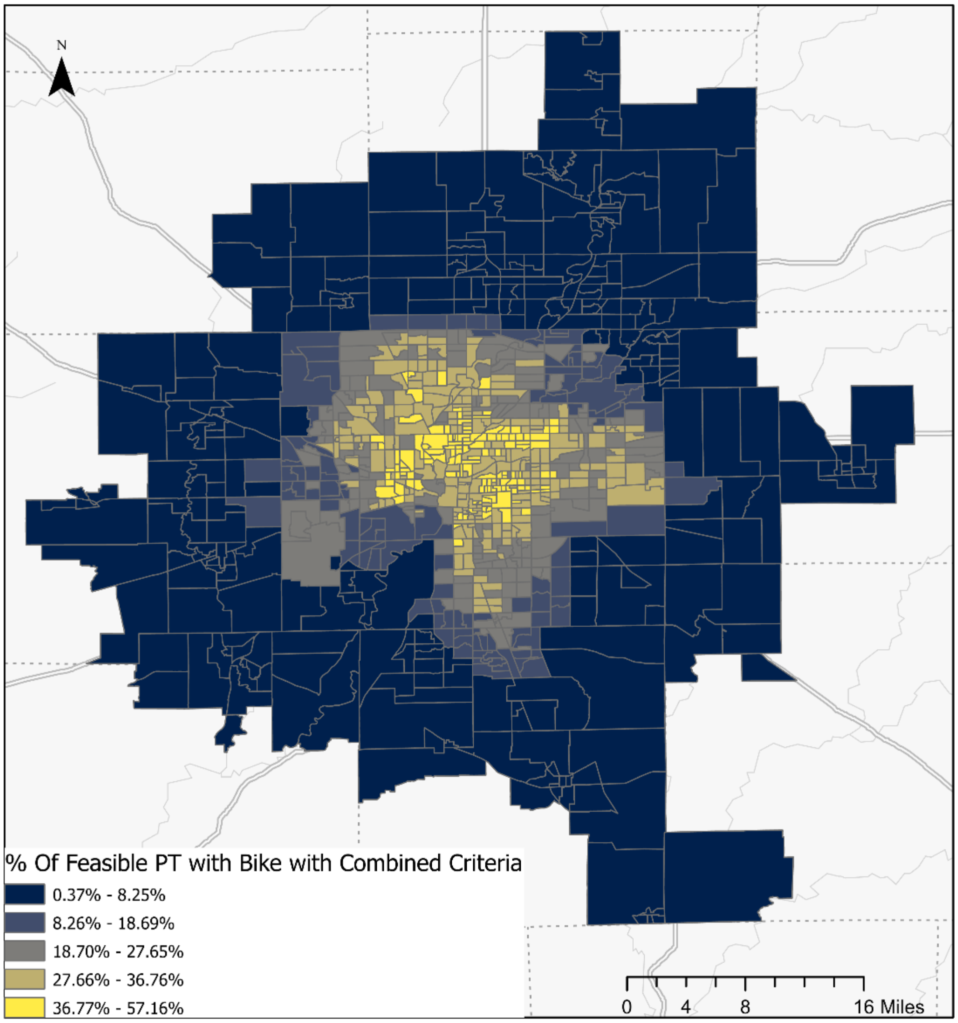

Figure 5-19. Percentage of feasible trips with public transit and biking as first & last mile based on the combined criteria in census block group level.

In summary, this thesis found that cost-based methods tend to favor faster trip modes such as car since the benefits from walking or similar active modes are harder to represent on trip level. Moreover, high initial costs of using certain modes (BSS, ride hailing) makes these modes incomparable to using car especially for shorter trips. Because of this when evaluating feasibility of certain trip modes quality criteria needs to be considered accounting for people who might prefer to use these costlier trip modes (for example for environmental reasons). Assessing the combination of both can give a more comprehensive idea of what are the available mobility option for people living in a certain area to use. Additionally, comparing the travel-demand-relevant access and traditional accessibility-based measures indicated that large percentage of trips that might assumed to be feasible with public transit with traditional accessibility-based studies are not feasible when disaggregated travel demand is incorporated. Lastly, the emerging mobility options such as bike sharing can satisfy almost half of the travel demand of people living near those stations with ease and comfort. Increasing the spatial coverage of these modes and decreasing the cost of using them can help increase the feasible mobility options available.

5.3 Equity Analysis

This chapter presents the results from the equity analysis methods and identifies disadvantaged census block groups (spatial equity), sociodemographic characteristics that have lower number of mobility options, and comparison of number of mobility options between pre-determined disadvantaged groups and the others.

5.3.1 Spatial Mobility Option Equity

First, to identify the differences between census block groups in terms of average number of feasible mobility options we created the maps depicted in Figure 5-20. As it can be seen from the figure using the cost function method, almost exclusively all suburban areas are identified as having a low number of mobility options. While quality criteria identified some areas as having more options there are only sporadic which means that it might be due to the outliers in these areas. As it can be seen from Figure 5-21, there are several regions very close to the downtown with a low number of mobility options. The transit lines almost overlap with all the census block groups with a high number of mobility options on Eastern, Western and Southeast Indianapolis, indicating these areas benefit from the public transit availability the most. However, Southwest Indianapolis has significantly fewer options compared with other regions in Marion County which could mean that individuals living in these areas have “access” to public transit in a traditional sense however, their travel demands can’t be met because the bus is not going to where individual aims to go. This is contradicting with the findings from (Griffin & Sener, 2016) which found that Indianapolis do not have spatial omission in terms of public transit. Figure 5-21 also shows bike sharing stations help with increasing the number of mobility options for the census block groups that have access to these stations. Key takeaway from this is that the presence of these alternative mobility options can help reduce the increased car ownership and prevent transport deserts.

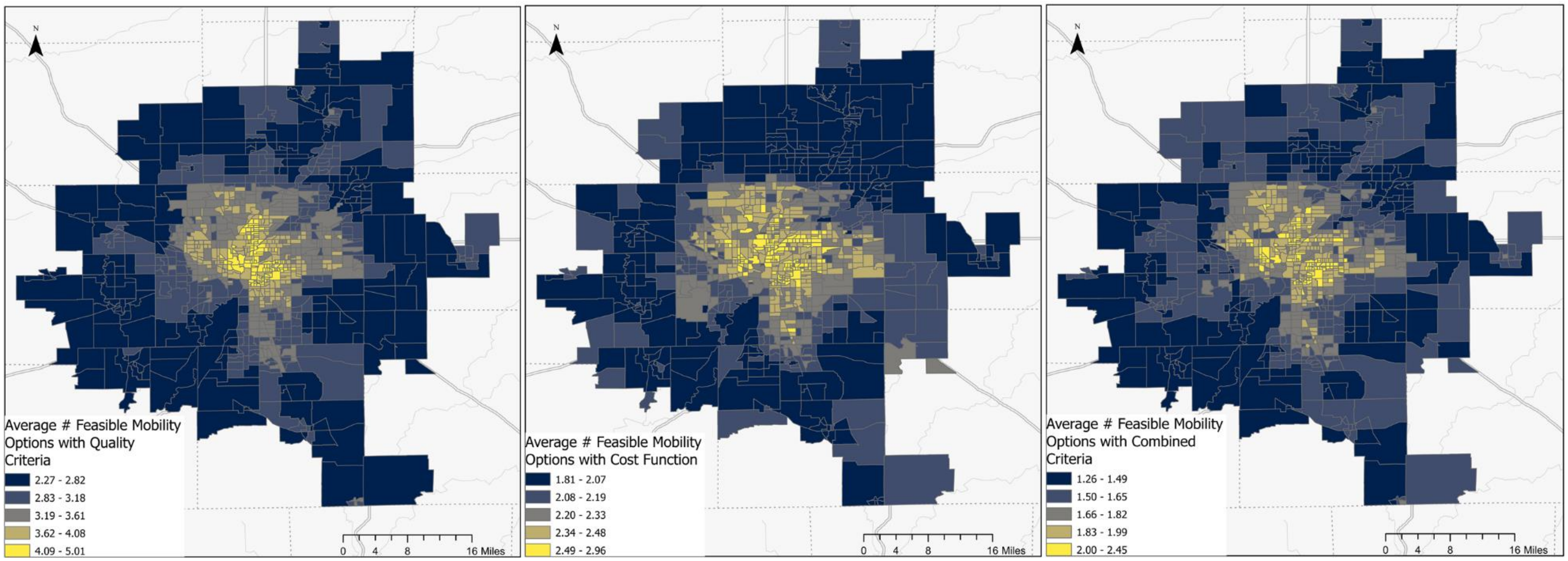

Figure 5-20. Average number of feasible mobility options based on quality criteria, cost function, and combined criteria in census block group level.

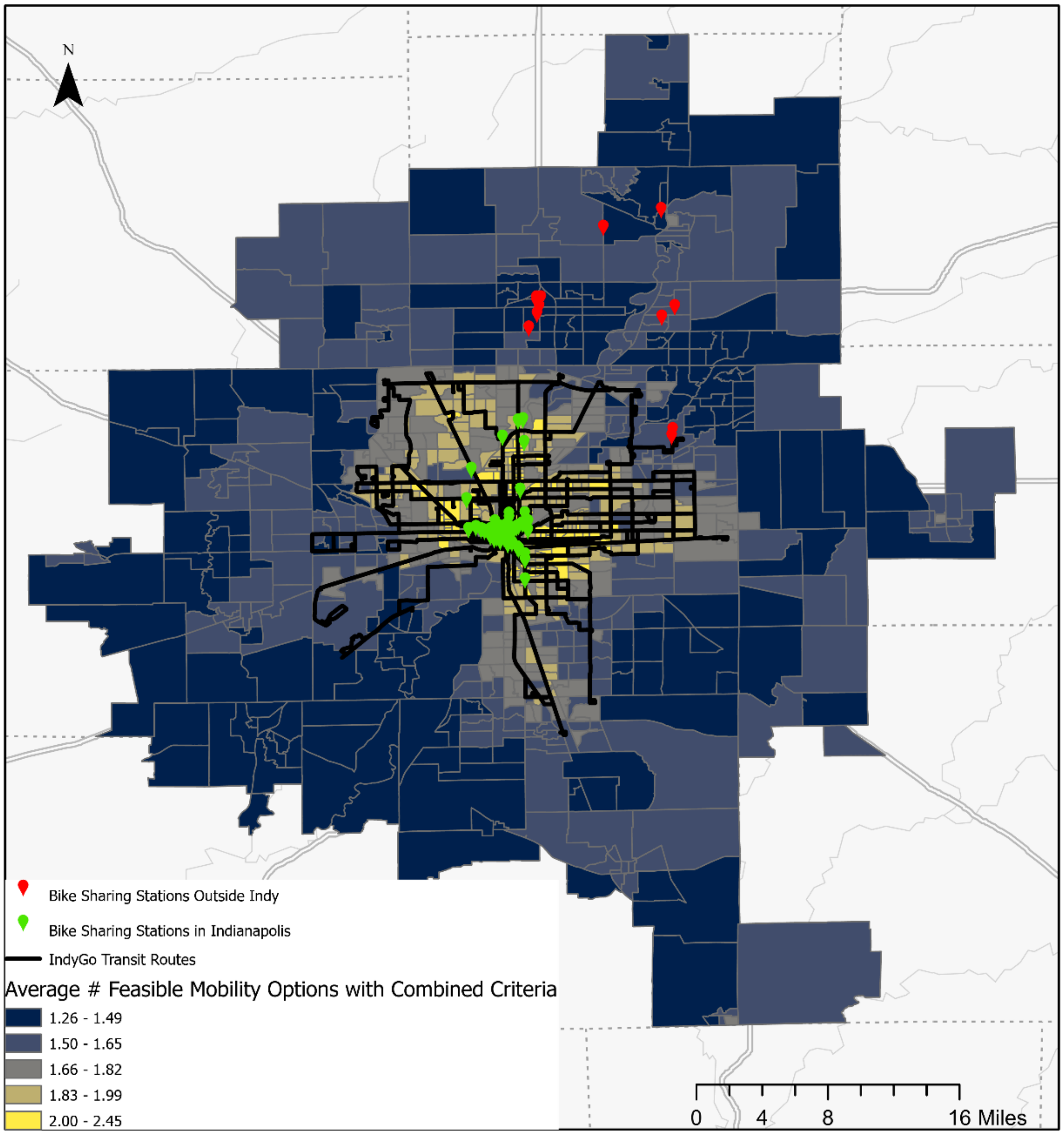

Figure 5-21. Average number of mobility options for residents in a census block based on the combined criteria, overlaid with public transit lines and bike sharing stations.

To check the clusters and outliers we created the Anselin Local Moran's I map for the combined criteria method which is represented in Figure 5-22. The Global Moran's I statistics for each trip mode and the number of mobility options also showed that each trip mode feasibility and number of mobility options are significantly clustered, meaning spatial autocorrelation is present. Please see Appendix A.6 for the Global Moran's I values and their significance level table. Suburban areas (census block groups outside of Marion County) had 12% fewer feasible mobility option with combined criteria than the urban core census block groups.

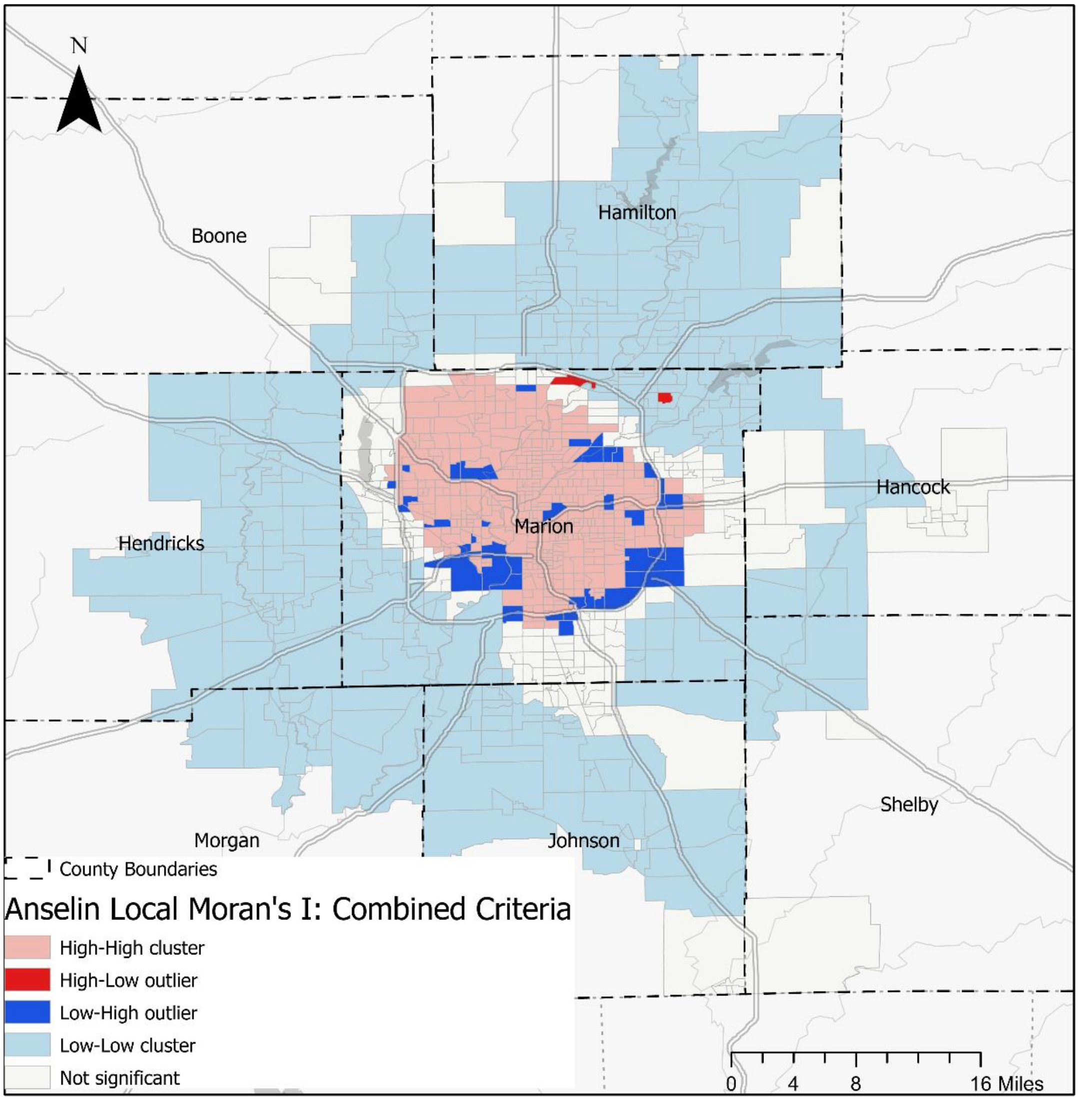

Figure 5-22. Anselin Local Moran's I cluster and outlier analysis with an average number of mobility options with combined criteria method

To understand the forced car ownership, this study mapped the census block groups with the lowest number of feasible options and census block groups with high vehicle ownership. Plotting the census block groups that are in the bottom 25% in terms of the number of mobility options with combined metric in Figure 5-23 showed that urban core areas are almost entirely excluded. Suburban areas on the other hand are almost exclusively comprised in the bottom 25%. This is also different from the findings from (Pyrialakou et al., 2016) which Northern Indianapolis and suburbs in that direction were not found to be clusters of high transport need. However, the areas in Morgan and Hancock Counties that are in bottom 25% were also identified as high need in the same study. Comparing the very high vehicle ownership areas (98% and above) and the census block groups in the first quantile in terms of feasible mobility options with combined criteria method shows that 64% of these census block groups with very high car ownership are also in the first quantile. The high car ownership in suburban areas (97% in suburban areas compared 89% in urban core) is an expected result because suburban areas by design are made to be automobile dependent as noted in Mattioli & Colleoni (2016). Based on this we can say that suburban areas are creating transport deserts where no alternative mobility options are available besides automobiles. This is a problem because except drive routes, other transportation infrastructure does not cover or benefit the suburban residential areas which also leads to automobile-oriented travel pattern of Indianapolis.

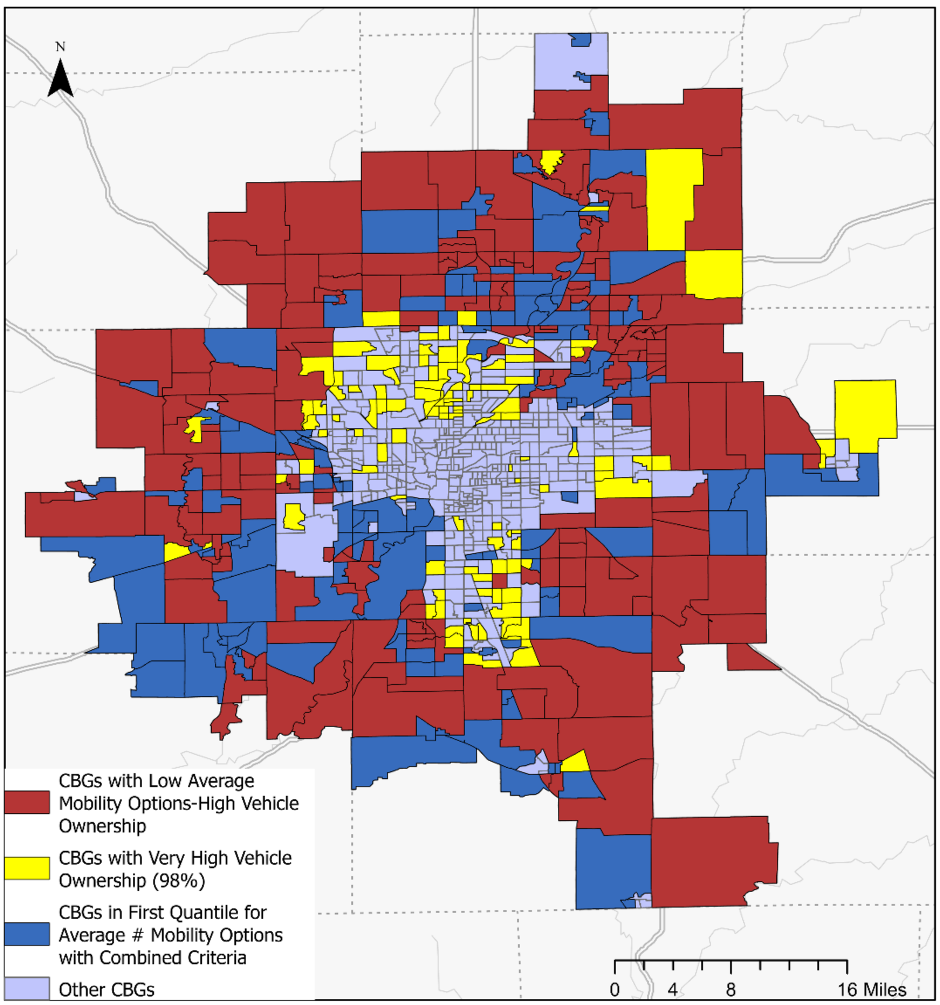

Figure 5-23. Census block groups in bottom 25% in terms of average number of mobility options available with combined criteria and census block groups with very high vehicle ownership

5.3.2 Sociodemographic Equity

To evaluate the equity in terms of sociodemographic characteristics, the results from the regression tree and linear regression model are presented. In Figure 5-24 the regression tree for sociodemographic characteristics mentioned in 4.2.3 evaluated with the combined number of mobility options. One way to interpret the regression trees shown in Figure 5-24 is to follow the left branch of each split. Following the leftmost branch in Figure 5-24 census block groups with high percentage of middle-class suburban white families have the lowest number of mobility options available for satisfying their travel demand, when using the combined metric to define feasibility. These characteristics are typical for the suburban areas in the empirical study area, implying that suburban areas have a low number of options. This however is likely not a problem considering high multiple vehicle ownership in these regions, the need for higher number of mobility options is minimal. (Please also see Appendix A.6 for both regression tree with quality criteria and cost function.)

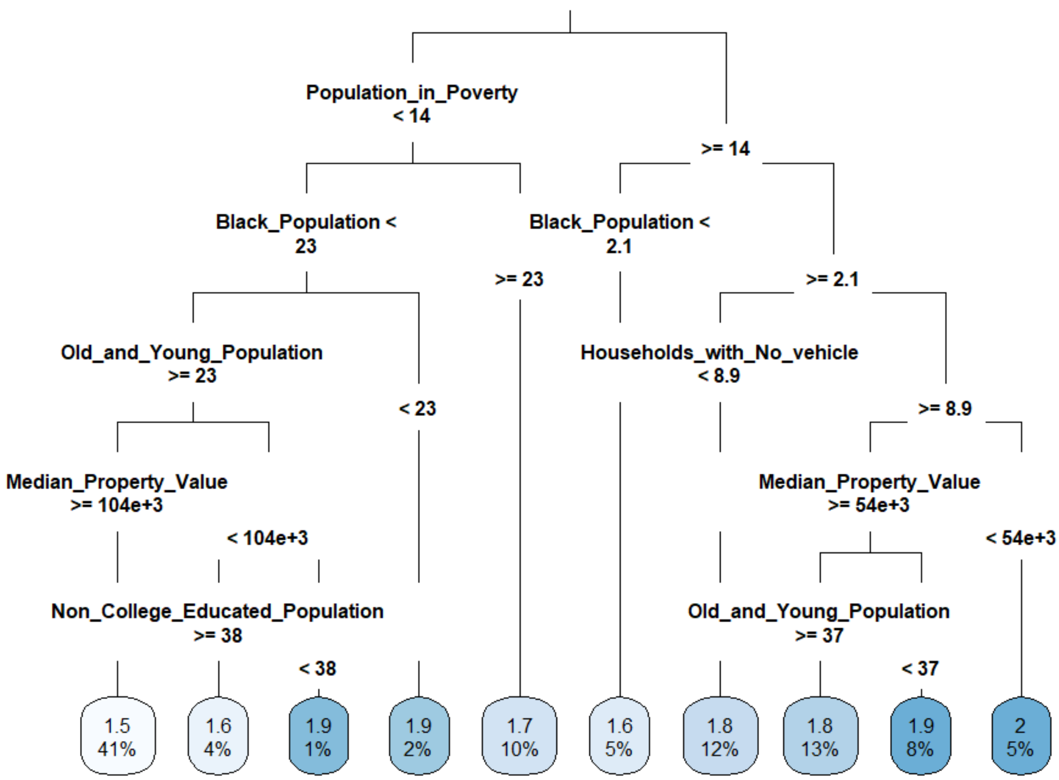

Figure 5-24 Regression tree with combined criteria mobility options and sociodemographic characteristic.

The importance ranks for each method which are listed in Table 5-2 below. The ranks in Table 5-2. indicate the importance of a variable to split the census block groups with low and high mobility options in each branch but it does not indicate if lower or higher values would be associated with lower or higher number of mobility options. These ranks allow us to distinguish which variables are more important with respect to certain method. Across all three methods economic variables are the highest ranked variables. Vehicle ownership and educational attainment follows these variables in terms of rank. For the cost function method, the percentage of black population is also more important compared to the other two methods.

Table 5-2. Variable importance rank for regression tree for each three methods (lower number indicates higher importance and higher the rank the earlier regression tree will split based on the criteria)

Variable	Combined criteria	Quality Criteria	Cost Function
% Of Population in Poverty	1	1	1
Median Income	2	2	2
% Of Households with No Vehicle	3	3	3
Median Property Value	4	4	5
% Of Non-College Educated Population	5	5	6
% Of People with Disability	6	6	7
% Of Black Population	7	7	4
% Of Young (0-19) and Old (65+) Population	8	8	8
% Of Hispanic or Latino Population	11	9	11
% Of Asian Population	9	10	9
% Of Population with No Schooling	10	11	10
% Of Native American Population			13
% Of Households with Room Occupancy > 1	12	-	12
% Of Hawaiian Population			14

One limitation with regression trees is that it is hard to distinguish if a certain sociodemographic characteristic has a negative or positive relationship with the number of mobility options. Therefore, multivariate multiple linear regression models with all three criteria were also build. After checking the correlation between variables, percentage of population with disability in poverty, percentage of population without disability in poverty, and percentage of

Non-White population variables were deleted due to high correlation with other poverty and race related variables. Since there is also high correlation between the three methods used to assess feasible mobility options multivariate analysis of variance (manova) for all three models were checked before finalizing the models. Based on the type 2 manova test Native American population, Hawaiian population, population with disability, population with no schooling and median property value were found to be insignificant thus removed from the model. Anova test between the full and reduced model also indicated taking these variables out does not create a significant change. The results from the multivariate multiple linear models for all three methods are presented in Table 5-3 below (Please see Appendix A.6 for the correlation importance graph after removing the highly correlated variables and results of type 2 manova test with Pillai statistic). The results from Table 5-3 indicates that education, high room occupancy, old and young population, and median income have a negative relationship with the number of mobility options. The percentage of household without vehicle variable having a positive relationship shows proves that forced automobile ownership exists in the study area since the results indicate higher the vehicle ownership lower number of mobility options. Interestingly, percentage of people in poverty had a positive relationship with mobility options. This can be explained by the higher proportion of low-income households in urban core compared with the suburban areas, where public transit options are available. Additionally, race and ethnicity related variables had positive or no significant relationship meaning that these groups are not negatively impacted by having a smaller number of options. Income and percentage of old-young population had a negative relationship among all three methods which could be explained by the suburban-urban core differences once again.

Table 5-3. Multivariate multiple linear regression coefficients for mobility options methodologies and sociodemographic variables (MPA Boundary)

Variable	**Combined Criteria Coefficients** (p-value in parentheses)	**Quality Criteria Coefficients** (p-value in parentheses)	**Cost Function Coefficients** (p-value in parentheses)
Intercept	$1.77\ (2.16 \times 10^{-16})^{***}$	$3.65\ (2.16 \times 10^{-16})^{***}$	$2.33\ (2.16 \times 10^{-16})^{***}$
% Of Households with No Vehicle	$0.0035\ (8.09 \times 10^{-6})^{***}$	$0.0101\ (3.51 \times 10^{-7})^{***}$	$0.0027\ (4.06 \times 10^{-5})^{***}$
% Of Black Population	$0.0018\ (5.81 \times 10^{-12})^{***}$	$0.0032\ (1.14 \times 10^{-6})^{***}$	$0.0022\ (2.16 \times 10^{-16})^{***}$
% Of Hispanic or Latino Population	$0.0025\ (0.0001)^{***}$	$0.0063\ (0.0001)^{***}$	$0.0016\ (0.002)^{**}$
% Of Population in Poverty	$0.0034\ (5.28 \times 10^{-8})^{***}$	$0.014\ (2.16 \times 10^{-16})^{***}$	$0.0025\ (7.12 \times 10^{-7})^{***}$
% Of Non-College Educated Population	$-0.001\ (0.026)^{*}$	$-0.0063\ (6.6 \times 10^{-8})^{***}$	$-0.0007\ (0.08)$
Median Household Income (in thousands)	$-0.0014\ (1.1 \times 10^{-6})^{***}$	$-0.0031(2.17 \times 10^{-5})^{***}$	$-0.0012\ (7.11 \times 10^{-7})$
% Of Households with Room Occupancy > 1	$-0.0064\ (0.002)^{**}$	$-0.0016\ (0.002)^{**}$	$-0.0052\ (0.002)^{**}$
% Of Young (0-19) and Old (65+) Population	$-0.0025\ (6.49 \times 10^{-5})^{***}$	$-0.0078\ (6.3 \times 10^{-7})^{***}$	$-0.0026\ (4.56 \times 10^{-7})^{***}$
	Combined Criteria Method	**Quality Criteria**	**Cost Function**
R^2	0.412	0.428	0.460
Adjusted R^2	0.406	0.422	0.454
Durbin-Watson Score	2.09 (0.244)	2.033 (0.086)	2.107 (0.168)

(Significance level, ***: 0.001, **:0.01, *: 0.05)

To understand if these groups are negatively impacted when urban core by itself was considered (since urban core and suburban characteristics are different), we selected all the census block groups inside the Marion County (only county with public transit system incorporated in the study) and evaluated the multivariate multiple linear regression trends as shown in Table 5-4. For these models focusing on Marion County similar variable selection process was adopted with manova to select the significant variables. Except the race and ethnicity variables (Asian or Hispanic population) the relationship between sociodemographic variables and the number of mobility options did not change implying that suburban-urban core differences is the main reason for the discrepancies regarding feasible mobility options. For all 6 of the models, we have checked the Variance Inflation Factor (VIF) to see if any variables had above value of 5. While no variable had above 5 value as (Johnston et al., 2018) pointed in their paper, VIF values above 2.5 also indicates considerable collinearity which is why income could be the cause for concern for models with all counties. However, since Durbin-Watson scores p-values are not less than 0.05, autocorrelation is not a cause for concern. It is also important to note that lower values R^2 and adjusted R^2 for all 6 of the models imply that there are either non-linear relationships between sociodemographic variables and number of mobility options or there are variables that we did not consider.

Table 5-4. Generalized linear regression results for mobility options methodologies and sociodemographic variables (Marion County)

Variable	**Combined Criteria** (p-value in parentheses)	**Quality Criteria** (p-value in parentheses)	**Cost Function** (p-value in parentheses)
Intercept	$1.8\ (2.16 \times 10^{-16})^{***}$	$3.6\ (2.16 \times 10^{-16})^{***}$	$2.3\ (2.16 \times 10^{-16})^{***}$
% Of Households with No Vehicle	$0.0038\ (1{,}43 \times 10^{-5})^{***}$	$0.0101\ (2.23 \times 10^{-6})^{***}$	$0.003\ (1{,}04 \times 10^{-6})^{***}$
% Of Black Population	$0.0008\ (0.006)^{**}$	$0.0002\ (0.7)$	$0.0011\ (9.68 \times 10^{-7})^{***}$
% Of Asian Population	$-0.0011\ (0.4)$	$-0.0087\ (0.008)^{**}$	$-0.0017\ (0.1)$
% Of Population in Poverty	$0.0035\ (3.59 \times 10^{-7})^{***}$	$0.0141\ (3.56 \times 10^{-16})^{***}$	$0.0025\ (3.51 \times 10^{-6})^{***}$
% Of Non-College Educated Population	$-0.0004\ (0.5)$	$-0.0037\ (0.009)^{**}$	$-0.0002\ (0.6)$
% Of Households with Room Occupancy > 1	$-0.0051\ (0.02)^{*}$	$-0.0124\ (0.02)^{*}$	$-0.0046\ (0.008)^{**}$
% Of Young (0-19) and Old (65+) Population	$-0.0022\ (0.001)^{***}$	$-0.0074\ (4.05 \times 10^{-5})^{***}$	$-0.0021\ (0.0002)^{***}$
	Combined Criteria Method	**Quality Criteria**	**Cost Function**
R^2	0.236	0.284	0.274
Adjusted R^2	0.226	0.275	0.265
Durbin-Watson Score	2.13 (0.132)	2.1 (0.248)	2.05 (0.588)

(Significance level, ***: 0.001, **:0.01, *: 0.05)

The adjusted R^2 values for all 6 multivariate multiple linear regression and regression trees are shown below in Table 5-5 for comparison. These results indicate that regression trees are better in identifying the census block groups with a low number of mobility options. In conclusion, based on our different methods it is hard to say certain sociodemographic characteristics are creating disadvantages in terms number of mobility options available. Moreover, the innate differences between suburban and urban core areas also seem to be the reason for the sociodemographic variables we see as indicators for a low number of mobility options in MPA boundary.

Table 5-5. Adjusted R^2 comparison between linear regression and regression tree with all counties and Marion County

	Multiple Linear Regression Adjusted R^2	**Regression Tree Adjusted R^2**
Combined Criteria (All Counties)	0.406	0.496
Combined Criteria (Marion County)	0.226	0.428
Quality Criteria (All Counties)	0.422	0.562
Quality Criteria (Marion County)	0.275	0.467
Cost Function (All Counties)	0.454	0.541
Cost Function (Marion County)	0.265	0.433

To understand if the disadvantaged census block groups have on average has a smaller number of mobility options, this study also compared all census block groups that fit with at least three of the seven criteria with all other census block groups. Wilcox test for the overall number of mobility options available with all three methods (combined criteria, quality criteria, and cost function) between two groups revealed that there is a significant difference between groups. Disadvantaged block groups had a higher average and median than the other block groups. This is likely because heavier concentration of disadvantaged census block groups in urban core where transit and bike sharing options are available. This means that expanding the urban areas with suburban regions could create a challenge with providing feasible mobility options for these areas and exaggerate the forced car ownership.

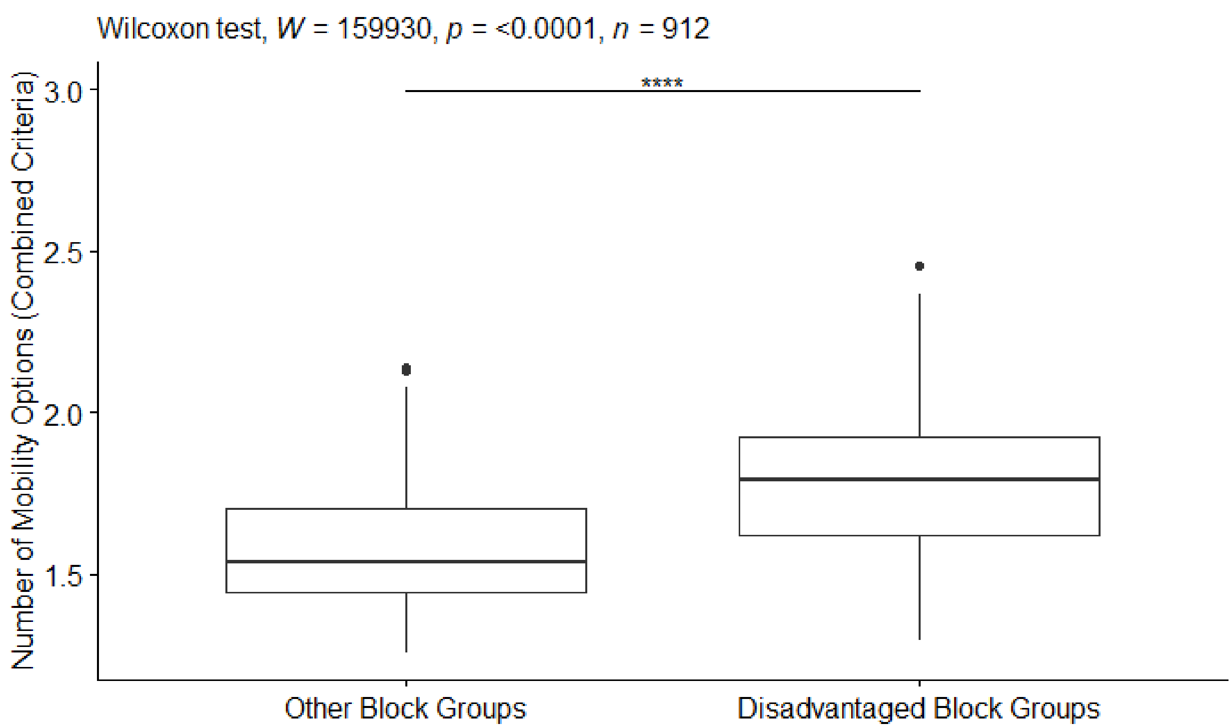

Figure 5-25. Box plot for disadvantaged block groups and other block groups with Wilcoxon test results

6. DISCUSSIONS AND CONCLUSION

In this thesis, we have proposed a new framework and travel-demand-relevant access metric for evaluating transportation equity using GPS data to construct individual travel demand and evaluate the number of feasible mobility options with respect to the quality of service and constraints from infrastructure. Our methods incorporated the individual heterogeneity in travel demand (collection of trips), considered built environment and system quality, and improved the traditional cost-based trip mode models by integrating better routing choice constraints with active travel modes. This thesis studied the mobility option equity with respect to the census block groups in Indianapolis Metropolitan Planning Organization's (MPO), Metropolitan Planning Area boundary in spatial and sociodemographic level. The key insights of this study can be listed as (1) it is important to consider travel-demand-relevant access to evaluate transportation equity because we found that both infrastructure constraints and travel demand to be substantially impactful; (2) suburban areas on average have 12% less mobility options available (by design) which forces high car ownership in these areas; (3) people with non-college educational attainment, households with more crowded rooms, and larger families are the negatively impacted disadvantaged groups while census block groups with high composition of white middle-class suburban families have the lowest number of options available; and (4) in general census block groups with several disadvantaged groups have a higher number of mobility options independent of the evaluation criteria, which indicates that transportation by itself can't improve the quality of life for these disadvantaged groups.

The results from this study can provide insightful policy suggestions to improve the transportation equity with current migration trends. If out-migration from urban core keeps the upward trend as we see during the pandemic (Frey, 2021) becomes the new trend for metropolitan areas, providing feasible and affordable transportation options for larger share of the population will increasingly become harder. This likely to cause even more dependence on automobiles while creating "transport deserts" where no alternative feasible trip modes are available. This increased dependence on automobiles are contradicting with the goal of reducing greenhouse gas emissions thus is a not desirable environmental outcome for the future. Consequently, continued subsidization and policymaking that makes suburbs an option are in stark contrast with equity and environmental goals. Furthermore, as the increasing trend in urbanization continues the disadvantaged groups will be forced into moving to outer urban areas due to rising urban core prices, gentrification, and lack of affordable housing. This will likely worsen the inequities for these communities since their access and opportunities are limited because of the transport deserts. Therefore, it is important to have a comprehensive understanding of the direct/indirect impacts of the existing and proposed transportation network elements and evaluate equity with respect to the observed travel demand with a combination of mobility and accessibility related measures. Additionally, understanding the limitations of the existing mobility options and improving them with relevant social justice philosophies rather than try to blindly achieve equality among all can benefit the groups with the most need. Ultimately it is important understand the role transportation plays in social and economic well-being and plan and act accordingly.

This study has several limitations. First, this study does not have the exact sociodemographic characteristics for each user because of the data limitation, thus we only could analyze equity in census block group level, assuming all users share the same mode choice preferences. It is possible that some people might not use certain trip modes due to bad maintenance and operation or simply due to preferences. Another important limitation of this study is that time cost was considered constant for everyone while it is likely to be more for people working in higher-wage jobs. Furthermore, since individual user characteristics can't be associated with individual trips, it is possible that some aggregation distorted the results toward certain direction. One potential improvement is to assign sociodemographic characteristics to everyone based on the distribution of the demographics at their identified home location area. While this model tries to incorporate preferences in identifying routes with certain travel modes (biking), preferences were based on characteristically different population and no heterogeneity between users was considered for mode choice nor routing. If relevant data is available, better cost/utility-based models can be developed with conjunction of quality of service and infrastructure constraint criteria. For trip modes that have a portion of the trip outside (walking, biking, bike sharing, and both public transit options) weather was also considered for quality criteria but since the weather forecast between August 21 and 31 in MPA did not have any rainy days, comparison regarding weather conditions were not included in quality criteria for those modes. Including weather as a feasibility criterion can change the overall picture especially for trip modes that large portions of the trip outside so future research can benefit from including weather condition constraints as well.

Moreover, this study missed some mode-specific attributes when quantifying the feasibility. For example, the health and other benefits from active modes such as walking, and biking were not included in cost function method. For car, parking cost and the time cost associated with walking from parking spot to final destination were also not included. Evaluating the parking cost and the walking to actual destination with out of vehicle time cost likely increase the percentage of feasible trips with cost function with other trip modes. For ride hailing trips wait time was taken as constant of 3.5 minutes but the areas with lower population density have likely higher wait times compared with the urban core areas. Last, this thesis did not identify the trip purpose and assumed all cost values to be associated with all-purpose trips. This can substantially change the results if trip purpose is also assigned based on the origin and destination locations. The cost function method likely won't be as significant for personal trips compared to commuting trips.

Some other minor limitation includes the trips with no feasible options based on the routing algorithm utilized in this study and the streetlight availability for nighttime walking. We considered streetlamp availability for safety and comfort for walking trips under dark but no publicly available data was found (OSM has a specific tag for streetlights but in the study area only a few of them marked). In the future, if the data is available this is something to be considered since similar to sidewalks, there has been even longer moratorium on street lights in Indianapolis (K. Dwyer & J. Ryan, personal communication, November 23, 2021).

Last, in terms of sociodemographic equity a more comprehensive definition of disadvantaged groups might be needed. In this study we mainly utilized the sociodemographic characteristics that were identified as disadvantaged in other transportation equity studies, but disadvantage based on the land use characteristics such as living close to highways can also provide important results regarding transportation equity regarding the home location of a user.

APPENDIX A. SUPPLEMENTAL INSTRUCTIONS

A.1 Empirical Setting

Since the MPA boundary shapefile was created in 2012 based on the 2010 census block boundaries, it does not completely match with census block group boundaries of respective 2017 American Community Survey (ACS) and 2020 Decennial Census datasets. Therefore, 'select by location' tool with 'have their center in' relationship in ArcGIS was utilized to select the census block groups that fall inside the MPA boundary. In Figure 3-1 two census block groups depicted in two colors: blue and orange. The blue block group was considered inside the boundary due to having its geographical center inside it even though most of its area was lying outside of the MPA boundary. The orange block group was not included in the boundary due to its geographical center being located outside the boundary even though it has a relatively large area that falls inside the boundary.

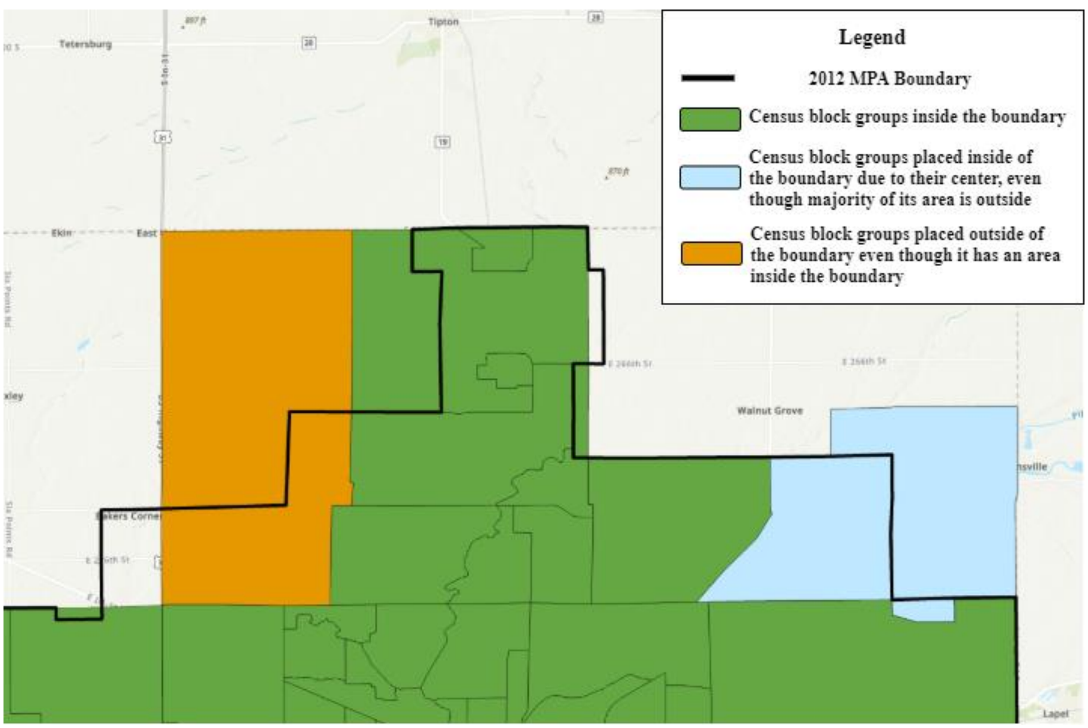

Figure A-1. Example selection operation for census block group inside the Metropolitan Planning Agency Boundary

A.2 GPS Trajectory Data Processing

Home Location Identification

When identifying the home location all GPS points from a user was evaluated based on the pre-assigned 7-character geohash value. In figure below top left picture shows all the data points of a user during nighttime (7 PM–7 AM). Bottom left picture is a zoomed in section of the larger map. The exact location of this boundary was left ambiguous to prevent identification of exact location of the home location. Bottom right picture shows how 9-character geohash values look like on the map. The yellow rectangle was the most frequent for this user so that grid is assigned as this user's home location. On top right the size of these grids shown regarding typical house around Indianapolis. This house is only for reference and it does not represent the exact house at this location.

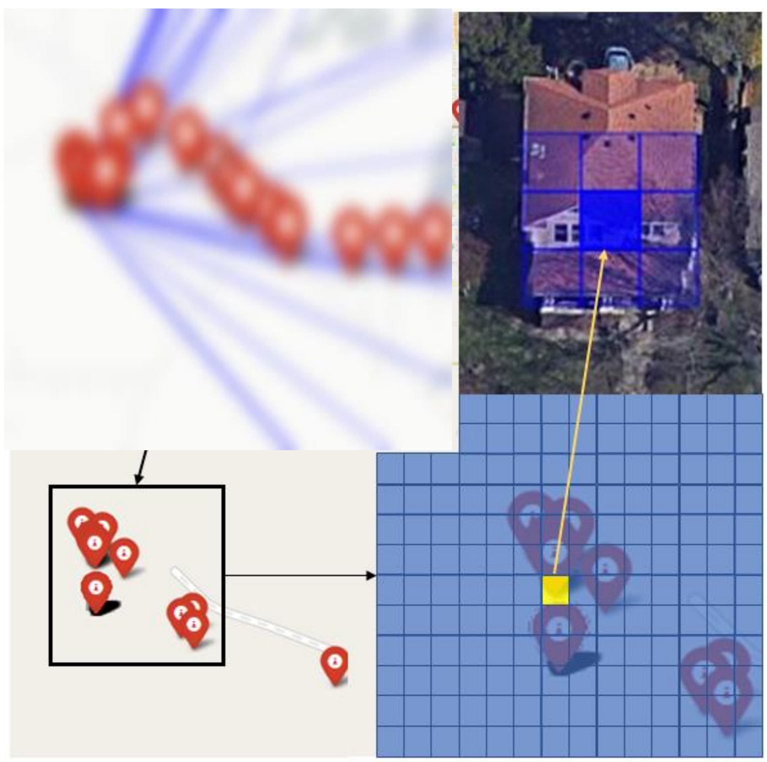

Figure A-2. Example home location identification process (map blurred for privacy protection)

A.2 Jump Point Identification

The figure below depicts one of the users GPS data points (red pins) sorted by timestamp. The consecutive points are connected with blue lines to show how the transitions between points look without any filtering. Two black boxes on the left figure show the two prevalent locations for this user's GPS points where a lot of sudden jumps were happening between trip trajectories. This characteristic for users were observed in all manually created user files which is why when filtering jump points we utilized the top two most frequent 7-character geohash values rather than just using the most frequent one. The exact 151m x 151m grid not shown to prevent identification of prevalent location for the user. The picture on the bottom right shows the pin locations on these likely jump points are spread out inside this grid. The top right picture size shows relative size of these grids compared to the large size buildings.

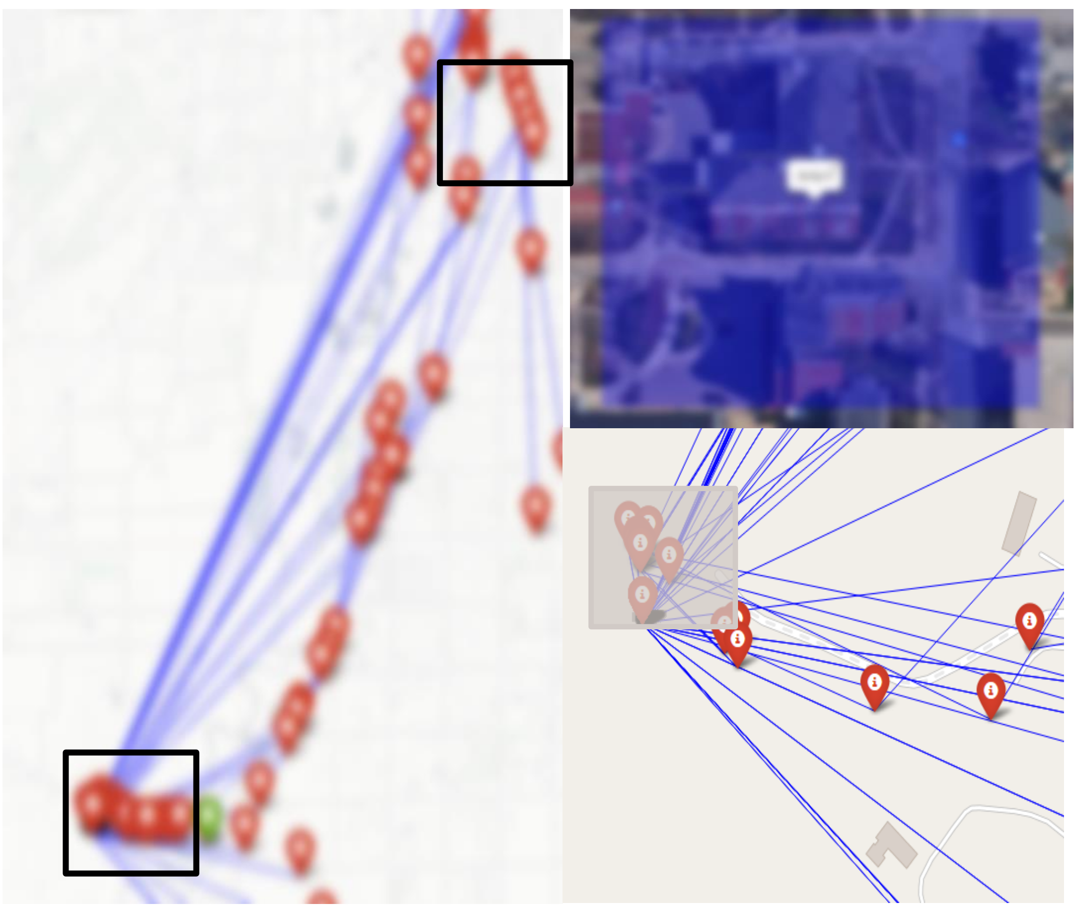

Figure A-3. Example for identifying jump points for users (map blurred for privacy protection)

Since SafeGraph reports this horizontal accuracy data for each GPS point, our first intuition was to use these values to determine GPS points that might have digressed from the actual path. The idea will use high horizontal accuracy values for finding out the potential GPS jumps that might have happened during a trip. To test this hypothesis, trip trajectories were manually created for 8 unique users by mapping the GPS points and time of the day a trip might have occurred to identify the GPS jump points. These points refer to the sudden gaps between two consecutive GPS points where the transition between the two points were unlikely to occur without an interference or error on the GPS. Jump points were then compared with the horizontal accuracy using logistic regression.

A.4 Mobility Options Comparability and Feasibility Evaluation

OSM Networks

When building the drivable, bikeable, and walking networks it is important understand which roads are allowed with each trip mode each tool (OSMNx and OTP).

OSMNx driving network drivable roads: Roads with *highway* tag equal to *motorway, motorway_link, trunk, trunk_link, primary, primary_link, secondary, secondary_link, tertiary, tertiary_link, unclassified, residential, living street, road, disused were* included in the driving network map of MPA boundary

OSMNx walking network walkable roads: Roads with *highway* tag equal to *trunk, trunk_link, primary, primary_link, secondary, secondary_link, tertiary, tertirary_link, unclassified, residential, service, footway, path, track, bridleway, living_street, steps, pedestrian, road, disused, crossing* were included in the walking network map of MPA boundary.

OSMNx biking network, bikeable roads: Roads with *highway* tag equal to *trunk, trunk_link, primary, primary_link, secondary, secondary_link, tertiary, tertiary_link, residential, service, cycleway, unclassified, path, track, bridleway, living_street, pedestrian, road, crossing, disused* were included in the walking network map of MPA boundary.

OPT driving network drivable roads: Roads with *highway* tag equal to *motorway, motorway_link, trunk, trunk_link, primary, primary_link, secondary, secondary_link, tertiary, tertiary_link, unclassified, residential, residential_link, service, track, byway, living street, road* were the road with open to car access.

Roads with open to pedestrians on OPT are roads with *highway* tag equal to *primary, primary_link, secondary, secondary_link, tertiary, tertiary_link, unclassified, residential, residential_link, service, track, byway, living street, road, bridleway, footway, pedestrian, path, cycleway, footway (footway = sidewalk, platform, crossing, steps, corridor, and public_transport = platform.*

Roads with open to pedestrians on OPT are roads with *highway* tag equal to *primary, primary_link, secondary, secondary_link, tertiary, tertiary_link, unclassified, residential, residential_link, service, track, byway, living street, road, bridleway, footway, pedestrian, path, cycleway, footway (*when *footway* tag is not equal to *sidewalk), platform, crossing, steps, corridor.*

Routing Configurations and Functions Used with OTP

Table A-1. Default and modified routing values for OTP.

	Default Values	Modified Values
Max transfers	12	2
Max walk distance	1500 m	1608 m (1 mile)
Max bike distance	1500 m	4023 (2.5 miles)
Walk speed	1.33 m/s	1.34 m/s
Max transfer distance between stations	2000 m	1608 m (1 mile)
Bike speed	4.9 m/s (11 mph)	5 m/s (11.2 mph)

A.5 Disadvantaged Census Block Groups

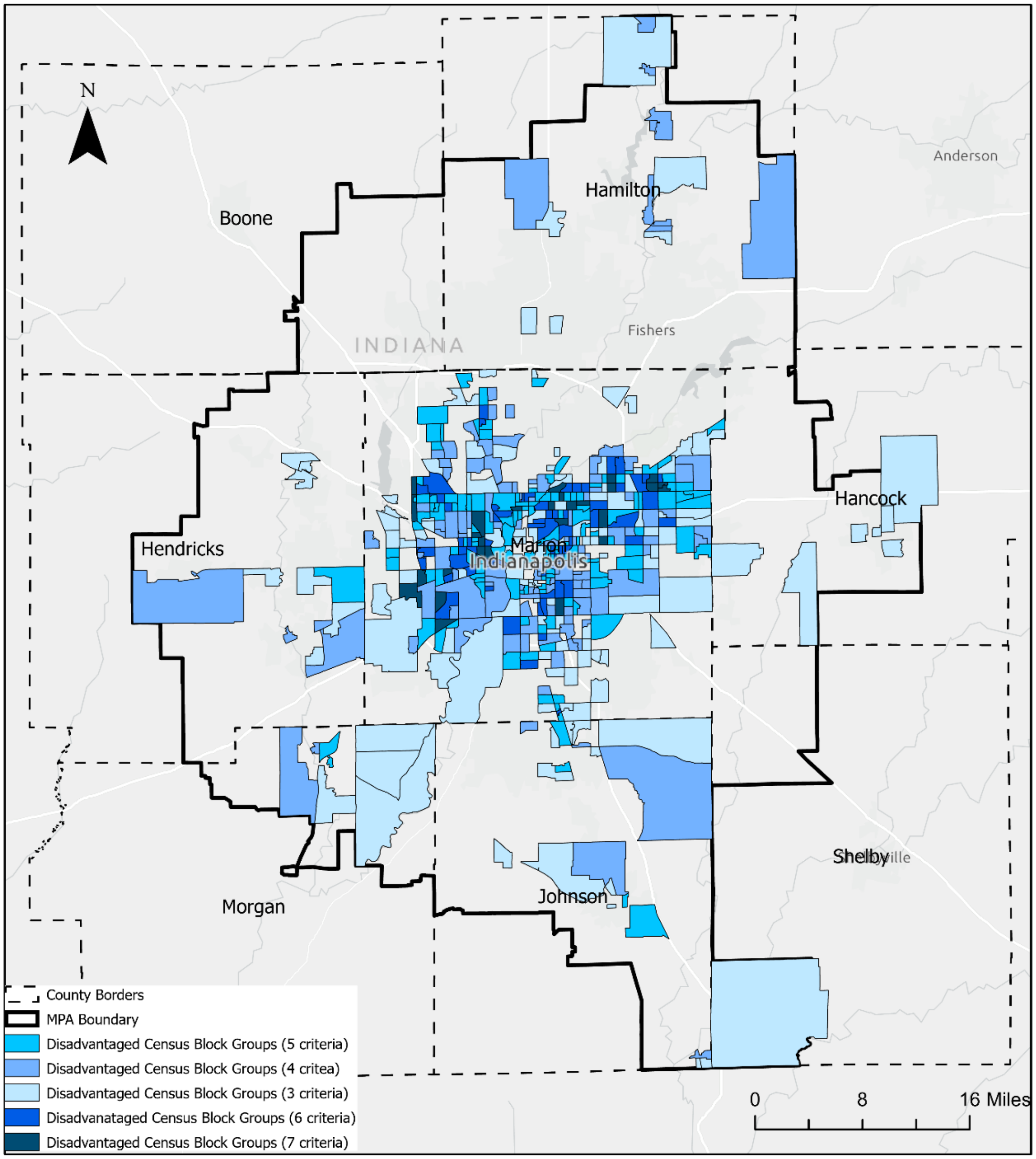

Figure A-4. Disadvantaged census block groups with sociodemographic criteria

A.6 Equity Analysis

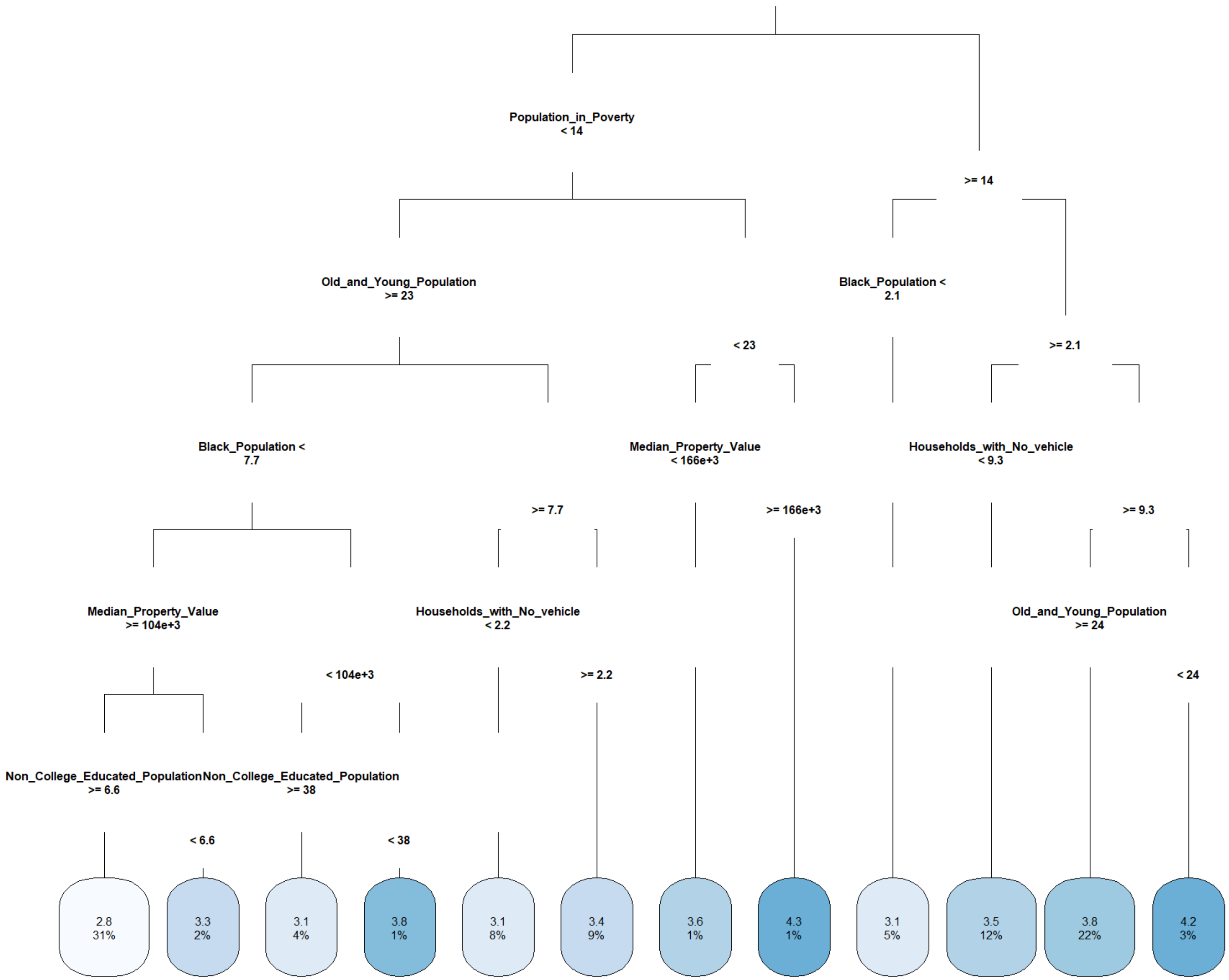

Figure A-5. Regression tree with quality criteria mobility options and sociodemographic characteristics

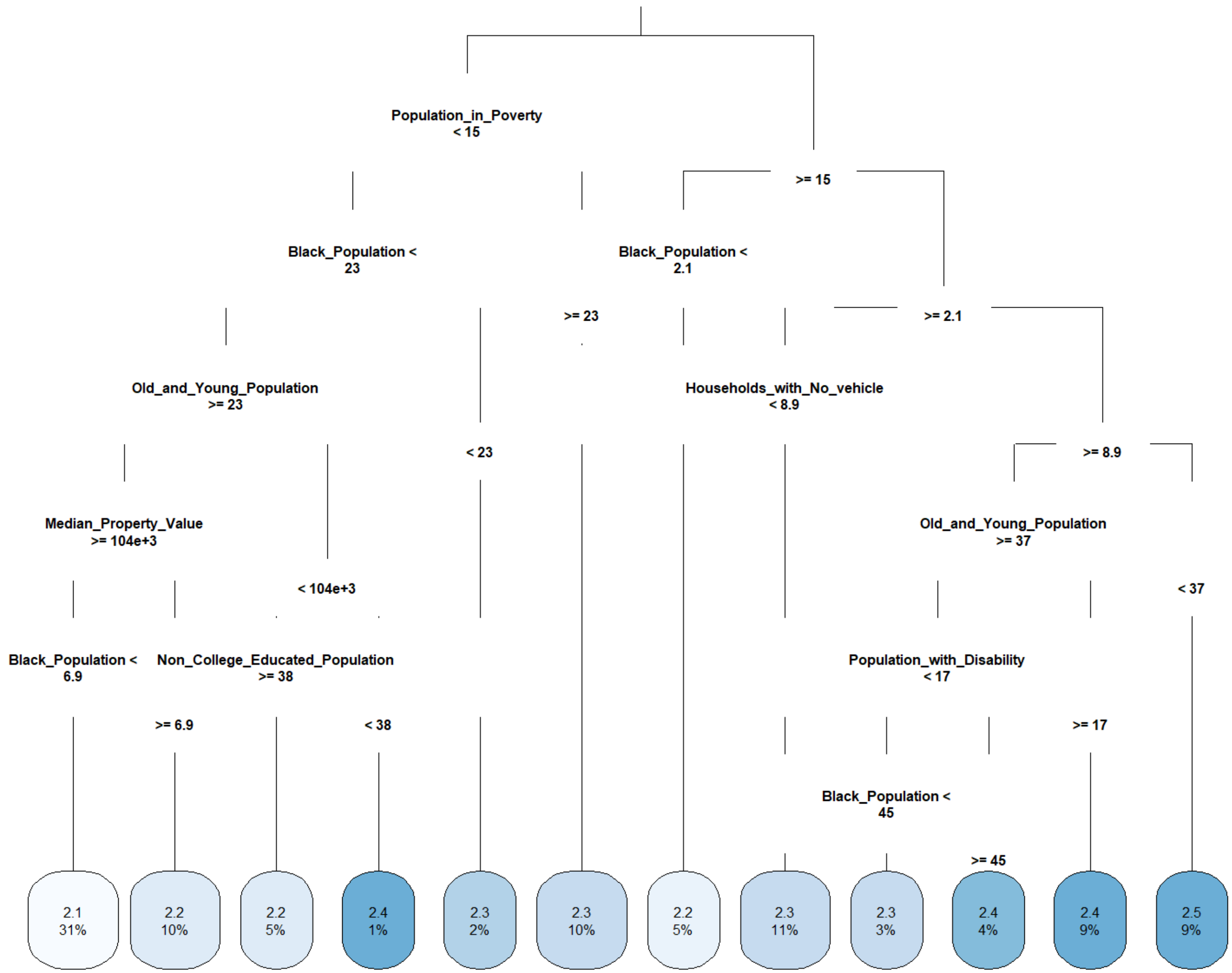

Figure A-6. Regression tree with cost function determined mobility options and sociodemographic characteristics

Table A-1. Global Moran I's Values (positive significant values indicate clusters)

	Quality Criteria	Cost Function	Combined criteria
Car	0.092***	0.146***	0.092 ***
Ride Hailing	0.093***	0.204***	0.235***
Bike	0.36***	0.369***	0.357***
Bike Sharing	0.468***	0.387***	0.39***
Walking	0.482***	0.436***	0.465***
Public Transit & Walk	0.717***	0.782***	0.728***
Public Transit & Bike	0.809***	0.799***	0.799***
Number of Options	0.66***	0.753***	0.689***

(Significance level, ***: 0.001, **:0.01, *: 0.05)

```
Type II MANOVA Tests: Pillai test statistic
                                    Df test stat approx F num Df den Df    Pr(>F)    
Households_with_No_vehicle           1  0.027442    8.418      3    895 1.606e-05 ***
Black_Population                     1  0.117694   39.796      3    895 < 2.2e-16 ***
Native_American_Population           1  0.001844    0.551      3    895  0.647457    
Asian_Population                     1  0.008702    2.619      3    895  0.049739 *  
Hawaiian_Population                  1  0.001622    0.485      3    895  0.692999    
Population_in_Poverty                1  0.091558   30.068      3    895 < 2.2e-16 ***
Population_with_Disability           1  0.008129    2.445      3    895  0.062663 .  
Population_with_No_Schooling         1  0.004189    1.255      3    895  0.288665    
Non_College_Educated_Population      1  0.035439   10.961      3    895 4.500e-07 ***
Hispanic_or_Latino_Population        1  0.016241    4.925      3    895  0.002128 ** 
Median_Income                        1  0.014488    4.386      3    895  0.004490 ** 
Households_with_High_Room_Occupancy  1  0.011973    3.615      3    895  0.012941 *  
Median_Property_Value                1  0.008600    2.588      3    895  0.051828 .  
Old_and_Young_Population             1  0.041669   12.972      3    895 2.673e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure A-7. Multivariate analysis of variance test results, insignificant p-value indicates those variables are not significant among all three models with three dependent variables (number of of feasible options with cost function, quality criteria, and combined criteria)

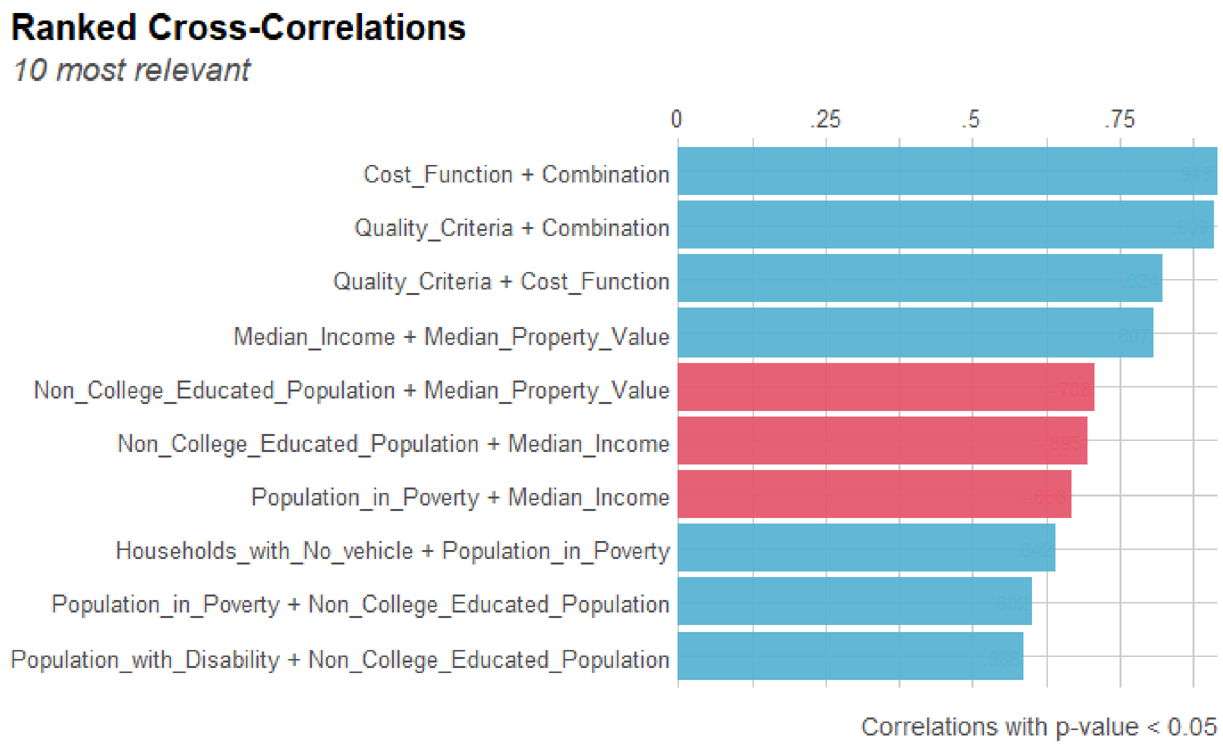

Figure A-8. Ranked cross-correlations graph for variables used in sociodemographic equity part

A.7 Addiional Maps

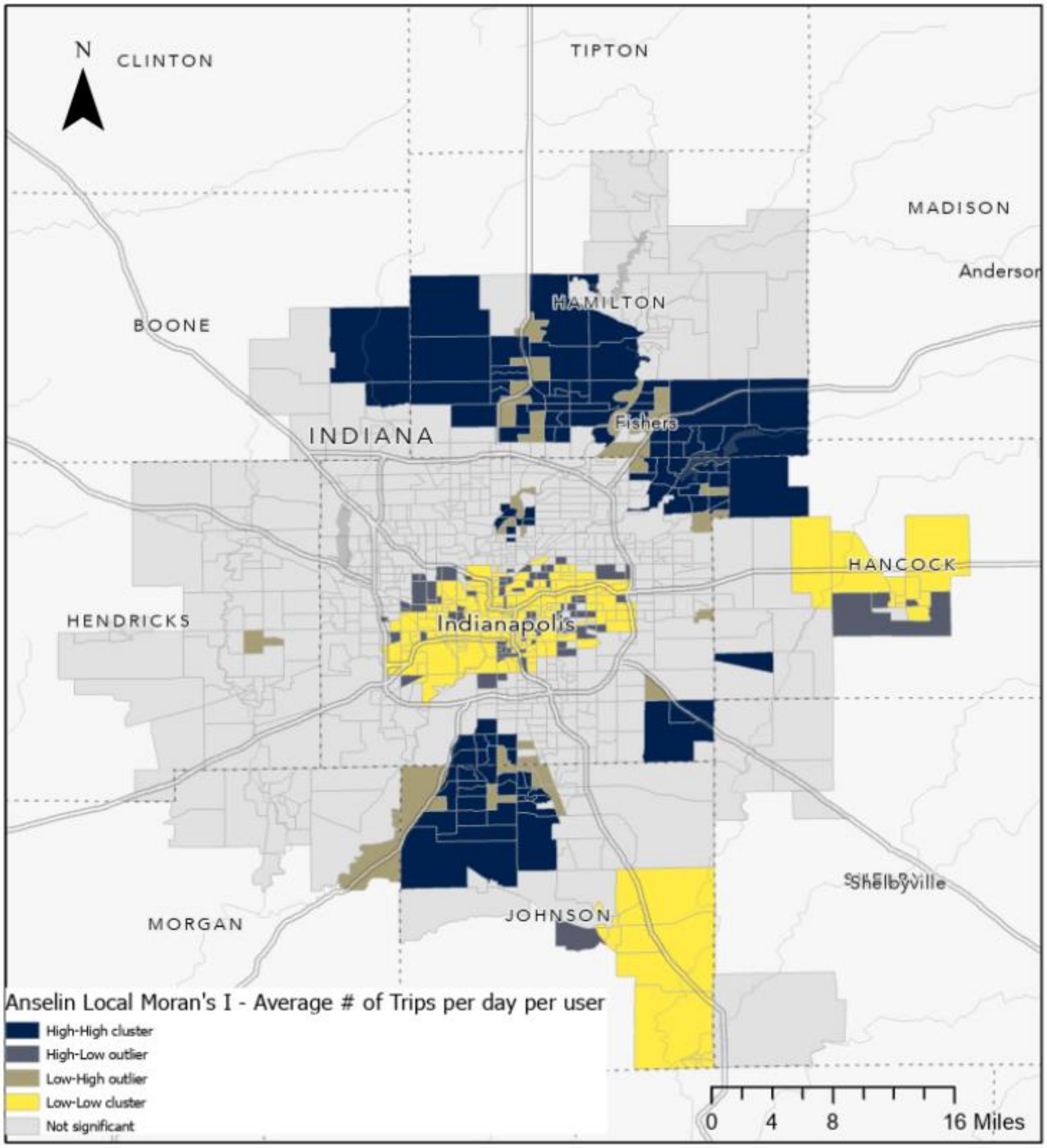

Figure A-9. Anselin Local Moran's I cluster and outlier analysis for average number of trips per day per user

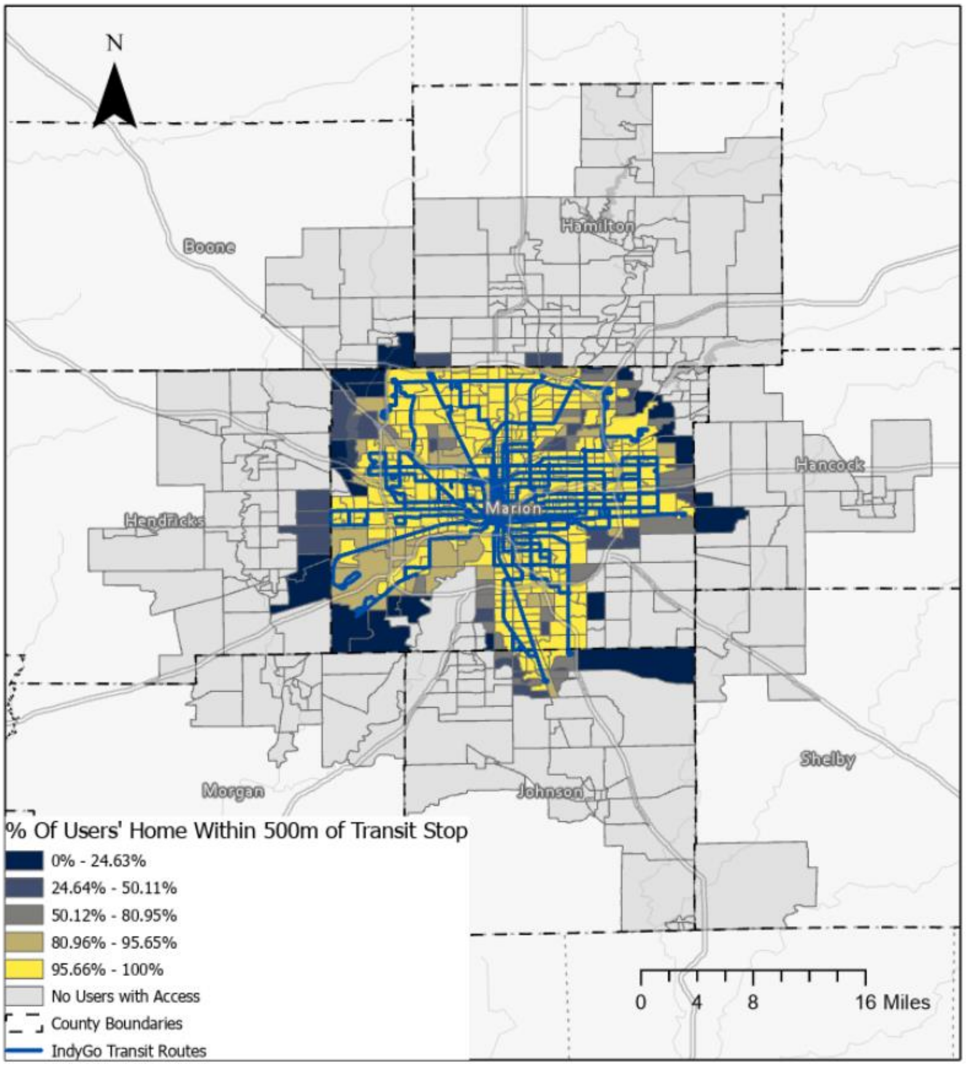

Figure A-10. Traditional proximity-based accessibility results

APPENDIX B. ADDITIONAL LITERATURE REVIEW

Table B.1. Literature review additional information.

Study	Inequity Measurement	Location	Socio-Demographic and Build Environment Characteristics Used / Location
(Shirmohammadli et al., 2016)	Horizontal Equity -spatial comparison (GINI Index) Vertical Equity - Focus on different groups)	Aachen, Germany	Age Employment Status Student status Vehicle ownership
(Griffin & Sener, 2016)	Horizontal Equity (Moran 1) Percent difference in transit service between low-wage and all transit accessible workers	Atlanta, GA/ Austin, TX / Dallas, TX/ Denver, CO / Houston, TX / Indianapolis, IN / Los Angeles, CA/ Seattle, WA/San Diego, CA	Income
(Karner, 2018)	Vertical equity—accessibility by income group	Phoenix metropolitan region, Arizona	Income
(Minocha et al., 2008)	Horizontal equity—spatial comparison (overlay analysis)	Chicago metropolitan area, Illinois	Vehicle ownership
(Welch & Mishra, 2013)	Horizontal equity (GINI index)	Washington–Baltimore region	-
(Chen et al., 2019)	Horizontal equity (GINI index) Vertical equity (comparison of descriptive statistics-ANOVA)	Tampa, Florida	Age Gender Income Race and Ethnicity

Table B.1.(continued) Literature review additional information.

(Mooney et al., 2019)	Spatial - sociodemographic equity: no specific mention but vertical equity is the most similar (Median and IQR analysis)	Seattle, Washington	Age Diversity index (race/ethnicity diversity) Education Income
(Smith et al., 2015)	Spatial and sociodemographic equity: no specific mention but vertical equity is the most similar (K-means test)	42 Bike Sharing Systems (BSS) across U.S., more detailed analysis with Chicago, Illinois	Economic hardship index (unemployment, dependency, education, income, race, percent occupied housing, and health insurance) Race and Ethnicity
(Meng & Brown, 2021)	Horizontal equity: spatial analysis (GINI index) (logistic regression and Tobit regression used for neighborhood characteristics)	32 U.S. cities that have both/either docked and dockless micromobility systems	Age Income Race and Ethnicity Vehicle Ownership Built environment variables: Employment density Population density Transit stop density
(Brown & Taylor, 2018)	Service equity across individuals: cancelation rate, wait times, (neither horizontal nor vertical equity descriptions describe the methodology very well but Vertical equity is more fitting)	Los Angeles, California	Age Income Race and Ethnicity Vehicle ownership Build environment variables: Employment density, road network density Number of transits stops Number of on and off-street parking Number of jobs in arts, food, and recreation Population density

Table B.1.continued Literature review additional information.

(Golub & Martens, 2014)	Spatial analysis by access poverty line (jobs accessible by transit users / jobs accessible by car)	San Francisco Bay Area, California	Income Minority status
(N. Chen & Wang, 2020)	Spatial and sociodemographic analysis (t-test for assessing the difference between quartiles 25^{th}, 50^{th} and 75^{th})	Fresno, California Cincinnati, Ohio	Age, Education, Income, Property Value, Race, Vehicle ownership
(Pyrialakou et al., 2016)	Vertical equity (Moran I used for spatial pattern identification and Needs gap assessment for transportation)	Indiana, U.S.	Age, Disabled status, Employment status, Income, Single parent/not, Vehicle ownership
(Braun et al., 2019)	Regression analysis between bike lanes and sociodemographic factors	23 U.S. cities	Composite socioeconomic index Education Income Race and Ethnicity

REFERENCES

ACS. (2018). *B25044 TENURE BY VEHICLES AVAILABLE*. American Community Survey (ACS). https://data.census.gov/cedsci/table?t=Transportation&g=0400000US18%241500000&tid=ACSDT5Y2019.B25044

Anselin, L. (2010). Local Indicators of Spatial Association-LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Executive Order No. 13985, 86 FR 7009 (2021). https://www.whitehouse.gov/briefing-room/presidential-actions/2021/06/25/executive-order-on-diversity-equity-inclusion-and-accessibility-in-the-federal-workforce/

Bike Score. (n.d.). *Bike Score Methdology*. https://www.walkscore.com/bike-score-methodology.shtml

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, *65*, 126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004

Bohannon, R. W., & Andrews, A. W. (2011). Normal walking speed: A descriptive meta-analysis. *Physiotherapy*, *97*(3), 182–189. https://doi.org/10.1016/J.PHYSIO.2010.12.004

Braun, L. M., Rodriguez, D. A., & Gordon-Larsen, P. (2019). Social (in)equity in access to cycling infrastructure: Cross-Chapteral associations between bike lanes and area-level sociodemographic characteristics in 22 large U.S. cities. *Journal of Transport Geography*, *80*. https://doi.org/10.1016/j.jtrangeo.2019.102544

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, *46*(10), 1730–1740. https://doi.org/10.1016/j.tra.2012.07.005

Brown, A. E., & Taylor, B. D. (2018). Ridehail Revolution: Ridehail Travel and Equity in Los Angeles [University of California, Los Angeles]. In *ProQuest Dissertations and Theses*. https://www.proquest.com/dissertations-theses/ridehail-revolution-travel-equity-los-angeles/docview/2056477010/se-2?accountid=13360 https://purdue.primo.exlibrisgroup.com/openurl/01PURDUE_PUWL/01PURDUE_PUWL:PURDUE? https://purdue-primo-prod.hosted.exlibrisgroup.com/openurl/PURDUE/purdue_services_page?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Global&atitle=&title=Ridehail+Revolution%3A+Ridehail+Travel+and+Equity+in+Los+Angeles&issn=&date=2018-01-01&volume=&issue=&spage=&au=Brown%2C+Anne+Elizabeth&isbn=978-0-438-01979-9&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Caltrans. (2022). *BENEFIT-COST ANALYSIS OF ACTIVE TRANSPORTATION PROJECTS (Cal-B/C AT)*. https://dot.ca.gov/-/media/dot-media/programs/transportation-planning/documents/data-analytics-services/transportation-economics/cal-bc/2022-cal-bc/guides/cal-bc-81-at-instructions-v1-a11y.pdf

CARB. (2018). *Low-Income Barriers Study, Part B: Overcoming Barriers to Clean Transportation Access for Low-Income Residents*. California Air Resources Board. https://ww2.arb.ca.gov/sites/default/files/2018-08/sb350_final_guidance_document_022118.pdf

Census. (2020a). *Indiana* [U.S. Census Bureau]. https://data.census.gov/cedsci/profile?g=0400000US18

Census. (2020b). *P1 Race* [U.S. Census Bureau].

Census. (2021). *Geographic Areas Reference Manual*. https://www.census.gov/programs-surveys/geography/guidance/geographic-areas-reference-manual.html

Chen, N., & Wang, C.-H. (2020). Does green transportation promote accessibility for equity in medium-size U.S. cites? *Transportation Research Part D: Transport and Environment, 84*, 102365. https://doi.org/10.1016/J.TRD.2020.102365

Chen, Z., Guo, Y., Stuart, A. L., Zhang, Y., & Li, X. (2019). Exploring the equity performance of bike-sharing systems with disaggregated data: A story of southern Tampa. *Transportation Research Part A: Policy and Practice*, *130*, 529–545. https://doi.org/10.1016/J.TRA.2019.09.048

Cohen, S., & Cabansagan, C. (2017). *A Framework for Equity in New Mobility*. TransForm.

Currie, G., & Senbergs, Z. (2007). Exploring forced car ownership in metropolitan Melbourne. *Social Research in Transport (SORT) Clearinghouse*.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269–271.

DOT. (2016). *The Value of Travel Time Savings: Departmental Guidance for Conducting Economic Evaluations Revision 2 (2016 Update)*. U.S. Department of Transportation. https://www.transportation.gov/sites/dot.gov/files/docs/2016%20Revised%20Value%20of%20Travel%20Time%20Guidance.pdf

DPW. (2018). *DPW Sidewalk Ratings*. https://data.indy.gov/datasets/dpw-sidewalk-ratings/explore?location=39.625674%2C-86.086631%2C9.94

Dwye) , K., & Ryan, J. (2021, November 23). *Indianapolis Star Transportation Reporter and Principal Consusltant at the History Concierge, LLC* [Personal communication].

eBikesHQ. (2020). *Over 450 Electric Bikes Compared! What does an ebike cost?* https://ebikeshq.com/cost-of-an-ebike/

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

Feng, T., & Timmermans, H. J. P. (2014). Extracting Activity-travel Diaries from GPS Data: Towards Integrated Semi-automatic Imputation. *Procedia Environmental Sciences*, *22*, 178–185. https://doi.org/10.1016/j.proenv.2014.11.018

Frey, W. (2021). Pandemic population change across metro America: Accelerated migration, less immigration, fewer births and more deaths. *Brookings Institution. See Https://Www. Brookings. Edu/Research/Pandemic-Population-Change-Acrossmetro-America-Accelerated-Migration-Less-Immigration-Fewer-Births-and-More-Deaths. Last Accessed*, *10*, 07–21.

Golub, A., & Martens, K. (2014). Using principles of justice to assess the modal equity of regional transportation plans. *Journal of Transport Geography*, *41*, 10–20. https://doi.org/10.1016/J.JTRANGEO.2014.07.014

Griffin, G., & Sener, I. (2016). Public Transit Equity Analysis at Metropolitan and Local Scales: A Focus on Nine Large Cities in the U.S.. *Journal of Public Transportation*, *19*(4). https://doi.org/10.5038/2375-0901.19.4.8

Guo, Y., Chen, Z., Stuart, A., Li, X., & Zhang, Y. (2020). A systematic overview of transportation equity in terms of accessibility, traffic emissions, and safety outcomes: From conventional to emerging technologies. *Transportation Research Interdisciplinary Perspectives*, *4*, 100091. https://doi.org/10.1016/j.trip.2020.100091

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). *Exploring Network Structure, Dynamics, and Function using NetworkX*. 5.

Hart, P., Nilsson, N., & Raphael, B. (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, *4*(2), 100–107. https://doi.org/10.1109/TSSC.1968.300136

Hillel, T., Bierlaire, M., Elshafie, M. Z. E. B., & Jin, Y. (2021). A systematic review of machine learning classification methodologies for modelling passenger mode choice. *Journal of Choice Modelling*, *38*, 100221. https://doi.org/10.1016/j.jocm.2020.100221

Indiana Pacers Bike Share. (2022). *Indiana Pacers Bikeshare Rates*. https://www.pacersbikeshare.org/top-nav-pages/join-now

IndyGo. (2021, August). *Fares & Passess*. https://www.indygo.net/fares-and-passes/

Ingram, D. D., & Franco, S. J. (2012). NCHS urban-rural classification scheme for counties. *Vital and Health Statistics. Series 2, Data Evaluation and Methods Research*, *154*, 1–65.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jensen, P., Rouquier, J.-B., Ovtracht, N., & Robardet, C. (2010). Characterizing the speed and paths of shared bicycle use in Lyon. *Transportation Research Part D: Transport and Environment*, *15*(8), 522–524. https://doi.org/10.1016/j.trd.2010.07.002

Johnson, R. A., & Wichern, D. W. (2014). *Applied multivariate statistical analysis* (Vol. 6). Pearson London, UK:

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity*, *52*(4), 1957–1976. https://doi.org/10.1007/s11135-017-0584-6

Jones, P., & Lucas, K. (2012). The social consequences of transport decision-making: Clarifying concepts, synthesising knowledge and assessing implications. *Journal of Transport Geography*, *21*, 4–16. https://doi.org/10.1016/j.jtrangeo.2012.01.012

Kamruzzaman, Md., & Hine, J. (2011). Participation index: A measure to identify rural transport disadvantage? *Journal of Transport Geography*, *19*(4), 882–899. https://doi.org/10.1016/j.jtrangeo.2010.11.004

Kamruzzaman, Md., & Hine, J. (2012). Analysis of rural activity spaces and transport disadvantage using a multi-method approach. *Transport Policy*, *19*(1), 105–120. https://doi.org/10.1016/j.tranpol.2011.09.007

Karner, A. (2018). Assessing public transit service equity using route-level accessibility measures and public data. *Journal of Transport Geography*, *67*, 24–32. https://doi.org/10.1016/J.JTRANGEO.2018.01.005

Lefrançois, R. (1998). Mobility Patterns and Attitudes Toward Driving a Car Among the Elderly Living in Small Towns and Rural Areas. *Rural Society*, *8*(1). https://doi.org/10.5172/rsj.8.1.17

Lewis, E. O. C., MacKenzie, D., & Kaminsky, J. (2021). Exploring equity: How equity norms have been applied implicitly and explicitly in transportation research and practice. *Transportation Research Interdisciplinary Perspectives*, *9*, 100332. https://doi.org/10.1016/J.TRIP.2021.100332

Litman, T. (2003). Measuring transportation: Traffic, mobility and accessibility. *ITE Journal*, *29*, 28–32.

Litman, T. (2021). Evaluating transportation equity. *World Transport Policy & Practice, 8*(2), 50–65.

Martens, K. (2017). *Transport Justice: Designing fair transportation systems*. Routledge.

Martens, K., Golub, A., & Robinson, G. (2012). A justice-theoretic approach to the distribution of transportation benefits: Implications for transportation planning practice in the United States. *Transportation Research Part A: Policy and Practice*, *46*(4), 684–695. https://doi.org/10.1016/J.TRA.2012.01.004

Mattioli, G. (2017). 'Forced Car Ownership' in the UK and Germany: Socio-Spatial Patterns and Potential Economic Stress Impacts. *Social Inclusion*, *5*(4). https://doi.org/10.17645/si.v5i4.1081

Mattioli, G., & Colleoni, M. (2016). Transport disadvantage, car dependence and urban form. In *Understanding mobilities for designing contemporary cities* (pp. 171–190). Springer.

McNeil, N. (2011). Bikeability and the 20-min Neighborhood: How Infrastructure and Destinations Influence Bicycle Accessibility. *Transportation Research Record: Journal of the Transportation Research Board*, *2247*(1), 53–63. https://doi.org/10.3141/2247-07

Meng, S., & Brown, A. (2021). Docked vs. Dockless equity: Comparing three micromobility service geographies. *Journal of Transport Geography*, *96*, 103185. https://doi.org/10.1016/j.jtrangeo.2021.103185

Merryid, K., & Bettinger, P. (2019). *Smartphone GPS accuracy study in an urban environment*. https://doi.org/10.1371/journal.pone.0219890

Minocha, I., Sriraj, P. S., Metaxatos, P., & Thakuriah, P. (Vonu). (2008). Analysis of Transit Quality of Service and Employment Accessibility for the Greater Chicago, Illinois, Region. *Transportation Research Record*, *2042*(1), 20–29. https://doi.org/10.3141/2042-03

Mooney, S. J., Hosford, K., Howe, B., Yan, A., Winters, M., Bassok, A., & Hirsch, J. A. (2019). Freedom from the station: Spatial equity in access to dockless bike share. *Journal of Transport Geography*, *74*, 91–96. https://doi.org/10.1016/J.JTRANGEO.2018.11.009

Morgan, M., Young, M., Lovelace, R., & Hama, L. (2019). OpenTripPlanner for R. *Journal of Open Source Software*, *4*(44), 1926. https://doi.org/10.21105/joss.01926

MPO. (2012). *Our Planning Area*. Indianapolis Metropolitan Planning Organization. https://www.indympo.org/who-we-are/our-planning-area

NCHRP. (2011). *NCHRP-184: Going the Distance Together: A Citizen's Guide to Context Sensitive Solutions for Better Transportation*. National Cooperative Research Program. https://ssti.us/wp-content/uploads/sites/1303/2012/04/nchrp_Citizens_Guide_to_Context_Sensitive_Solutions.pdf

Niemeyer, G. (2008, February 26). Geohash.org is public! *Labix Blog*.

Nussbaum, M. C. (2001). *Women and Human Development: The Capabilities Approach* (Vol. 3). Cambridge University Press.

OpenStreetMap. (2022). https://www.openstreetmap.org/

Páez, A., Scott, D. M., & Morency, C. (2012). Measuring accessibility: Positive and normative implementations of various accessibility indicators. *Journal of Transport Geography*, *25*, 141–153.

Preston, J., & Rajé, F. (2007). Accessibility, mobility and transport-related social exclusion. *Journal of Transport Geography*, *15*(3), 151–160. https://doi.org/10.1016/j.jtrangeo.2006.05.002

Pucher, J., & Renne, J. L. (2005). Rural mobility and mode choice: Evidence from the 2001 National Household Travel Survey. *Transportation*, *32*(2), 165–186. https://doi.org/10.1007/s11116-004-5508-3

Pyrialakou, V. D., Gkritza, K., & Fricker, J. D. (2016). Accessibility, mobility, and realized travel behavior: Assessing transport disadvantage from a policy perspective. *Journal of Transport Geography*, *51*, 252–269. https://doi.org/10.1016/j.jtrangeo.2016.02.001

Reck, D. J., Martin, H., & Axhausen, K. W. (2022). Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility. *Transportation Research Part D: Transport and Environment*, *102*, 103134. https://doi.org/10.1016/j.trd.2021.103134

Reichmuth, D. (2019). *Air Pollution from Cars, Trucks, and Buses in the U.S.: Everyone is Exposed, But the Burdens are not Equally Shared*. https://blog.ucsusa.org/dave-reichmuth/air-pollution-from-cars-trucks-and-buses-in-the-u-s-everyone-is-exposed-but-the-burdens-are-not-equally-shared/

Schuessler, N., & Axhausen, K. (2009). Processing raw data from global positioning systems without additional information. *Transportation Research Record*, *2105*, 28–36. https://doi.org/10.3141/2105-04

Sheller, M. (2018). Chapter 1: What Is Mobility Justice? In *Mobility Justice: The Politics of Movement in an Age of Extremes*. Verso.

Shen, L., & Stopher, P. R. (2014). Review of GPS Travel Survey and GPS Data-Processing Methods. *Transport Reviews*, *34*(3), 316–334. https://doi.org/10.1080/01441647.2014.903530

Shirmohammadli, A., Louen, C., & Vallée, D. (2016). Exploring mobility equity in a society undergoing changes in travel behavior: A case study of Aachen, Germany. *Transport Policy*, *46*, 32–39. https://doi.org/10.1016/j.tranpol.2015.11.006

Smith, C. S., Oh, J.-S., & Lei, C. (2015). *Exploring the equity dimensions of U.S. bicycle sharing systems*. Western Michigan University. Transportation Research Center for Livable….

Thyer, W. (2021, April 10). Visualizing and Analyzing Bicycle Infrastructure using OSMnx. *Visualizing and Analyzing Bicycle Infrastructure Using OSMnx*. https://williamthyer.github.io/posts/2021/4/bike_networks/

United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. https://www.un.org/sustainabledevelopment/sustainable-development-goals/

UPHAIL. (2022). *Uber,Taxi, Lyft in Indianapolis, IN*. https://uphail.com/us/in/indianapolis/

USGS. (2021). *USGS 1 meter Indiana Statewide Lidar Data*. https://apps.nationalmap.gov/downloader/#/

Vale, D. S., Saraiva, M., & Pereira, M. (2015). Active accessibility: A review of operational measures of walking and cycling accessibility. *Journal of Transport and Land Use*. https://doi.org/10.5198/jtlu.2015.593

Vallina-Rodriguez, N., Crowcroft, J., Finamore, A., Grunenberger, Y., & Papagiannaki, K. (2013). When Assistance Becomes Dependence: Characterizing the Costs and Inefficiencies of A-GPS. *SIGMOBILE Mob. Comput. Commun. Rev.*, *17*(4), 3–14. https://doi.org/10.1145/2557968.2557970

Veness, C. (2014). *Geolib*. https://www.movable-type.co.uk/scripts/geohash.html

Walkways. (2016). *Indianapolis / Marion County Pedestrian Plan*. Walkways. http://indywalkways.org/wp-content/uploads/2015/10/Indianapolis_Pedestrian-Plan_DRAFT_web_Pages.pdf

Wang, Z., He, S. Y., & Leung, Y. (2018). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, *11*, 141–155. https://doi.org/10.1016/j.tbs.2017.02.005

Welch, T. F., & Mishra, S. (2013). A measure of equity for public transit connectivity. *Journal of Transport Geography*, *33*, 29–41. https://doi.org/10.1016/J.JTRANGEO.2013.09.007

Wilcox, R. R., Char1in, V. L., & Thompson, K. L. (1986). New monte carlo results on the robustness of the anova f, w and f statistics. *Communications in Statistics-Simulation and Computation*, *15*(4), 933–943.

Winters, M., Teschke, K., Brauer, M., & Fuller, D. (2016). Bike Score®: Associations between urban bikeability and cycling behavior in 24 cities. *International Journal of Behavioral Nutrition and Physical Activity*, *13*(1), 18. https://doi.org/10.1186/s12966-016-0339-0

Zandbergen, P. A., & Barbeau, S. J. (2021). Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones THE JOURNAL OF NAVIGATION. *THE JOURNAL OF NAVIGATION*, *64*, 381–399. https://doi.org/10.1017/S0373463311000051

Zheng, Y., & Zhou, X. (2011). *Computing with Spatial Trajectories*.