SPATIO-TEMPORAL VARIABILITY AND DEMOGRAPHIC CHARACTERISTICS OF TRANSIT-BASED JOB ACCESSIBILITY: A GIS ASSESSMENT OF THE PUBLIC TRANSIT SYSTEM IN FLAGSTAFF, ARIZONA

By Antonio Henrique Caldeira Jorge Neves

A Thesis
Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Science
in Applied Geospatial Sciences

Northern Arizona University
May 2018

Approved:
Ruihong Huang, Ph.D., Chair
Rebecca Dawn Hawley, Ph.D.
Brian Craig Petersen, Ph.D.

ABSTRACT

SPATIO-TEMPORAL VARIABILITY AND DEMOGRAPHIC CHARACTERISTICS OF TRANSIT-BASED JOB ACCESSIBILITY: A GIS ASSESSMENT OF THE PUBLIC TRANSIT SYSTEM IN FLAGSTAFF, ARIZONA

ANTONIO HENRIQUE CALDEIRA JORGE NEVES

Accessibility measures the ease at which an individual can access a desired location. It is a major aspect in transportation planning, and transit systems are extensively used to improve accessibility. Well-designed public transit systems enable a high level of access to socioeconomic opportunities. This is especially important to socially disadvantaged populations due to their higher need for transit services to maintain a basic level of mobility. Increased transit-based accessibility can potentially diminish social exclusion rates and improve the well-being of these population groups. This research analyzes the spatio-temporal variability and the socio-demographic characteristics of transit-based job accessibility in Flagstaff, Arizona. This study employs a temporally-enabled schedule-aware simulation based on the transit system's General Transit Feed Specification and the city's street network. Gravity-based measurements were used in the calculation of accessibility from origin locations to job opportunities. This accessibility calculation considers the supply and demand location characteristics, the travel time impedance between them, and temporal variations in transit service frequency and availability. Statistical analyses were used to measure the relationship between accessibility and the individual socioeconomic attributes. The results served as the quantitative basis for discussing the social aspects of the city's transportation system.

Keywords: *transit, accessibility, social equity, network analysis, gis, gtfs*

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. Ruihong Huang, for the valuable guidance and support provided during my time as his student. His vast knowledge and extensive experience in GIS, spatial analysis, and accessibility have considerably added to the quality of this research project, and to my experience as a graduate student. I would also like to thank the other members of my committee, Dr. Rebecca Dawn Hawley and Dr. Brian Craig Petersen, for the assistance and encouragement provided during the development of this project. Their useful and constructive recommendations were critically important to broaden my understanding of the research topic.

I must express my gratitude to Northern Arizona University and the Department of Geography, Planning, and Recreation for providing the financial assistance which allowed me to successfully complete the graduate program and this research project. I must also acknowledge my fellow graduate student colleagues for their friendship and the countless hours spent together in the office. A special thanks to Emily Garrett, Corryn Smith, Daniel Foley, Monica Steele, Alex Arkowitz, Emily Melhorn-Kane, Olivia RoDee, and Lorna Thurston.

The assistance and support provided by NAIPTA and the NAU Campus Shuttle Services was also greatly appreciated. This work would not have been possible without the GTFS feeds these agencies provided me with. I would like to thank Alicia Becker from NAIPTA for being in touch since the beginning of this project and for making a great effort to help me with the project's data requirements.

Finally, my gratitude goes to my family, especially my parents Rossana and José Alberto, and my grandparents Maria Geralda and José Jorge, who unfailingly provided me with all the support I could have ever needed or asked for. They have shown me the value of education since the beginning, and they are responsible for everything I ever accomplished in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The world is becoming progressively smaller with the continuous technological improvements in transportation. This is not, however, absolutely accurate. It only holds true if individuals can both access and afford those services. At the 1999 ESRI User Conference, Waldo Tobler stated that "while the world is shrinking it is also shriveling" (Miller 2004: 286). Tobler meant that the world is indeed becoming smaller with the introduction of new technology, but the intensity at which it occurs is spatially and socially variable, depending both on accessibility and socioeconomic aspects. Accessibility is defined as "a measure of spatial separation of human activities" (Morris, Dumble, and Wigan 1979: 91). Within the specific context of this research, accessibility can be referred to as a measure of how easy it is for someone to reach a desired location using a specific combination of transportation modes.

Socioeconomic segregation is a reality in the United States, especially at the micro level that includes cities and neighborhoods (Massey, Rothwell, and Domina 2009). Segregation refers to the spatial separation of specific groups by social status, where relatively higher proportions of members of one group are geographically clustered. Race, income, education, and age are examples of social attributes on which segregation is based. These social traits are not only linked to spatial separation but they are also associated with the incidence level of socioeconomic disadvantages, including those related to mobility and accessibility. For instance, although car ownership is estimated to include 92% of all American households, only 60% of the lowest income quintile have access to a private vehicle (Lucas 2012). These levels are also different if racial aspects are considered, where white Americans are far more likely to own a car than those belonging to other races (Lucas 2004). These facts suggest that, overall, certain population groups have a higher need for public transit than others. And when these facts are not considered by

transportation authorities, groups that heavily rely on public transit will experience limited access to socioeconomic opportunities, such as jobs, education, health services, and recreation facilities. This lack of access to resources, opportunities, and services has a negative effect on the well-being of individuals and is referred to as social exclusion (Levitas et al. 2007).

Public transportation planning focuses on improving the connection between individuals and locations of interest. For this reason, accessibility is of primary importance for transportation planning, and public transit is widely used in solving accessibility problems, as it is affordable, environmentally friendly, and aids in decreasing congestion in saturated urban centers (Tribby and Zandbergen 2012). In addition, improving overall accessibility through public transit is also a way of reducing the negative effects of socioeconomic segregation. Tribby and Zandbergen (2012) address the use of transit systems as a tool to overcome social exclusion by facilitating the access to social and economic opportunities. Accessibility improvements can be reached in a few ways: reducing transportation costs and travel times to build up physical mobility, introducing information technology to improve social interactions and stimulate virtual accessibility, and land management measures that aim at the decentralization of facilities, opportunities and activities (Preston and Rajé 2007).

Equity is a common concern in public transportation planning, where transportation agencies and local governments are interested in providing services that are as comprehensive and fair as possible. The concept of equity, however, has more than a single connotation and can be branched into two major categories: horizontal and vertical equity (Shirmohammadli, Louen, and Vallée 2016). The main assumption in horizontal equity is that all individuals are equal and that the distribution of resources should be done in a way that both costs and benefits are impartially shared among individuals. In transportation, the result of equally weighting all individuals would

be the distribution of public transit resources purely according to the population distribution. On the other hand, vertical equity is associated with directing a greater proportion of benefits toward previously identified disadvantaged groups. In this case, the transportation resources are aimed at disadvantaged population groups with a higher need for those resources.

1.1 - PUBLIC TRANSPORTATION AND ACCESSIBILITY IN FLAGSTAFF

A number of local transportation services and resources are currently available in the City of Flagstaff (FMPO 2017). The Northern Arizona Intergovernmental Public Transportation Authority (NAIPTA) and Northern Arizona University (NAU) are the major providers of these services. NAIPTA's Mountain Line operates several bus fixed-routes throughout the city, including the Mountain Link, a high-frequency bus-rapid transit system that connects downtown Flagstaff, the NAU campus, and the Woodlands Village neighborhood. The Campus Shuttle Service at NAU provides free bus fixed-route service within the NAU campus along with SafeRide, a late-night bus service that operates between downtown Flagstaff and campus on weekend nights from 9:00 pm to 2:00 am.

NAIPTA also provides a complementary paratransit service aimed at individuals who are unable to benefit from the fixed-route bus services. NAIPTA operates the Mountain Lift, a curb-to-curb paratransit service where passengers are picked up and dropped off at the curb or driveway in front of their home or destination. This service is available within ¾ of a mile of any Mountain Line route. Taxi voucher programs are also available for eligible clients. In addition, all the vehicles in the NAU Campus Shuttle are currently fully equipped and meet all the requirements to transport individuals with disabilities.

Maintaining a basic level of mobility can be challenging for certain population groups, especially people with low incomes, individuals with disabilities, and senior citizens (FMPO

2017). The Flagstaff Metropolitan Planning Organization's 2017 Coordinated Public Transit and Human Services Transportation Plan was put together with the objective of developing a framework that improves the coordination between various transportation service providers to enhance the availability of these services to disadvantaged populations. Although various public transportation options are available in Flagstaff, there are still unmet transportation needs and gaps. A comprehensive list of needs and gaps is presented by the transportation plan. The most important and relevant for this study are highlighted below:

- Temporal gap: the fixed-route service operation hours do not meet the needs of many workers that need extended hours to access their jobs, especially on weekends;
- Spatial gaps:
    - Some Flagstaff neighborhoods are not properly served by fixed-route buses;
    - A number of human service agencies and senior housing projects are not accessible by the fixed-route bus service;
    - No transit service connects Doney Park, Mountainaire, Kachina Village, and Bellemont to Flagstaff and NAU for work and school journeys;
- Infrastructure gap: pedestrian and bicycle access to bus stops needs to be improved.

The transportation plan also lists the main goals and the associated efforts of the transportation plan given the current needs and service gaps. One of the objectives that can be highlighted is the intention to improve accessibility to jobs, education, and services by target population groups by increasing service frequencies and expanding coverage to underserved areas.

1.2 - RESEARCH STATEMENT AND PURPOSE

This research proposes to evaluate the City of Flagstaff's public transit system with respect to accessibility and social aspects. It intends to quantify the spatial and temporal variability of

accessibility to employment opportunities by public transit and analyze the distribution of accessibility across different socioeconomic groups. The objective is to correlate the accessibility levels with the demographic parameters assumed to determine social advantage or disadvantage and the need for public transit. Those results are expected to provide a quantitative basis for an evaluation of the social aspects of the transit-based job accessibility scenario.

The results would provide important information regarding the transit system's effectiveness in promoting social equity and reducing social exclusion. This is significantly relevant for the public transportation decision-making process and can provide valuable insight for the appropriate improvement of the city's public transit system according to the guidelines established by the FMPO (2017).

Geographic Information Systems (GIS) and spatial analysis techniques were the methods used to produce the results. From a methodological point of view, the goal of this research is to use freely available open datasets as input parameters in the development of a geoprocessing tool for the calculation of accessibility, where the measurements can be easily further correlated with demographic information. The objective is to simplify the assessment by compiling the analytical framework used to derive the accessibility measurements into a reusable tool. This way, the same procedure could also be appropriately applied to other locations if the data input requirements are met. The intention is to also embed a reasonable level of flexibility so that the process can be easily implemented in different study areas and benefit other transportation agencies and local governments.

1.3 - RESEARCH QUESTIONS

This research will focus on answering the following questions:

1) What is the spatial and temporal variability of the transit-based accessibility levels experienced by the city's residents to employment opportunities?
2) What is the spatial relationship between the accessibility measurements and the socio-demographic parameters that determine the need for public transit?
3) Does the accessibility scenario in Flagstaff reflect horizontal or vertical equity? In other words, are the individuals equally weighted or the resources are being distributed according to socioeconomic condition and the need for transit service?
4) Are there any locations with a relatively high need for public transit that are not currently covered by the transit system?

1.4 - RESEARCH SCOPE

This research is limited to the geographic context of the City of Flagstaff and the FMPO region (Figure 1). Given that the central idea of the study is the calculation of accessibility to jobs by public transit, it focuses on the area served by the public transit system within the city limits. However, a few neighborhoods of interest are located outside this perimeter and therefore a larger extent within the FMPO region boundary will be used in the demographic characterization of the area.

The accessibility calculation will not consider transportation modes other than public transit, walking, or a combination of both. Therefore, the results produced can only be treated as relative accessibility measurements given that different results would be obtained if other transportation modes were included in the analysis. For a given location, the accessibility measurement can be understood as the relative access to all potential employment opportunities within a specified travel time threshold by using the combination of transportation modes specified

above. In addition, the calculation will only include origin locations that are served by public transit. Those locations will be defined by a distance threshold from public transit stops.

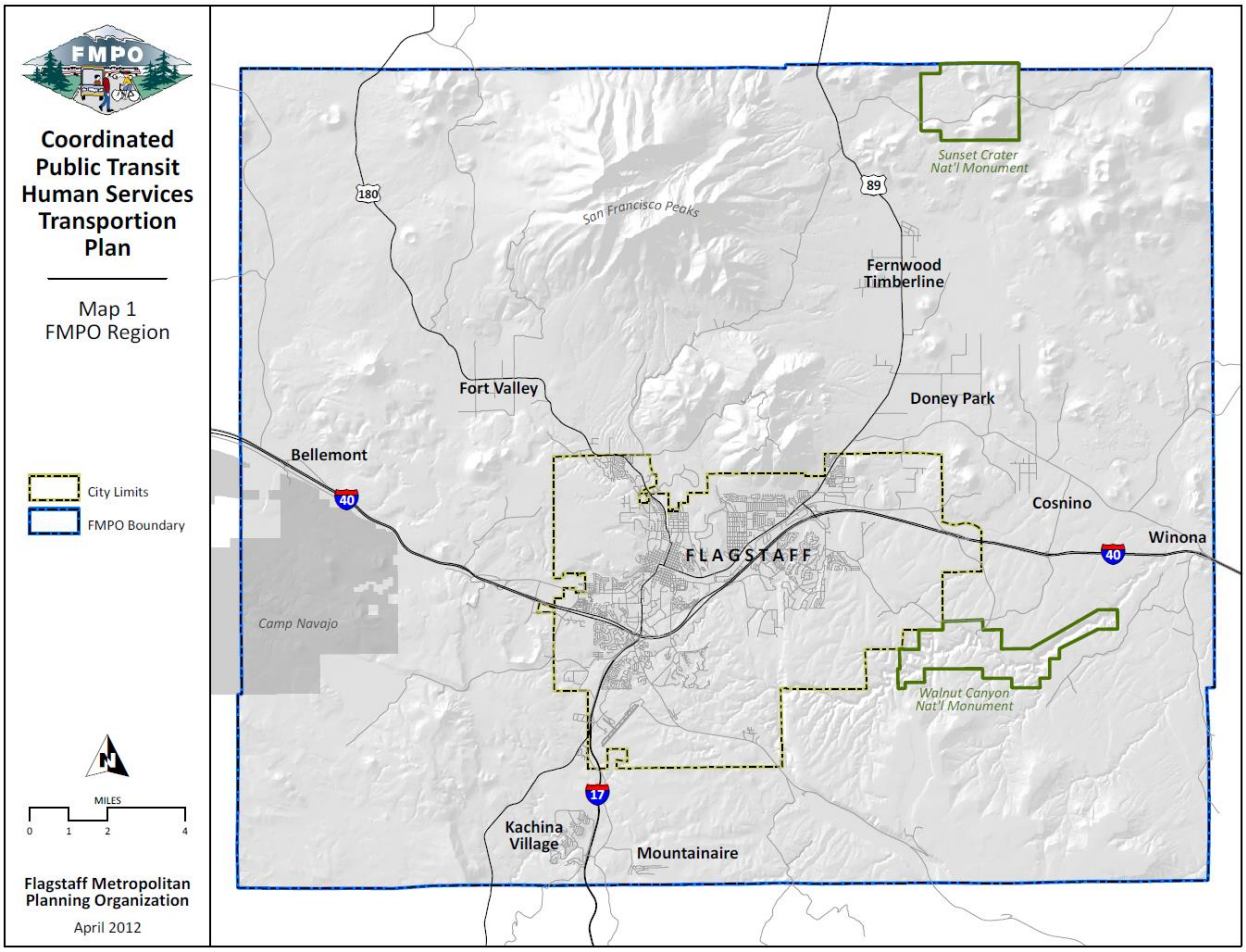

Figure 1. Flagstaff City Limits and the boundary of the FMPO region (FMPO 2017).

This study intends to analyze the current configuration of the public transit system and identify areas with an assumed high need for public transit that are currently underserved or not served at all. Therefore, although the demographic analysis will involve a larger extent than the accessibility calculations, the correlation between accessibility and socioeconomics will only be performed for the public transit service area, where the accessibility measurements will be taken.

CHAPTER 2: LITERATURE REVIEW

This chapter aims at the review of the existing literature on accessibility, and at presenting the relevant theoretical basis used to guide the methodology employed in this study. The specific terminology on accessibility will also be introduced. First, the general and more specific concepts of accessibility will be presented. Second, there will be a discussion on the social role of transportation systems. A summary of the accessibility measurement methods and perspectives will follow. Finally, there will be a review on the latest methodological approaches on accessibility, where the current knowledge gaps will be highlighted.

2.1 - ACCESSIBILITY

Hansen (1959) defines accessibility as a measurement of the ease at which people can interact with places. Accessibility is generally contrasted with mobility. While mobility emphasizes the transportation system itself, accessibility is also concerned with the land use parameters, i.e., the spatial distribution of the specific origins and destinations served by the system (Bhat et al. 2000). Therefore, mobility indicators only include criteria related to the transportation system, while accessibility would evaluate the transportation system from the perspective of the users (Ikhrata and Michell 1997). In other words, accessibility is not only concerned with the means to get to the destinations but also with the location of the destinations themselves. From a public transit point of view, mobility would measure the movement speed of a transit system. On the other hand, accessibility would not only consider that but would also account for the location of the relevant origins and destinations. Therefore, accessibility fundamentally refers to the relationship between spatio-temporal mobility and proximity to the opportunity landscape, which includes employment, health services, stores, and recreation facilities (Páez et al. 2010).

This research is concerned with the concept of transit-based accessibility, which is defined as “the ease of travel for an individual to reach a desired destination via public transit” (Shahandashti, Liu, and Zhang 2017: 2). This study will use a more specific definition of accessibility that is commonly adopted by transportation agencies: “accessibility is a measure of the ease of an individual to pursue an activity of a desired type, at a desired location, by a desired mode, and at a desired time” (Bhat et al. 2000). In this specific case, the desired activity is employment and the desired transportation mode is a combination of walking and traveling with transit.

2.2 - THE SOCIAL ASPECTS OF ACCESSIBILITY

Although the social function of a public transportation system is explicitly recognized by the local transportation authority (FMPO 2017), a more significant interest in the connection between transport policies and its socioeconomic effects is only recent, having started in the 1990s (Beyazit 2011). Within the context of accessibility and transportation planning, it is important to understand how the allocation of transportation resources impact different social groups with distinct mobility and accessibility needs. Neutens (2015) highlights the knowledge gaps after a comprehensive literature review on accessibility, and finds that there is a need for more studies that attempt to quantify social disparities in accessibility, as demonstrated below.

Traditional planning policy addresses transportation planning from an economic point of view (Litman and Brenman 2012). The conventional objectives are congestion reduction, increased mobility, savings in travel expenses, and transport safety. There has also been an increasing concern with environmental aspects, such as the conservation of natural resources, emission reductions, and protection of habitats (Litman and Burwell 2006). Many transportation

practices still dismiss the social impact of the distribution of transport resources, generally focusing on efficiency and environmental indicators (Shirmohammadli, Louen, and Vallée 2016).

Transport surveys often reveal that transport disadvantage is usually experienced by the most underprivileged population groups (Lucas 2012). Transport disadvantage is directly related to social exclusion. Lack of access to transportation by underprivileged groups may lead to the further inability of taking advantage of essential goods, services, and opportunities, ultimately leading to social exclusion (Figure 2). Therefore, it is important that social aspects are addressed by transportation planning policies.

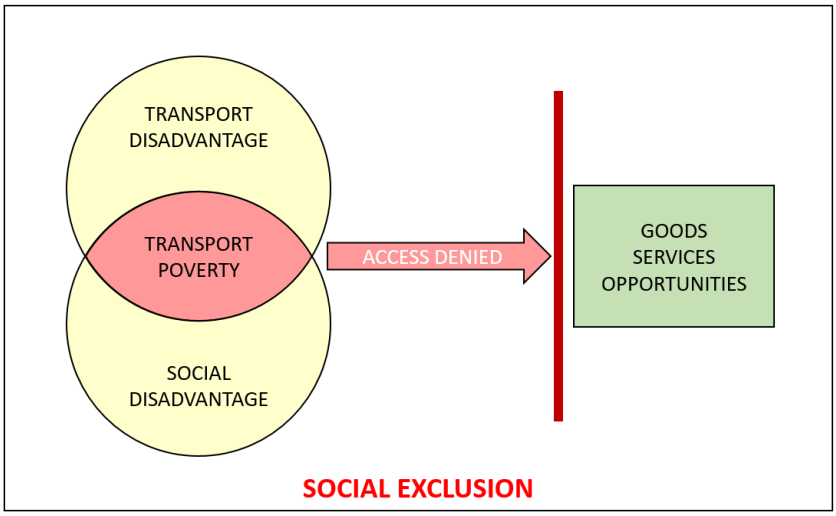

Figure 2. Diagram showing social exclusion as an implication of the simultaneous effect of transport and social disadvantage (modified from Lucas 2012).

Lately, several studies have focused on the social aspects of transportation and accessibility. Haas et al. (2006) observed a negative relationship between income and the percentage of income spent on transportation from a compilation of 2000 census data for 28 metropolitan areas in the US. The results show that the percent amount spent on transportation decreases as income increases. The average proportion of income spent by low income households (<$20,000) was 56%, while the classes with the top three highest incomes ($50,000 to $75,000;

$75,000 to $100,000; and $100,000 to $250,000) only spent 18%, 13%, and 8% respectively on average. This evidence shows that low income households are significantly more sensitive to the affordability of transportation.

Tribby and Zandbergen (2012) measured the travel time savings achieved with the implementation of two bus rapid transit (BRT) lines in Albuquerque, New Mexico. The results suggest a significant increase in accessibility in the area, as shown by the travel time savings. However, the improvements are negatively correlated with the transit users' index developed in the study. This index reflects the social need for transit service, based on income, private vehicle ownership, and age. Therefore, the results indicate that, although the transit improvements would help in decreasing congestion and pollution, the outcomes of this intervention did not mostly benefit the population groups assumed to have a higher need for transit service.

Similarly, Delmelle and Casas (2012) evaluated the consequences of the implementation of a BRT system in Cali, Colombia. The study addressed accessibility to the BRT system itself, and also to hospitals, recreation areas, and libraries. The accessibility to the transit system was measured by the walking times to stops or stations. The accessibility results were compared to the city's socioeconomic structure, determined by income, housing characteristics, and urban context. It was found that accessibility was the greatest for the middle class, while the lowest accessibility was experienced by the lowest and highest social strata. However, the most prosperous social class would less likely experience low accessibility due to higher access to private vehicles. The lowest socioeconomic strata is more likely to be dependent on the transit system and therefore may experience lower access to the city's opportunities.

Grengs (2010) analyzed job accessibility by disadvantaged populations in Detroit, Michigan. The target population groups were defined based on poverty rates and the proportions

of African American and Hispanic individuals. The study focused on three inner-city neighborhoods where the target population group was concentrated. The results suggested that, although the inner-city has the advantage of being in close proximity with a substantial amount of employment opportunities, good accessibility was conditioned to car ownership. Therefore, the residents of these neighborhoods are not geographically disadvantaged, but are deprived of opportunities due to low private vehicle ownership rates and poor transit service.

Milan and Creutzig (2017) use qualitative methods to measure the changes in quality of life due to a new transit development based on a participatory planning process in Medellín, Colombia. The measurements are based on a questionnaire that focuses on the perceived changes in social capital, well-being, and public infrastructure across different geographic zones, income levels, and gender. It was found that equity improved overall, and that well-designed transit interventions can potentially improve the lives of underprivileged groups in addition to being more environmentally friendly and aiding in decreasing congestion.

The evidence demonstrating that disadvantaged populations are more dependent on public transit justifies the increasing interest in the social aspects of transport-related accessibility by academia and transportation authorities. It is not only important to recognize the existence of social disparities and poor accessibility, but also to measure when, where, and the intensity at which they occur. A well-conceived transportation plan that takes those aspects into account has a greater potential to contribute for the achievement of social equity, ultimately improving the perspective and quality of life of disadvantaged population groups.

2.3 - ACCESSIBILITY MEASUREMENT METHODS

An accessibility measurement is an estimation of the ease at which specific locations of interest can be accessed. A significant number of accessibility measurement methods have been

developed over time, and different review articles focus on particular types of measurements or applications (Lei and Church 2010; Neutens et al. 2010; Páez, Scott, and Morency 2012; Neutens 2015; van Wee 2016). A more general and comprehensive categorization that applies to most measurement techniques is presented in the reviews by Geurs and Ritsema van Eck (2001), and Geurs and van Wee (2004). These authors describe the fundamental components of accessibility and the different perspectives from which accessibility can be measured. Accessibility measurements involve four basic types of components:

- The *land-use component* refers to the opportunity landscape or supply locations (jobs, schools, supermarkets, hospitals, recreation sites), and the location of the demand for the opportunities (where people live).
- The *transportation component* reflects the travel impedance (time, distance, monetary cost) that exists between the origins (demand) and destinations (supply). The travel impedance results from the characteristics of the transportation system available (street network density, road conditions, number of lanes, travel speed, transit fare, reliability, transit service frequency) and the spatial distribution of the land-use component.
- The *temporal component* describes the temporal variability associated with opportunity availability. It accounts for the times at which different opportunities are available, and for the variations inherent to the transportation component, such as the differences in levels of congestion and transit service frequency at different times of the day and different seasons of the year.
- The *individual component* distinguishes different types of demand according to the individuals' characteristics (age, income, education, household type), abilities (physical

condition and availability of transportation means), and desired opportunities (depends on the needs of an individual according to its characteristics).

Ideally, an accessibility measurement would incorporate all four elements in the analysis (Geurs and van Wee 2004). In practice, accessibility applications focus on one or more components, depending on the perspective from which the measurements are taken. Accessibility can be measured from four different perspectives: infrastructure, location, person, and utility.

Infrastructure-based measurements focus on the transportation system's performance, such as the level of congestion, average travel speed, and average delays. These are conventionally used in transportation planning and policy (Litman and Brenman 2012). Levine et al. (2012) analyze whether it is more efficient to have a transportation network that is dense, decreasing distances from origins to destinations, or a transportation network with high travel speeds. The results obtained suggest that the density effect allows for greater overall accessibility. A disadvantage of this perspective is that it focuses on the transportation component, and ignores the land-use component. In other words, it measures the efficiency of a system or its travel speed, but it does not consider whether the demand locations are being correctly connected with their specific supply locations.

Location-based measurements evaluate the accessibility to spatially dispersed opportunities from origin locations. They could be used, for instance, to calculate the number of opportunities within a specified travel impedance threshold, such as the time it takes to travel to the nearest supermarket (Farber, Morang, and Widener 2012), or the number of jobs located within walking distance from origin locations (Wang and Chen 2015).

Potential accessibility measures, or gravity-based measures (Hansen 1959), are location-based estimations that have been widely used in the literature. Gravity models estimate

accessibility from each demand location to all possible supply locations, where the influence of the opportunities decreases with travel impedance. Gravity-based measurements allow for the assessment of the joint effect of the spatial distribution of the land-use element and the characteristics of the transport element (Geurs and van Wee 2004). In addition, the distance decay parameter models the common behavior of a person when traveling with transit. People are less likely to be willing to travel longer distances. These measurements have been used as social indicators and are suitable for analyzing accessibility to socioeconomic opportunities for different social groups.

Accessibility can be analyzed at the individual level with *person-based* measures. The objective is to quantify the opportunities that an individual is able to access at a given time. These measurements are based on space-time prisms (Hägerstraand 1970). This framework takes into account personal possibilities and limitations, including time budgets and transportation mode availability and performance. Djurhuus et al. (2016) used SQL (structured query language) programming and GIS to integrate transit timetable data into a multimodal network to enable accessibility measurements at the individual level. Trips by bus, train, light rail, metro, ferry, walking, and cycling were included in the network. Their idea made possible the obtainment of accessibility areas for any address in the Capital Region of Denmark at any given time. An issue is that these measurements are demand-oriented, and usually data on the individual's time budget and destination choices is not available at the individual level (Thill and Horowitz 1997).

Utility-based measures analyze the economic outcomes of the level of access to spatially distributed opportunities. This measurement is based on economic theory and assumes that, given all the possible transport alternatives, a person will choose the alternative associated with maximum possible utility (Koenig 1980). Miller (1999a) developed a space-time accessibility

measurement based on utility-maximizing choice behavior. This calculation considers time budgets, opportunity attractiveness, travel impedances, and behavioral aspects. Although it is computationally possible to implement such measurements, the necessary data at the individual level is still difficult to obtain.

There is a wide range of variations of accessibility metrics that have been applied under the categories presented above. The next section is concerned with the latest methodological advances in accessibility research, and the knowledge gaps that should guide future studies. In addition, there will be a discussion on how this study intends to address the limitations of the current methods and contribute to the accessibility research agenda.

2.4 - ACCESSIBILITY RESEARCH AGENDA

Despite the significant accomplishments achieved by the extensive application of accessibility measurement methods, the current approaches used are still significantly affected by two major spatial issues (Neutens 2015): the *modifiable areal unit problem* (MAUP) and the *uncertain geographical context problem* (UGCoP); (Kwan 2012). Although there is an increasing interest in the relationship between social equity and transportation, the spatial distribution of accessibility across different population groups has also not yet been fully explored (van Wee and Geurs 2011).

The MAUP is a source of statistical bias that occurs when point measurements are aggregated into arbitrary geographic boundaries (Openshaw 1984). Different results are obtained depending on how those boundaries are defined (Miller 1999b). In transportation analysis, the existing data is commonly organized in arbitrarily defined geographic zones, such as census blocks or census tracts (Viegas, Martinez, and Silva 2009). Measurements based on large-scale aggregate

data are unable to capture smaller-scale variations within social groups and neighborhoods (Omer 2006).

The UGCoP refers to the uncertainty associated with the spatial delineation of area-based attributes that define the geographic context of a study (Kwan 2012). Geographers are unable to precisely establish spatio-temporal characteristics for the physical and social determinants that influence the phenomenon studied in area-based measurements. Therefore the area, timing, and duration of the attributes used in accessibility studies are always subject to some deviation from the true geographic context.

Due to those issues, future accessibility research should advance toward more disaggregate (person-based) and temporally integrated metrics; more refined geocomputational tools; and indicators that analyze the distribution of accessibility across different socioeconomic groups in an attempt to quantify social disparities (Neutens 2015). Neutens et al. (2010) propose that further studies should attempt to develop techniques that articulate more dimensions without making overly restrictive generalizations. In other words, there should be an effort toward the creation of an ideal measurement that does not significantly ignore any of the four basic components defined in the reviews by Geurs and Ritsema van Eck (2001) and Geurs and van Wee (2004).

Geurs and van Wee (2004) suggest that future work should aim at the development of more advanced accessibility measurements that are still relatively easy to be interpreted by researchers and policy makers. In addition, given that the interest in the social aspects of accessibility is relatively recent, there is a need for studies that look at the actual social benefits of transportation policies shaped by accessibility analysis (van Wee and Geurs 2011; van Wee 2016). The current studies are still focused on mapping and measuring accessibility-related social disparities. The outcomes of the application of those studies have not yet been analyzed.

With regard to transit-based accessibility measurements, Lei and Church (2010) show that many studies concentrated on the physical access to the transit system (i.e., distance or time it takes for a person to reach a transit stop or station), ignoring the entire journey from an origin to a desired destination. Ideally, as demonstrated by Tribby and Zandbergen (2012) and Djurhuus et al. (2016), transit travel time calculations should use a door-to-door approach, including walking times to transit stops, waiting times at the stops, travel times on the transit network, and any transfers that might be necessary.

To avoid the requirement of building complex GIS-based multimodal transit networks, Chen et. al (2017) use an internet mapping service to obtain realistic door-to-door travel time calculations for the city of Nanjing, China. Significant improvement was achieved with travel time calculations involving full travel chains, departure and arrival times, fluctuations in service availability at different times of a day, and traffic conditions.

In addition, simplifying assumptions have usually been made in studies that model the travel times between demand and supply, such as transfer and waiting times, and average travel speeds that ignore transit timetable data (Lei and Church 2010). Finally, another potential avenue for future research involves the use of GIS-based tools to forecast the accessibility changes caused by the introduction of potential new routes, modifications in the existing routes, or expansion of the street network. Such application would be extremely valuable for transportation agencies that intend to evaluate the impact of the proposed alternatives for future service changes.

Omer (2006) attempted to overcome the methodological issues (MAUP) associated with the use of aggregate datasets by using house-level socio-demographic point data in his accessibility analysis. His accessibility calculation was based on simple distance measurements (buffer and straight line) from residential locations to urban parks, and accounted for the quality of service

provided by the park (surface area normalized by the number of people assumed to be served by the park). The results were found to provide a significantly more accurate and sensitive assessment compared to the usual aggregated approach at the neighborhood level. Twenty six neighborhoods were revealed to experience inappropriate levels of accessibility at an aggregate resolution, while only nine neighborhoods were found to be underserved at the individual level analysis. Although the advantages of using socio-demographic data at the micro-level are significant, the use of georeferenced data at this level of detail raises critical issues concerning privacy and confidentiality of personal information.

Given that socio-demographic data at the micro-level is most commonly unavailable, Lovelace, Ballas, and Watson (2014) employed spatial microsimulation techniques to synthetically "reconstruct" the microdata from a combination of non-spatial survey data and geographic aggregate demographic data. The method works by exhaustively searching for the optimal combination that links the survey data with the aggregated counts. This alternative was found to be successful in modeling the intra-zone variability within aggregated census datasets. This kind of approach would have even been useful in the implementation of the new transit routes in Albuquerque (Tribby and Zandbergen 2012), in which the intervention could have been designed to facilitate access to the city's opportunities by low income residents.

El-Geneidy et al. (2016a) analyzed the temporal variability of transit-based accessibility to jobs and compared the results with socioeconomic data in an attempt to map the incidence of social disparities in Toronto, Canada. The study shows that the accessibility levels fluctuate over the course of a day, demonstrating the limitations associated with using a static measurement to represent overall accessibility. They also analyze the distribution of those measurements across different socioeconomic groups (based on income, unemployment rate, and proportion of recent

immigrants), and conclude that socially disadvantaged areas experience equal or better accessibility compared to socioeconomically prominent areas.

Similarly, Farber, Morang, and Widener (2014) use transit timetable data (GTFS feeds) to account for temporal variability in their study on transit-based accessibility to supermarkets in Cincinnati, Ohio. The travel time from each census block to the nearest supermarket was calculated for every minute of the day, revealing significant variability according to the level of transit service. In addition, the results were combined with census data on race, income, and age in an effort to search for social disparities. Exploratory statistical analyses were used to summarize those measurements and estimate the relationships between accessibility and socioeconomic parameters. It has been found that, although the degree of inequality between social groups is minor, many Cincinnatians can only reach a supermarket in 20 minutes or less during 20% of the day.

In an attempt to incorporate individual perception and behavior relative to public transit systems, El-Geneidy et al. (2016b) studied the effects of accounting for transit fares on accessibility measurements, which is a variable generally excluded from accessibility studies. Their study translated accessibility into a dollar value by incorporating hourly wages and transit fares in the calculations. They were able to demonstrate that accessibility is sensitive to transit fares and significant differences arise when transit fares are not considered, commonly leading to overestimation. In addition, the results can be easily interpreted by policy makers and the general public, given that a monetary value was used to represent accessibility.

Farber et al. (2014) also analyze transit fares as constraints by comparing the accessibility levels provided by flat fare with distance-based fare schemes. They demonstrated that distance-based fares can potentially lower the transportation costs of highly transit dependent populations, providing important information for public transit policies regarding the transit fare component.

Cheng and Bertolini (2013) use a modified job accessibility measurement that incorporates the effects of competition, job diversity, and distance decay. The method combines a probabilistic technique with a spatial interaction model (gravity-based). Employment data at the postcode level was disaggregated into a gridded arrangement with a cell size of 500 meters. The results suggest that job diversity has an impact on job opportunity, and that it should be considered in urban planning strategies. Mixed land use approaches can increase the variety of activities available and diversify job opportunities, ultimately improving job accessibility.

Jin and Paulsen (2018) investigate the impacts of job accessibility on unemployment rates and household income in the Chicago metropolitan area at the census block group level during 2000-2010. A gravity-based measurement was used to calculate accessibility to job opportunities in 2000 and 2010. Their findings suggest that job accessibility has an important social role. Specifically, improvements in job accessibility for African Americans lead to decreases in unemployment rates. Likewise, increased access to employment opportunities for low-income households not only increase employment, but also raises household income.

Although accessibility research has recently achieved significant progress, studies are usually unable to overcome all the existing limitations due to differences in data availability, scale of analysis, and methodological perspectives. For that reason, current research more often focuses on addressing one or more issues, but not all of them simultaneously.

The main difficulty in the accessibility field is related to the unavailability of disaggregate socio-demographic data and the associated effects of the MAUP and the UGCoP, especially when the objective is to map and measure social disparities in accessibility (Neutens 2015). In addition, although it is important that accessibility measurements are based on advanced geocomputational techniques, there should also be an effort towards the production of results that are still simple and

relatively easy to be interpreted by transportation planners (Geurs and van Wee 2004). Until recently, transit-based accessibility measurements focused on the physical access to the system, and generalizing assumptions have been made for the other steps of a typical trip with transit (Lei and Church 2010). This study attempts to address these issues and contribute to the current state of knowledge in a variety of ways:

- Travel time calculations are based on a door-to-door approach, and include walking times from start points to bus stops, waiting and boarding times at the stops, travel times with transit, any necessary transfers, and walking times from final stops to the destination;
- A transit-enabled network that combines the services of two distinct transit service providers (NAIPTA and NAU) was developed to simulate transit trips from multiple origins to multiple destinations. Not only can this network dataset be used for accessibility calculations, but also to solve any other transit-related network analysis problem that is relevant for transportation planning. This is especially relevant in a coordinated effort setting;
- The accessibility calculations are temporally-enabled and schedule-aware, accounting for the variations of accessibility over space and time. Instead of a single static measurement for a specific time of the day, the measurements were systematically repeated and summarized to account for fluctuations in transit service availability;
- The most recent socio-demographic datasets at the finest possible resolution available (census blocks and block groups) for the study area were used to minimize the effects of the MAUP and UGCoP;

- The social aspects of accessibility were addressed. The accessibility measurements were correlated with socio-demographic data to analyze the distribution of the results across different socioeconomic groups;
- The study not only provides useful information that can be easily interpreted by public transportation agencies and policy makers, but also establishes an analysis framework that can be readily adapted and reused according to the needs of different regions and transportation agencies.

CHAPTER 3: METHODOLOGY

This chapter starts by introducing the specific concepts, tools, and datasets used in the development of the analysis framework of this study. This background information is essential for understanding the specific methods employed in this research. Then, the particular methods and tools used in the study, and the purpose of the associated results, will be explained.

3.1 - BACKGROUND INFORMATION

3.1.1 - GIS and Spatial Analysis

"Almost everything that happens, happens somewhere. Knowing where something happens can be critically important" (Longley et al. 2011: 4). Location is a critical parameter to consider in many problems facing society. Geographic information is a category of information that not only records events, activities, and measurements, but also keeps track of where they took place on Earth's surface. Geographic data is the composite of spatial data (location) and attribute data (ESRI 2017a). Geographic Information Systems, or GIS, are comprehensive computer-based systems for storing, managing, displaying, and analyzing geographic information.

GIS solves geographic problems and summarizes information through spatial analysis and spatial modeling. Spatial analysis is a process that combines procedures, tools, and calculations to compile complex relationships and datasets into something simpler (Galati 2006). The objective is to expose relationships, patterns, and anomalies that are not immediately clear (Longley et al. 2011). A spatial model attempts to replicate or reproduce real world phenomena, suggesting their functionality and behavior according to specific conditions (Werner 1985). The excessive complexity of the real world prevents us from immediately and directly comprehending it. Therefore, we must look for similarities, regularities, and patterns to formulate spatial models.

Accessibility and transportation planning are inherently spatial, and location is a critical parameter that must be considered in the analysis. GIS, spatial modeling, and a number of spatial analysis techniques are employed in this study. A transportation network will be modeled and the movement of people through the network will be simulated. The data generated will be analyzed and a quantitative measurement of accessibility over space will be obtained. Specific details about the spatial analysis techniques used in this study will be presented in the following section.

3.1.2 - Transportation Network Analysis

Networks are sets of interconnected lines and points that represent real-world features. Networks are used to model the flow of goods and services, and they have been largely used to model utility and transportation infrastructure within GIS (Longley et al. 2011). The data model used to implement networks is simple (Zeiler 1999). Two fundamental elements constitute a network: nodes (or junctions) and the edges that connect them (Figure 3). Nodes or junctions are used to model street intersections, fuses, switches, water valves, confluence of streams, and other locations of interest. Edges are representations of streets, pipelines, streams, and transmission lines. These two basic components and the relationship between them establish the foundation for modeling the network behavior.

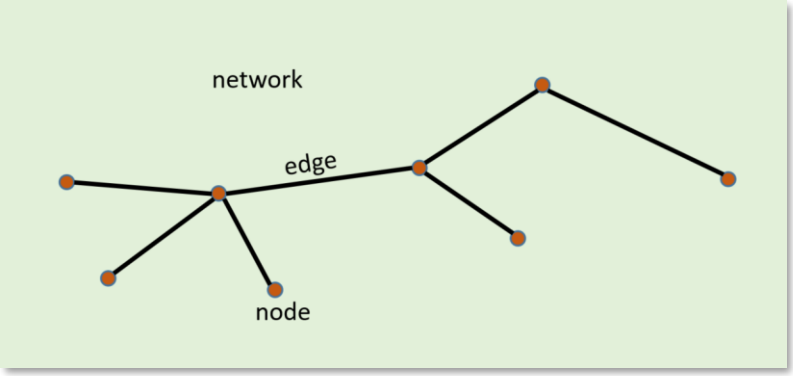

Figure 3. Basic elements of a network: edges and nodes.

A transportation network models the movement of people, vehicles, or goods (Bell and Iida 1997). In a transportation network, streets are represented as edges and intersections are modeled as nodes. If a transportation network includes a transit component (multimodal network), nodes are also used to represent transit stops and edges can also behave as transit lines (Figure 4). More specific details about transit-enabled networks and the functionality of the network used in this research will be presented in the next section.

Rules and attributes control how objects can move through the network. An attribute environment is used to establish costs, descriptors, restrictions, and hierarchy in a network (ESRI 2006). For instance, every edge (street or transit line) in the network is associated to a cost or impedance. This attribute determines how demanding it is to traverse the edge. This is user-defined and can be determined by the length of the street segment, the time it takes to travel through the segment according to the associated speed or travel mode, or the time it takes for a bus to go from one stop to another. Direction of movement can also be restricted in a network, and that is especially useful to model one-way streets. In the network data model, that would be equivalent to allow movement from a given node to another through an edge, but not the other way around.

Connectivity rules determine the allowed connections between edges and nodes (Zeiler 1999). Not all nodes are connected by edges, and not all edges that intersect in two-dimensional view define nodes (Goodchild 1998). An intersection of edges without nodes means that there is no connectivity between the edges. For instance, in Figure 4, although some transit lines intersect streets at certain locations (highlighted in blue circles), movement from the “street network” to the “transit network” is only allowed at a transit stop (highlighted in green circles). Multimodal transportation networks include more than a single transportation mode, such as walking, driving,

or transit. Connectivity rules are the basis for modeling multimodal networks, where different transportation modes are separated by connectivity groups (ESRI 2006).

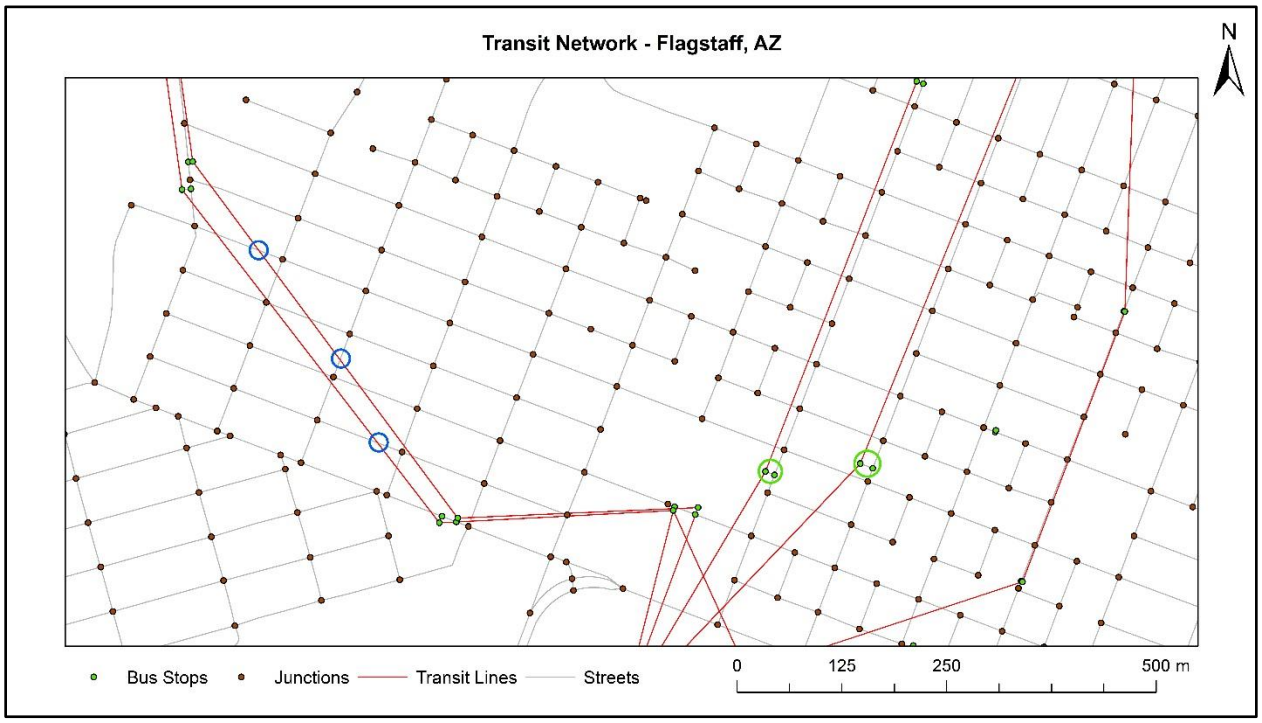

Figure 4. Network model example: transit-enabled transportation network.

The connectivity in a network allows for analysis and problem solving. Network analysis can be performed with the ArcGIS Network Analyst extension (ESRI 2017c). The Network Analyst extension uses network datasets and network analysis layers (Figure 5) to solve network problems. Network datasets are collections of interconnected features (junctions, edges, and turns) that model undirected flow. Undirected flow refers to the fact that the objects moving through this type of network have their own will and can choose to move in any direction (Zeiler 1999). Network analysis layers are created from network datasets, and they provide the framework for configuring the calculation parameters of a specific network problem to be solved.

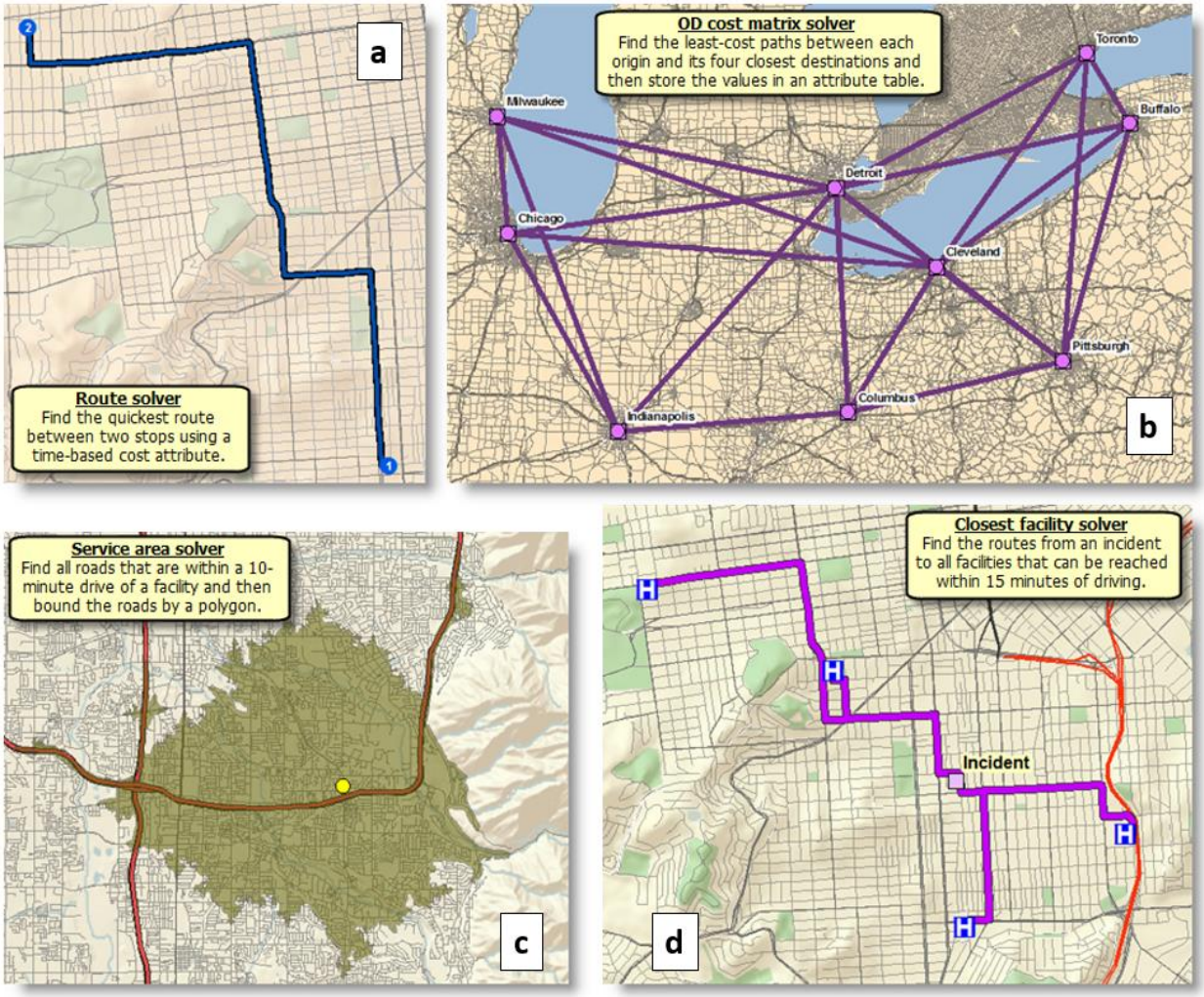

Figure 5. Types of network analysis layers (modified from ESRI 2017b): Route (a), OD cost matrix (b), Service area (c), Closest Facility (d). Copyright © 2017 Esri.

Common transportation network problems can be solved with the Network Analyst extension (Figure 5). A route layer can be used to find the best path from one location to another (Figure 5a), or the best way a delivery or service vehicle can visit several locations, improving customer service and minimizing transportation costs. Given an origin location and a set of facilities (potential destinations), the closest facility or the facility that minimizes travel time can be determined by solving the specific network analysis problem (Figure 5d). Service areas around any location can be calculated to determine the extent that encompasses all locations within a

specific travel time or distance from the origin location (Figure 5c). For instance, service areas can be used to determine all addresses located within a five-minute drive from a fire or police station. An origin-destination cost matrix (OD cost matrix) can be created to calculate the travel time or distance from multiple origins to multiple destinations (Figure 5b).

These network analysis tools can be applied to a number of applications in accessibility research. Service areas or OD cost matrices can be used to determine the number of people served by a facility, such as supermarket, hospital, or school. For example, 30-minute drive service areas can be generated for local supermarkets, and the number of people that need to travel more than 30 minutes to reach a supermarket can be determined. In this study, OD cost matrices will be used to determine the number of jobs that are accessible within a one-hour travel time threshold from multiple locations.

3.1.3 - General Transit Feed Specification (GTFS)

The General Transit Feed Specification (GTFS) establishes a universal data format for public transit timetables and the associated geographic attributes (Google Developers 2016). The GTFS format was developed with the objective of establishing a common system for public transit agencies to distribute their transit schedules. A common specification allows computing applications to consume transit datasets in an interoperable way. In other words, with the standardization of transit schedule data, an application developed to perform a specific analysis for the transit system of a given location will also be able to perform the same analysis for any other transit system with GTFS data available. This allows for public transit agencies to benefit from any GTFS-based application if their GTFS feed is available. For instance, Google Maps uses GTFS feeds from transit agencies to incorporate transit as a transportation mode in their direction

calculations. One can easily look up transit directions for many different cities in many different countries.

A GTFS feed is a collection of text files with tabular data. Each file contains a table with information about a feature of the transit system: stops, routes, trips, fares, stop times, transfers, and other related schedule data. The GTFS feed for the local public transit system will be used in this project to implement schedule-aware simulations of trips by public transit.

3.2 - METHODS

Most of the data processing and analysis will be performed within a GIS environment, specifically ESRI's ArcGIS for Desktop platform. Python scripting will also be used to automate tasks within ArcGIS, and R programming (RStudio) will be used for statistical analyses. The analysis will consist of four major components: 1) the construction of a transit-enabled network dataset; 2) the collection and processing of census geographic data and the associated socioeconomic attributes; 3) accessibility calculation through network analysis; and 4) statistical analyses to summarize the measurements and correlate accessibility with socioeconomics.

3.2.1 - Building the Network Dataset

The basic requirement of this research is creating a transit-enabled network dataset for the city's transportation network that includes the transit systems operated by both NAIPTA (Mountain Line) and NAU (Campus Shuttle Services). Although the shuttle service operated by NAU is mostly used by students, NAU is the largest employer in the City of Flagstaff (COF 2018), and NAU employees and student workers use this service to get to their workplace within the university campus. Besides, this service is connected to the Mountain Line network in some bus stops within the NAU campus, and it would be important that transfers between these systems are incorporated in the network dataset. Therefore, this transit service is significantly relevant for the

realistic simulation of work trips with transit in Flagstaff. Updated GTFS feeds were provided by NAIPTA and NAU, and the street network dataset was obtained from the City of Flagstaff Open GIS Data Portal (COF 2016).

A custom ArcGIS toolset called "Add GTFS to a Network Dataset" (Morang 2016) was used to merge the GTFS feed with the street network. Add GTFS to a Network Dataset allows the user to “translate” GTFS public transit feeds into ArcGIS network dataset features to enable transit-based schedule-aware analyses using the Network Analyst tools, like best route, closest facility, location-allocation, service area calculations, and origin-destination cost matrices.

This tool processes the street network and the associated GTFS feeds to produce the necessary spatial elements for the creation of a multimodal network dataset. It essentially reads the tables that make up the GTFS dataset and creates spatial features for the transit lines and stops. These features, together with the street network, are used as sources in the creation of the network dataset.

The ArcGIS Network Analyst extension was used in the creation of a multimodal network dataset from the data elements generated by the previous procedure. Figure 6 presents a simplified diagram that summarizes the behavior of this transit-enabled network and the calculations performed when simulating a trip with transit. The network functionality allows travel time calculations that account for walking times (assuming a walking speed of 3 mph) from start points to public transit stops, waiting and boarding times at the stops, travel times with transit, any necessary transfers, and walking times from stops to the destination.

As illustrated in Figure 6, the network dataset is composed of two distinct but connected networks: the street network and the transit network. The calculation of a trip from an origin location A to a destination B includes: walking time from start point A to stop 1, waiting and

boarding times at stop 1, travel time with transit from stop 1 to stop 2, walking time from stop 2 to stop 3, waiting and boarding time at stop 3, travel time with transit from stop 3 to stop 4, and finally the time it takes to walk from stop 4 to point B (Figure 6).

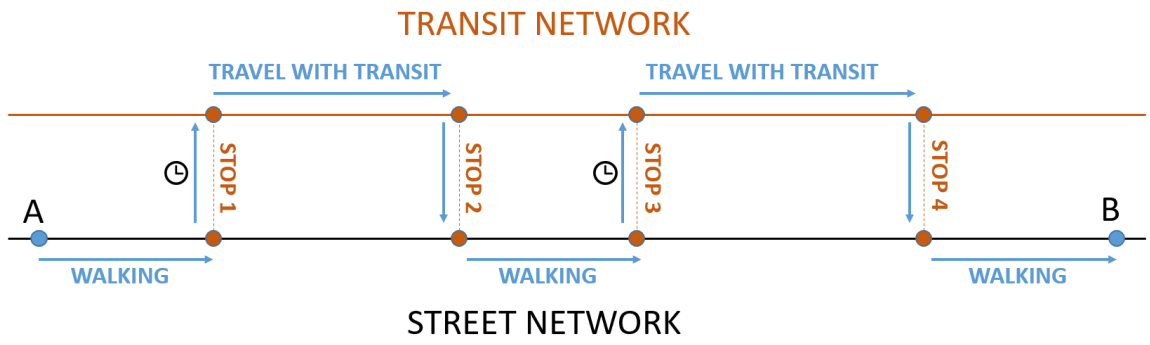

Figure 6. Diagram showing the transit-based network dataset's functionality.

3.2.2 - Census Data Collection and Processing

Census blocks, block groups, and the associated relevant socioeconomic attribute data were collected (USCB 2017a). These datasets were used in the socioeconomic characterization of the study area and further correlation with the accessibility measurements. An index was developed to estimate the need for transit service for work trips, and correlations with individual variables were also performed. Attribute data on total population, age, income, employment, commuting characteristics, disability status, and vehicle ownership were used.

The need for public transit was estimated from attribute data on total population, age, income, disability and vehicle ownership. The estimation was based on a modification of the transit users' index used in the study by Tribby and Zandbergen (2012). An average of the variables' (income, car ownership, and age) proportions was used in that study, where the index ranges from 0 to 1. The closer the index is to 1, the higher the need for transit due to a greater relative amount of lower income residents, no car ownership households, and non-driving age individuals. This index needed to be adapted for better compatibility with the study area characteristics. The original index was based on a larger study area (Albuquerque, New Mexico) and used attribute data at the

block group level. Flagstaff is a smaller city and block groups are too coarse to be used for this purpose, therefore census blocks are more appropriate due to their finer resolution.

At the block level, there are two issues: the unavailability of data on income and car ownership; and blocks where the population is very low. Therefore, it would be unrealistic to use the proportion-based method mentioned above because some census blocks would end up being associated with high proportions despite their population being very low. To account for that, the average proportion was multiplied by the total population in the block. The proportion of individuals with a disability was also incorporated in the index. Data on income, car ownership, and disability were obtained by joining the proportions at the respective block group to each census block.

Census blocks were also processed to be used as origin locations in the network analyses. The procedure included transforming the polygon feature class that represents the census blocks into centroids. This centroid point dataset was then snapped to the street network. The procedure is illustrated in Figure 7.

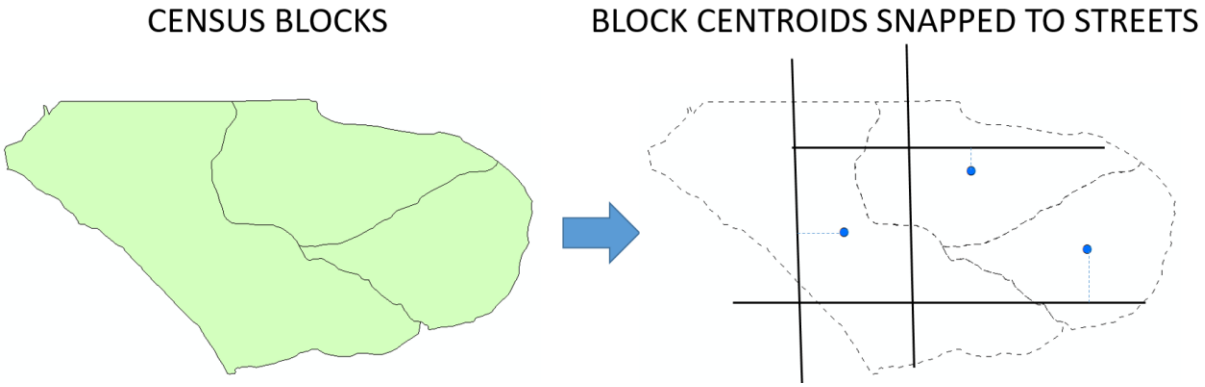

Figure 7. Census block centroids were computed and snapped to the nearest street location.

In addition to calculating the accessibility levels experienced by locations with a higher or lower need for public transit, the study also intends to address the Title VI requirements (FTA 2017). These requirements aim to ensure that the public transportation services are provided in a

way that individuals are protected from being discriminated based on race, color, or national origin. Therefore, accessibility was also correlated with the data on race and national origin.

3.2.3 - Accessibility Analysis

The functionality of the transit-enabled network was used to measure accessibility to the city's employment opportunities by public transit. This study will employ a location-based potential accessibility measure based on a gravity model (Hansen 1959). The accessibility value at a given time and origin location is equal to the sum of the attractiveness of all socioeconomic opportunities divided by the respective travel impedance associated with those opportunities. Figure 8 presents the equation used and a diagram that illustrates the gravity model. The accessibility at an origin location i (A_i) is equal to the sum of the attractiveness (a_j) at every single opportunity (supply location) j divided by the travel time with transit from location i to opportunity j (t_{ij}).

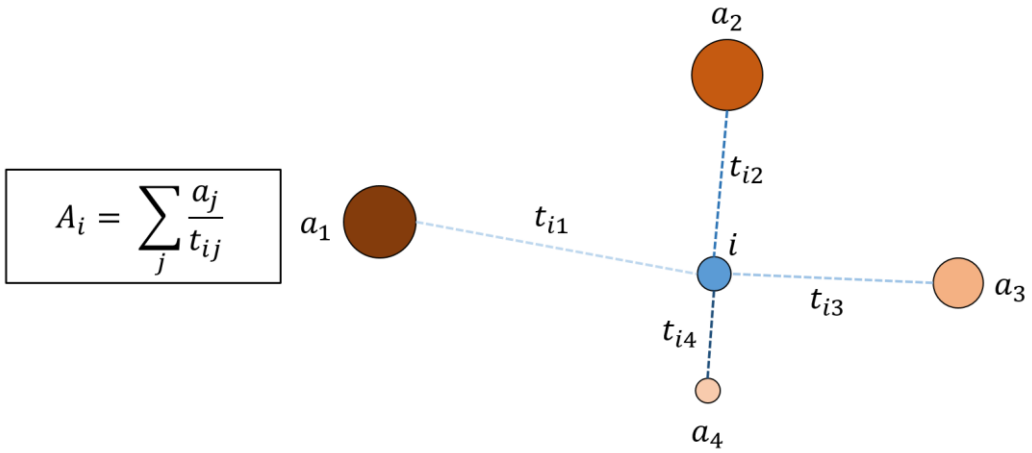

Figure 8. Equation and diagram describing the gravity model mechanism (Hansen 1959).

The travel impedance measurement (t_{ij}) will be obtained by calculating the travel time with transit from the city's census block centroids (Figure 9) to the opportunities. To ensure that only the actual transit users were included in the analysis, only the populated census block centroids

located within a quarter mile walking distance to a bus stop were used in the calculation. A total of 753 locations were included.

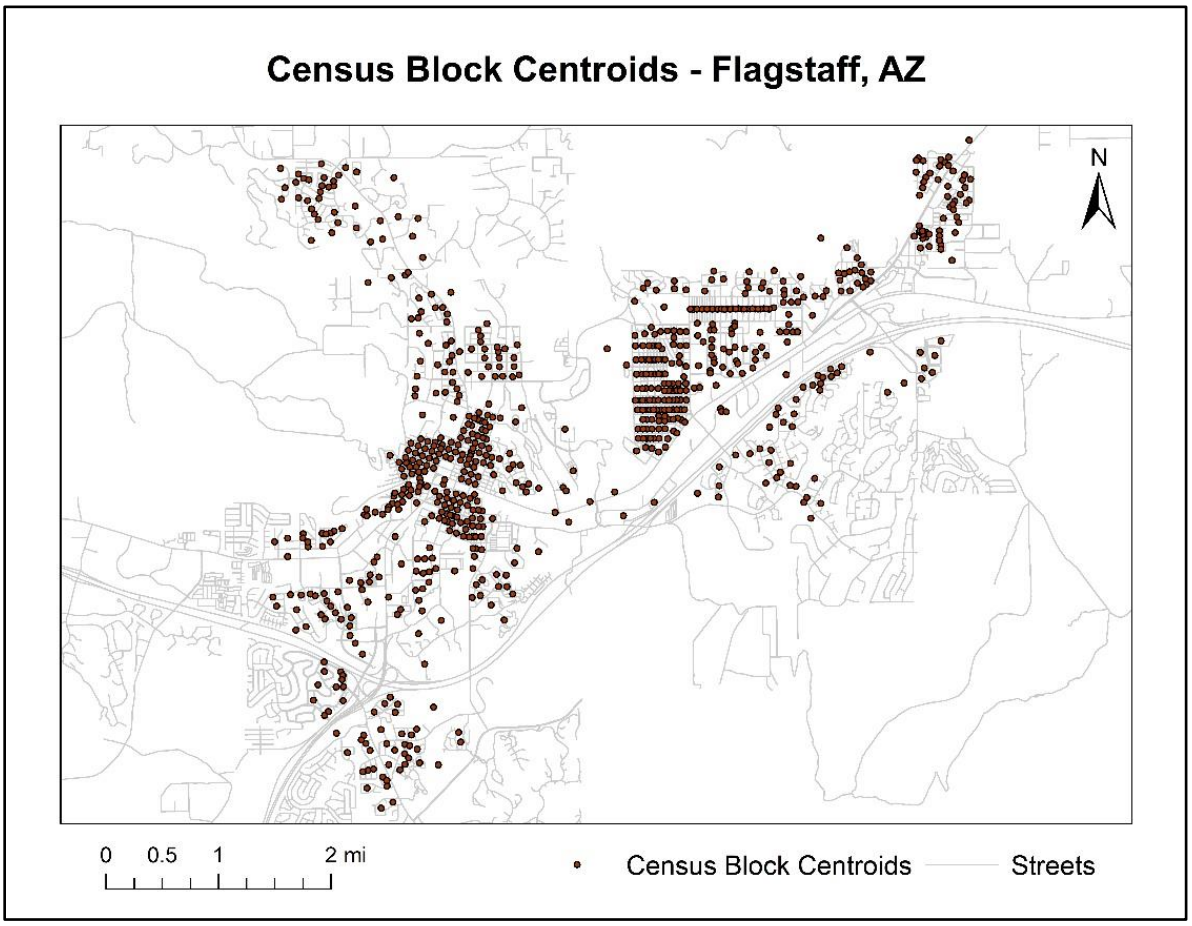

Figure 9. Flagstaff's census block centroids located within walking distance to a bus stop.

The travel time measurements were obtained through an OD Cost Matrix Analysis in the network dataset (Figure 10). The OD cost matrix identifies and measures the fastest routes from various origins to multiple destinations and records the results in an attribute table (ESRI 2016). This tool was used to calculate the travel time with transit from each census origin location to every opportunity location. In Figure 10, the straight lines are not straight-line distance measurements. They are simply a representation of a trip from the origin to a specific destination.

Each line has a corresponding attribute value for the travel time with transit measured in minutes, used as the travel impedance variable in the accessibility calculation.

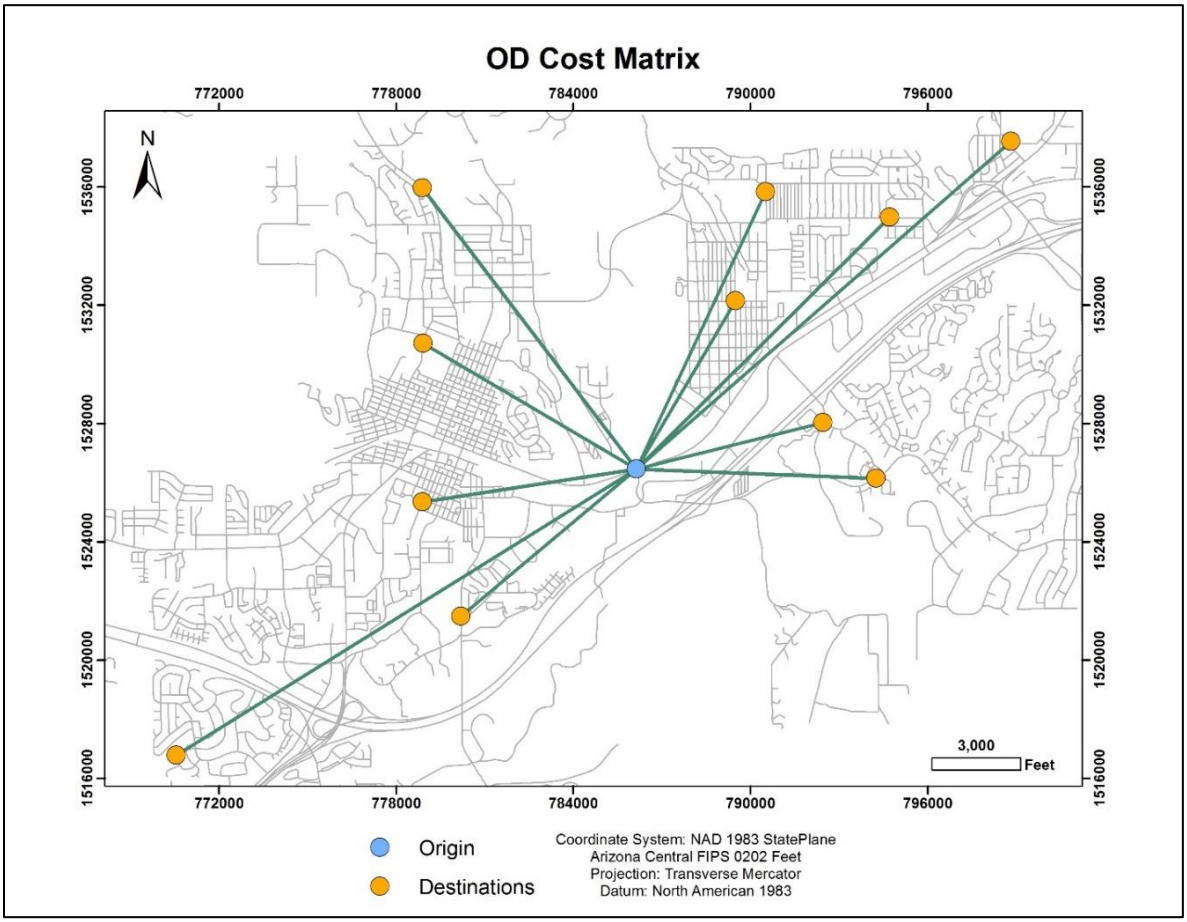

Figure 10. OD cost matrix from a single origin to multiple locations.

In this specific situation, employment opportunities are assigned the supply locations, while the number of jobs in each location is the measure of attractiveness. The employment opportunities were obtained from the Longitudinal Employer-Household Dynamics (LEHD) Origin–Destination Employment Statistics (LODES) data product (USCB 2018). This dataset contains aggregated attribute information on the number of jobs at each census block in a given year (Figure 11). The LEHD framework gathers earnings records and QCEW (Quarterly Census of Employment and Wages) data quarterly (Graham, Kutzbach, and McKenzie 2014). In a given

year, a job is included in the dataset if a person has positive income both in the reference quarter (April to June) and in the previous quarter (January to March).

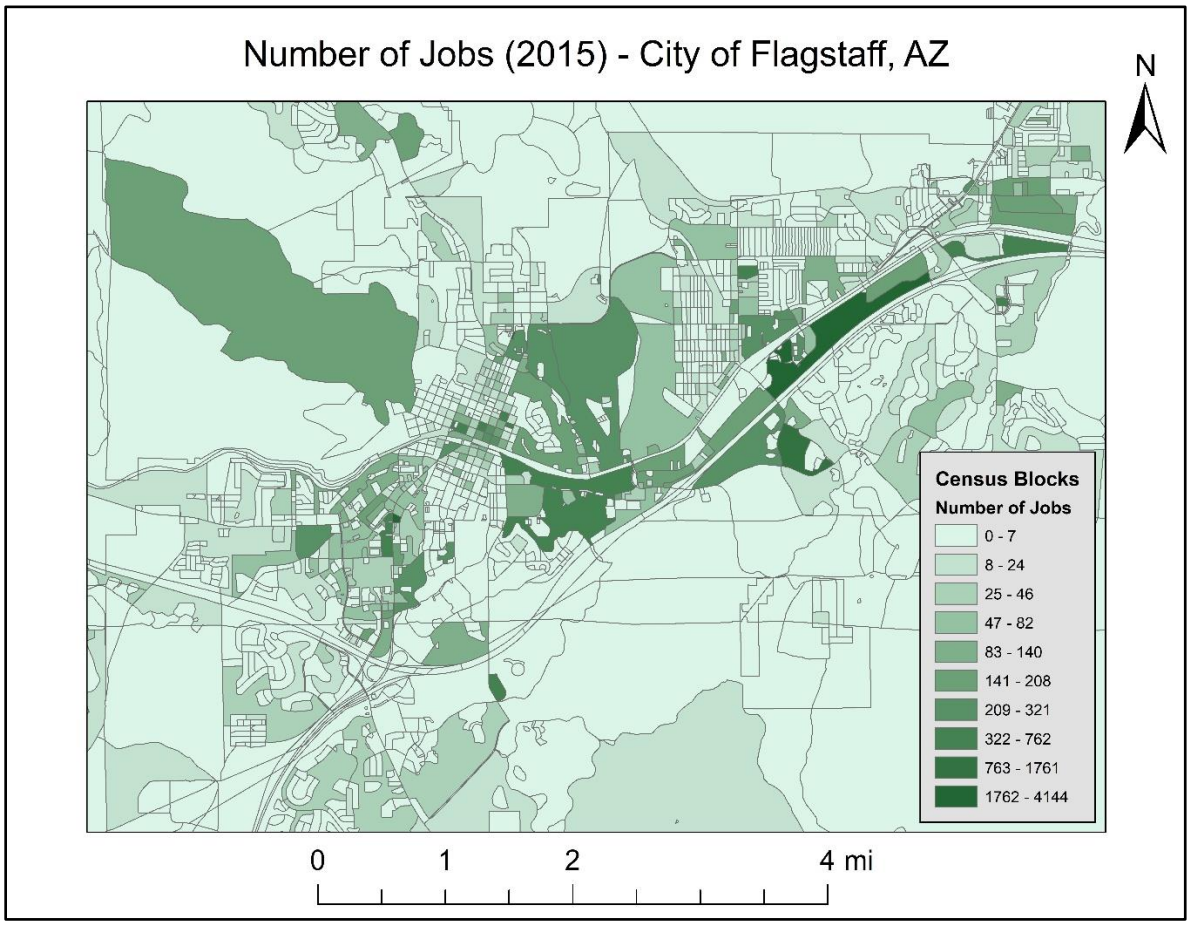

Figure 11. Census blocks symbolized according to the number of jobs within each block.

The gravity model normalizes the influence of the opportunities in the calculation of accessibility. The harder it is to reach an opportunity, the higher the travel impedance value by which the attractiveness measure (number of jobs) is divided. Therefore, opportunities with very high or very low attractiveness values have a less substantial weight in the calculation if they are significantly distant from the origin locations. This distance decay is important to account for both people's perceptions of a transit system and competition for jobs. People are usually less likely to travel longer distances, and distant opportunities are given a lower accessibility value.

Due to the size of the city and the spatial coverage, connectivity, and service frequency of the transit system, a maximum commuting time of one hour was assumed to be realistic. Several trips that require transfers can often take approximately one hour. For any given origin location, opportunities located outside this travel time threshold were not included in the accessibility calculation. Calculating OD cost matrices for large datasets and multiple times in a day can be computationally inefficient and impractical (Shahandashti, Liu, and Zhang 2017). The travel time threshold also aided in avoiding excessively long processing times and extremely large output datasets.

As previously mentioned, this study intends to capture the temporal variability of accessibility in the study area. Variations in service availability lead to travel time differences throughout the day and for different days of the week. Therefore, the accessibility measurement needs to be repeated for multiple times in a day and different days of the week. A sampling scheme was designed based on the timetables for the different transit routes operated by NAIPTA's Mountain Line.

The Mountain Line routes run on two schedules: a weekday schedule and weekend schedule (NAIPTA 2018). The weekend schedule also applies to holidays. In addition, the minimum headway (interval between two consecutive services) associated with those routes is 10 minutes during the week and 20 minutes on weekends. Therefore, to best capture the variations in service availability and consequently accessibility, the accessibility measurements were performed for a single weekday (Monday) and a single weekend day (Saturday). Given the minimum headways, a 5-minute sampling frequency was used. There is no need to repeat the analysis for additional days given that the schedule associated with the GTFS feed is the same for different weekdays or weekend days.

It would have been extremely time consuming and possibly impractical to manually perform all the individual calculations without any kind of automation. To avoid that, a Python script tool (Appendix A) was developed to automate the repetitive calculations. This tool automatically performs all the OD cost matrix and accessibility calculations, summarizing the results in a table. For each census block, the output table includes a record with attribute fields for the accessibility value at any given time.

3.2.4 - Statistical Analysis

Statistical procedures were used to summarize the individual accessibility measurements obtained in the previous step. The mean and the coefficient of variation (CV) of those measurements were calculated. The mean provides an assessment of the overall accessibility level experienced at each census block. The higher the mean, the better the overall accessibility. The coefficient of variation (standard deviation divided by the mean), in turn, measures the variability of accessibility relative to the mean at each location. Higher CVs are associated with greater variations in transit service availability. Therefore, the ideal situation where the public transit system provides high access to opportunities is characterized by most locations being associated with high means and low CVs. This means that, for most places and at most times, the accessibility level is high and does not vary significantly.

Average measurements that account for the temporal variability of accessibility were obtained for each census block for both weekdays and weekends. Socioeconomic attributes are also available for each census location. The next step involved analyzing the relationships between these two variables. An overall mean accessibility measurement was obtained to simplify this comparison. A weighted average of the weekday and weekend mean accessibility values was used. Given that the mean weekday accessibility is experienced five times in a week, it was given a

weight of 5. Similarly, a weight of 2 was assigned to the weekend mean accessibility. Graphing procedures and correlation coefficients were used to summarize the key aspects of those relationships.

Bar charts were used to produce visual representations of those relationships. Accessibility values were displayed on the y-axis, while the number of people associated with a given socioeconomic variable was plotted in the x-axis. For instance, to analyze the relationship between accessibility and income, the hypothetical bar chart presented in Figure 12 could be produced. In this graph, the accessibility experienced by different ranges of low income household numbers can be obtained.

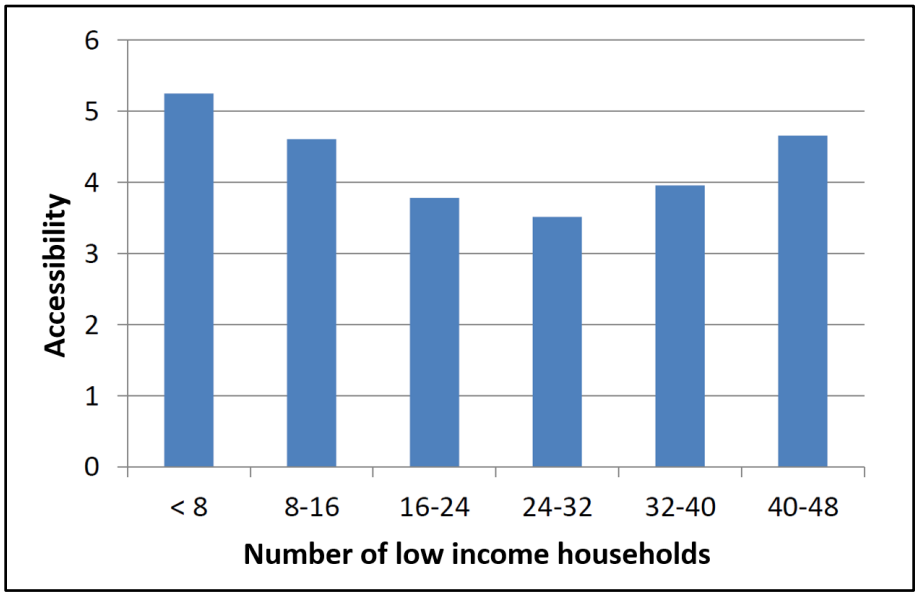

Figure 12. Hypothetical bar chart showing the relationship between accessibility and the number of low income households.

Relationships between accessibility and the individual socioeconomic variables can be inferred by visual inspection. For instance, in Figure 12, the distribution of accessibility shows no apparent trend. In this case, the bar chart characterizes a situation where the transit-based accessibility levels are more likely impartially distributed relative to household income.

Correlation coefficients were also calculated so that the relationships inferred by visual inspection could be measured. Given that the variables measured are not normally distributed and there is no homoscedasticity in the data, the assumptions required for regression models and the Pearson's correlation coefficient are violated. For that reason, Spearman rank-order correlation coefficients (r_s) were calculated instead. Pearson's coefficient measures the linearity of the relationship between two variables, meaning that the change in values in one variable is associated with a proportional change in the other variable. On the other hand, the Spearman correlation coefficient quantifies the monotonic relationship between two variables. In a monotonic relationship, the variables also change values simultaneously, but not necessarily at a constant rate.

This coefficient measures the direction and strength of the correlation between two variables and assumes values that range from -1 (perfect negative association) to +1 (perfect positive association). For instance, if the accessibility levels increase with the number of low income households, the coefficient will assume a positive value for the relationship between these two variables. Figure 13 shows hypothetical examples of coefficients calculated for different types of relationships between two variables.

If the measurements in both variables increase simultaneously (top left graph in Figure 13), the correlation coefficient will assume a value closer to +1. If the measurements in one variable increase while the values in the other variable decrease (top right graph in Figure 13), the coefficient will be negative and closer to -1. For intermediate situations, where the nature of the relationship is not very clear (bottom graphs in Figure 13), the coefficient assumes more moderate values. Given the nature of this study's datasets, and that it is impractical for a transit system to completely satisfy the demand of all locations associated with a high need for transit, weaker relationships are more common outcomes in this research.

The results of this step were used as the quantitative basis for a discussion on the social equity aspects of the city's public transit system. In addition, the results are expected to provide valuable insight that could benefit the current coordinated transportation plan established for the city (FMPO 2017). The information obtained in the analysis is relevant for the objectives and efforts of the transportation plan, given the current needs and gaps in the transit service.

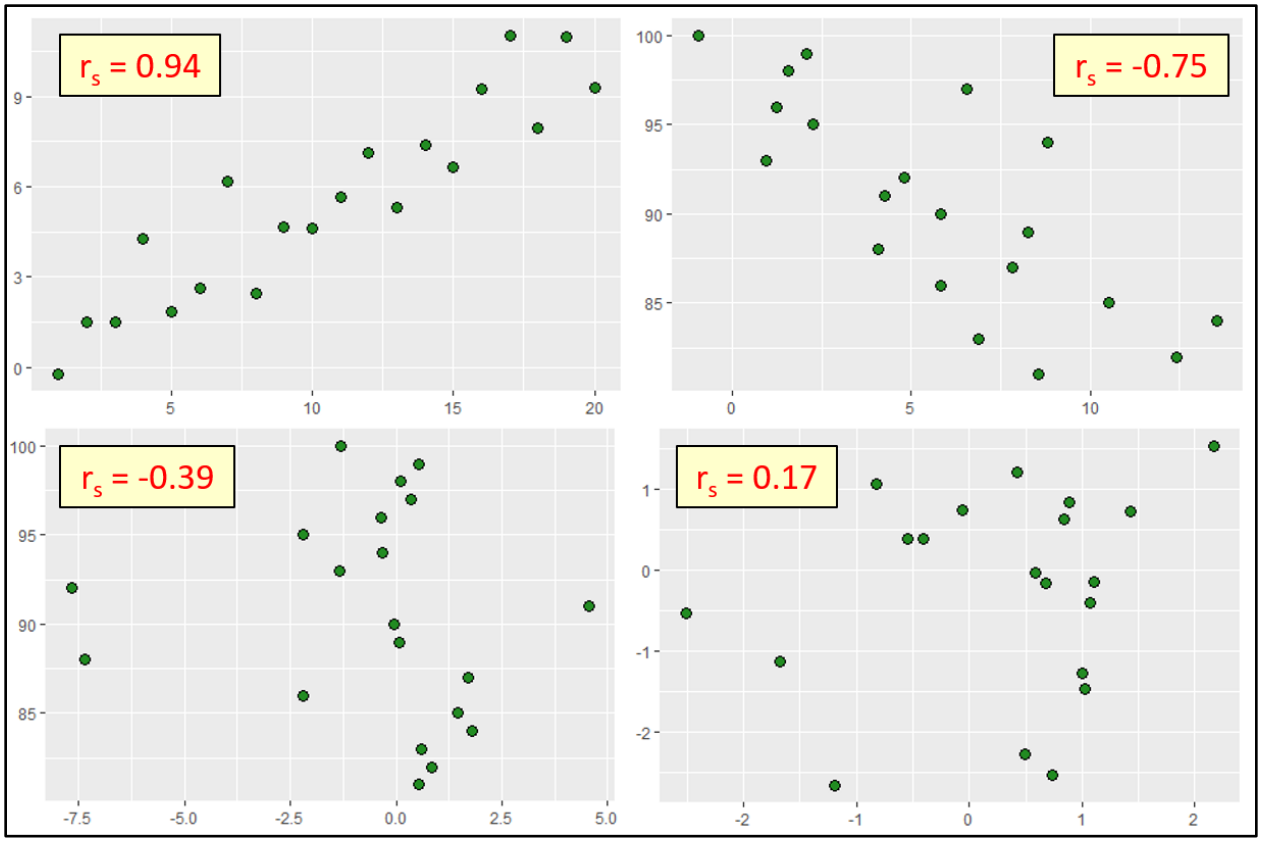

Figure 13. Spearman correlation coefficients for different types of paired data relationships.

CHAPTER 4: RESULTS AND DISCUSSION

This chapter starts by focusing on the entire FMPO region and looking at the need for transit service in the area. The transit users' index is used to compare the need for transit in the area covered by the transit system against the area not served by public transit. Then, the geoprocessing tool developed to automate the accessibility calculation is presented and the workflow associated with its functionality is described. The next section concentrates on the temporal component of accessibility, and shows how the average accessibility levels for the area covered by transit fluctuates over the course of a day due to differences in the bus service frequency. Finally, the spatio-temporal variations in accessibility are addressed and the results are correlated with the socioeconomic attributes that define the need for public transit.

4.1 - ANALYZING THE NEED FOR TRANSIT SERVICE

Since this study addresses transit-based accessibility, it mainly focuses on the area covered by the public transit system. This area is defined by the census blocks whose centroids are located within walking distance from a bus stop. However, one of the objectives of this research is to also identify areas with a relatively high demand for transit that are not currently served by the system. Therefore, the transit users' index was calculated for the entire FMPO region (Figure 14). The FMPO region includes the City of Flagstaff and a number of surrounding unincorporated communities, with a total population of approximately 85,000 people (FMPO 2017). This calculation used data from the 2010 census blocks and 2016 American Community Survey (ACS) estimates at the block group level (USCB 2017a). According to this dataset, in the FMPO region, 59,467 people are served by the transit system, while 24,555 are not.

The attribute fields with the proportions of low income households (income less than $20,000/year), individuals with a disability, non-driving age individuals (younger than 19 and

older than 65 years old), and households with no private vehicles available were joined from the census block groups to the respective census blocks. For each census block, the transit users' index was obtained by averaging these proportions and then multiplying by the total population. In the map presented in Figure 14, darker tones represent areas with a higher need for transit service, while the blue polygon defines the area served by the public transit system.

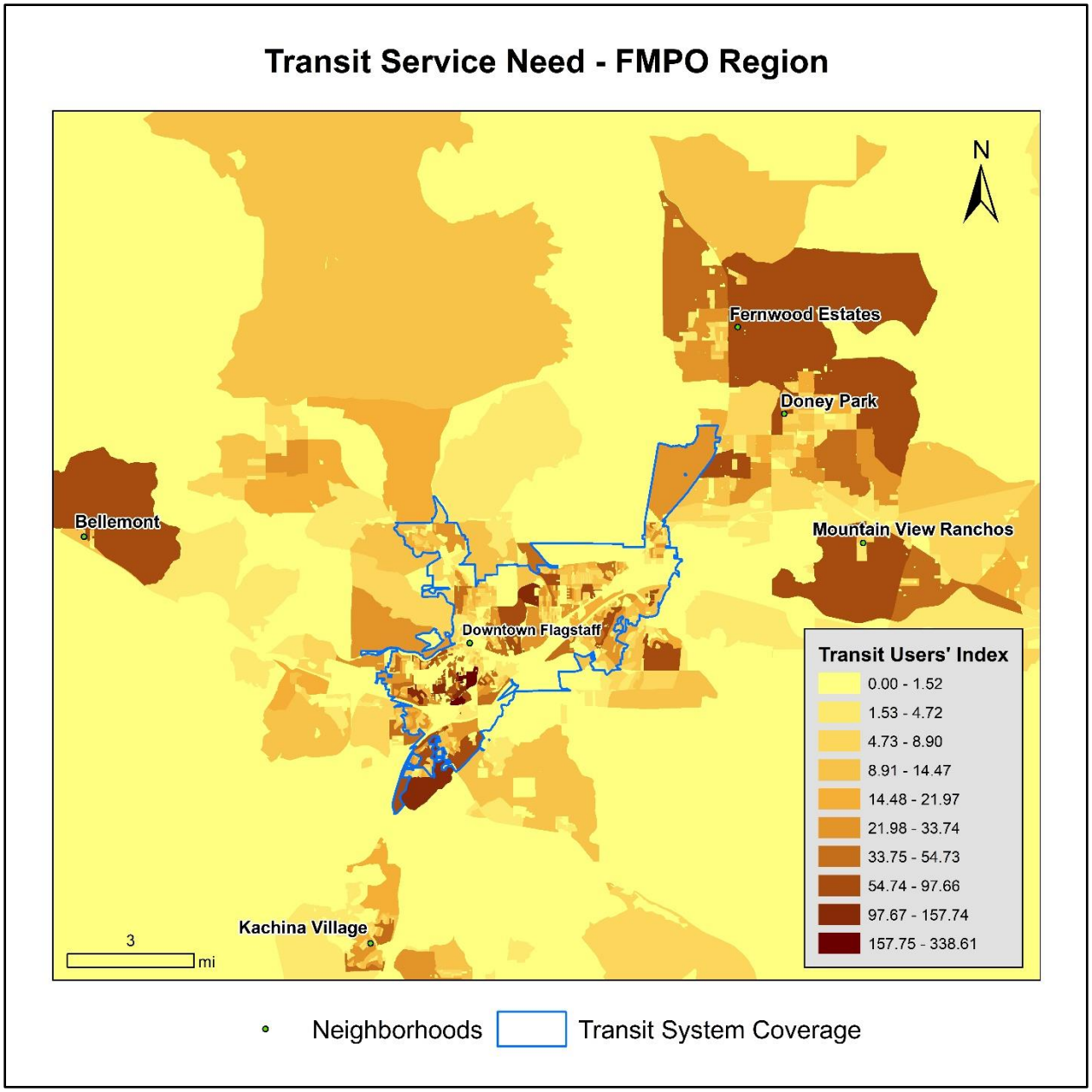

Figure 14. Map showing the transit users' index for the FMPO region.

Important information can be extracted from this map (Figure 14). The results match one of the spatial gaps identified by the transportation plan (FMPO 2017): no transit service connects Doney Park, Kachina Village, Mountainaire, and Bellemont to Flagstaff for work journeys. The relatively high value for the index in these areas suggest that there is a significant portion of the population that could benefit from the availability of public transit service. In addition to these neighborhoods, visual inspection of the map also indicates that a couple of other areas deserve attention, such as Fernwood Estates and Mountain View Ranchos.

Census-based polygons can be misleading because their extent covers much more area than the extent where people actually live. Large polygons with a high value for the transit users' index can give the wrong impression about where people with a need for transit live. For ease of communication and interpretation by stakeholders and the general public, the index based on the census boundaries was summarized for the neighborhood polygons (Figure 15). A spatial overlay procedure was performed for that purpose. The census block polygons were clipped to the neighborhood boundaries, and a spatial join was performed to obtain the average index within each neighborhood. Table 1 presents the average index for the neighborhoods covered and not covered by transit.

Table 1. Average transit user's index for both the areas served and the areas not served by the transit system.

Average Transit Users' Index	
Area served by the transit system	14.95
Area not served by the transit system	5.52

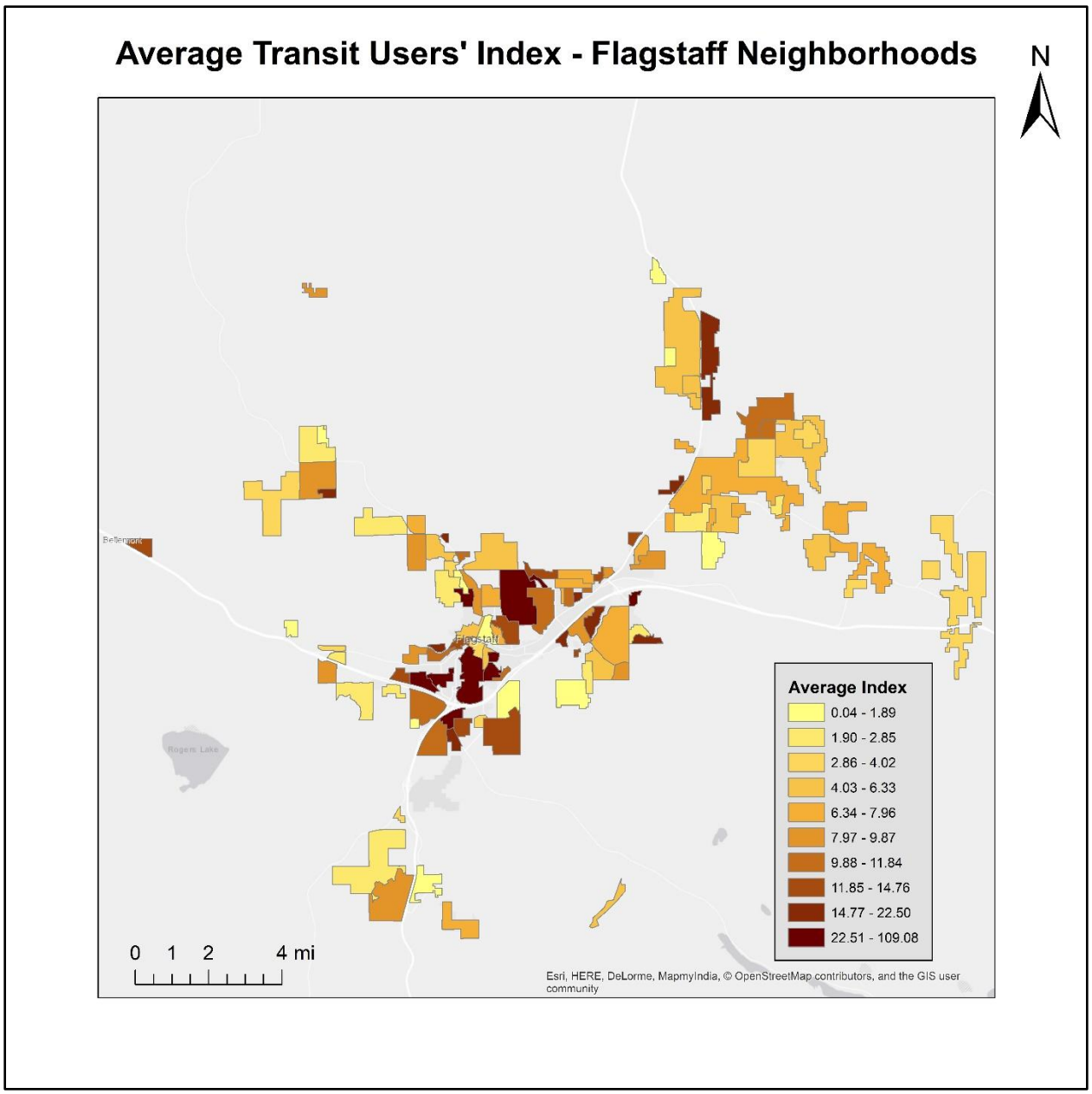

Figure 15. Map showing the average value for the transit users' index by neighborhood.

As shown in Table 1, it would be expected that that the need for transit is, on average, higher in the area served by the transit system than in the area not served by the transit system. A total of 57 neighborhoods have transit service, while 54 do not. A t-test on the difference in means for the two categories yields a p-value = 0.00006007, showing that the difference is statistically significant at any reasonable confidence level. It would be useful to compare those average values

to the average index obtained for the neighborhoods currently not served by the transit system to see if some of them have a higher need for transit than the ones within the served area. Table 2 shows the average index for some of the neighborhoods outside the transit coverage area in descending order.

Table 2. Average transit users' index for the neighborhoods currently not served by the public transit system.

Average Transit Users' Index	
Fernwood Estates	18.73
Fort Valley Estates	17.96
Walnut Meadows	17.34
Flagstaff Meadows (Bellemont)	14.76
Sinagua Heights	12.60
Pioneer Valley	11.64
Sunset Crater Estates	11.11
Kachina Village	9.87
Baderville	9.83
Snowbowl Estates	9.32
Amberwood	9.22
Westwood Estates	8.87
Macann Estates	7.90
Doney Park	6.73
Mountainaire	6.56

Table 2 shows that some of the neighborhoods (Fernwood Estates, Fort Valley Estates, and Walnut Meadows) have an average index that is higher than the mean index for the areas covered by transit. Bellemont has an average value approximately equal to the mean of the covered areas. It is also likely that some of these neighborhoods have higher averages than some of the transit-covered ones. This is important information that could be used to prioritize the neighborhoods in

an eventual expansion of the current transit service. Although the results obtained are generalizing and do not cover all the specific variables that determine the need for transit, they can provide important information with respect to the relative need of the different neighborhoods in the city.

Although it would be terrific to expand the transit system coverage so that high quality transit services are available everywhere, public transportation agencies have to comply with a limited budget. For that reason, public transit planning involves balancing coverage and ridership (Walker 2015). Ridership refers to the number of people who actually use the public transit system on a frequent basis, and that is directly related to the quality of the service. With a limited budget, the expansion of the system's spatial coverage may come at the expense of service quality.

For instance, consider the hypothetical situation where NAIPTA decides to create additional bus routes to connect Fernwood Estates, Bellemont, and Kachina Village to the current transit network. Assuming that the budget remained constant, the redistribution of resources that occurs with the creation of those routes would probably decrease the service frequency in other areas. When service frequency decreases: the system becomes less flexible, where waiting times are longer and it is harder to make connections between different routes.

Suppose that someone has to be at work at 9:00 AM. With the new bus schedule, this person now realizes that it is only possible to get to work either at 8:15 AM or 9:15 AM. The decreased frequency creates a 45 minute wait at the destination. In addition, when transfers are necessary for an individual to reach the desired destination, the waiting times at transfer stops are also likely to increase. As a result, although the new transit routes expand the system's spatial coverage, they might also significantly decrease the service quality, making it harder for individuals to reach their desired destinations. In other words, increasing accessibility in some places compromises the accessibility experienced at other places.

Consequently, the "accessibility budget" is also limited. This is not to say that transportation agencies should focus on providing high quality services at the expense of broad spatial coverage. But unfortunately there are minimum quality requirements that must be met to ensure that the accessibility and ridership levels are reasonable. After all, there is not much use in a system associated with excessively long travel times and low frequencies. With limited budgets, transportation agencies can only do their best to balance the coverage and quality components.

Nevertheless, assuming that NAIPTA eventually finds it necessary to rearrange transit resources to connect those areas to Flagstaff, some measures would be important to ensure that the ridership for these new routes is satisfactory. Using different means to advertise the new service would help in increasing awareness of the routes added to the transit system. This is critical to ensure that the transitional period of low ridership is as short as possible. It would also be important to coordinate with housing initiatives, colleges, universities, and employers to ensure that individuals that rely on transit are informed about the availability of transit service to commute from these neighborhoods to Flagstaff. These efforts and the transitional period of low ridership would also add to the cost of those service changes, so they need to be accounted for in the decision making process associated with these service changes.

There are limitations associated with the characteristics of the datasets used in calculation of this index. Although the problem is partially minimized with the use of age data at the block level, the intra-zone variability within the census block groups is ignored and the MAUP effects are still relevant. However, this generalization was necessary given the unavailability of comprehensive socio-demographic data at the block level. This is especially limiting for studies in smaller cities like Flagstaff. In addition, ACS datasets are based on estimates and cannot be regarded as census datasets, given that the attribute values are obtained from smaller samples.

From a methodological point of view, however, the index overcomes a limitation that was not addressed in the study by Tribby and Zandbergen (2012). The use of a standardized index based on proportions can be misleading because underpopulated areas may be assigned to relatively high values. This problem can be minimized by multiplying the proportions by the total population or the target population.

Consider the hypothetical situation where 100 people live in block group A, while 1000 people live in block group B (Figure 16). Assume that A and B have the same area, and that 80% of the individuals in A and 30% of the individuals in B have a high need for transit. With the proportion-based approach, the index for A would be 0.8, while the index for B would be 0.3. Therefore, A would have higher need for transit than B. However, there are 300 individuals in B with a social need for transit and only 80 in A. From a transportation planning point of view, and considering the ridership variable, block group B is significantly more relevant and therefore should be assigned a higher transit users' index.

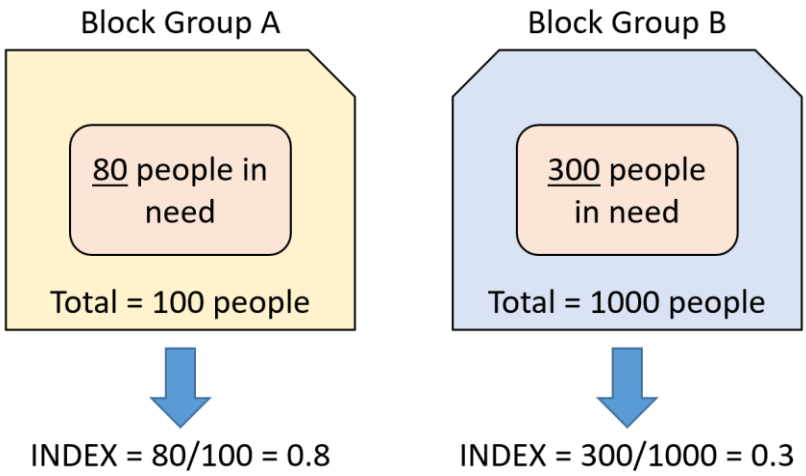

Figure 16. Diagram showing how proportion-based indexes can be misleading.

Another issue is related to the fact that the data at the block level dates back to 2010. Although the overall distribution of the population is not likely to undergo any drastic changes, the fluctuations that occurred in the last 7-8 years were not accounted for. Despite the highlighted

limitations, this is currently the most suitable open data available for this type and scale of analysis. Nevertheless, this framework could be reused in a larger metropolitan area where block groups can more appropriately describe the spatial distribution of the population.

4.2 - SPATIO-TEMPORAL VARIABILITY OF ACCESSIBILITY

4.2.1 - Automating the Accessibility Calculation

A Python script tool named *CalculateAccessibility* (Appendix A) was developed to automate the calculation of accessibility in 5-minute intervals. Figure 18 includes a simplified flowchart that describes the execution path of the tool. A general description of this geoprocessing workflow will follow. This tool requires five input parameters (Figure 17):

1) **OD Cost Matrix Layer:** the OD cost matrix network analysis layer that references the transit-based network dataset on which the calculation will be based. This preconfigured layer includes the origins and destinations (Figure 19), the travel impedance used in the calculation (travel time with transit), and the default cutoff value (60 minutes). In addition to the day of the week that is defined in the beginning of the script, the only thing that the tool changes in this layer along the execution path is the time of the day every time it solves the network analysis problem.
2) **Output Geodatabase:** the output geodatabase where the output line feature classes will be saved. Every time the OD cost matrix layer is solved, it outputs a line feature class that contains the travel times from each origin to each destination.
3) **Weekend:** a Boolean value (true or false) that determines whether the analysis will be performed for a weekday or a weekend day. If it is set to true, the tool will perform the accessibility calculation based on the weekend schedule. The default value is false.

4) **Census Blocks Jobs:** point feature class with the census blocks centroids and attribute fields for the employment data. This feature class contains a field with a unique key identifying each census block and another field for the number of jobs.

5) **Census Blocks Table:** blank table that only includes an attribute field with a unique key that identifies each census block. The accessibility measurements for each census block at each time of the day will be recorded in this table.

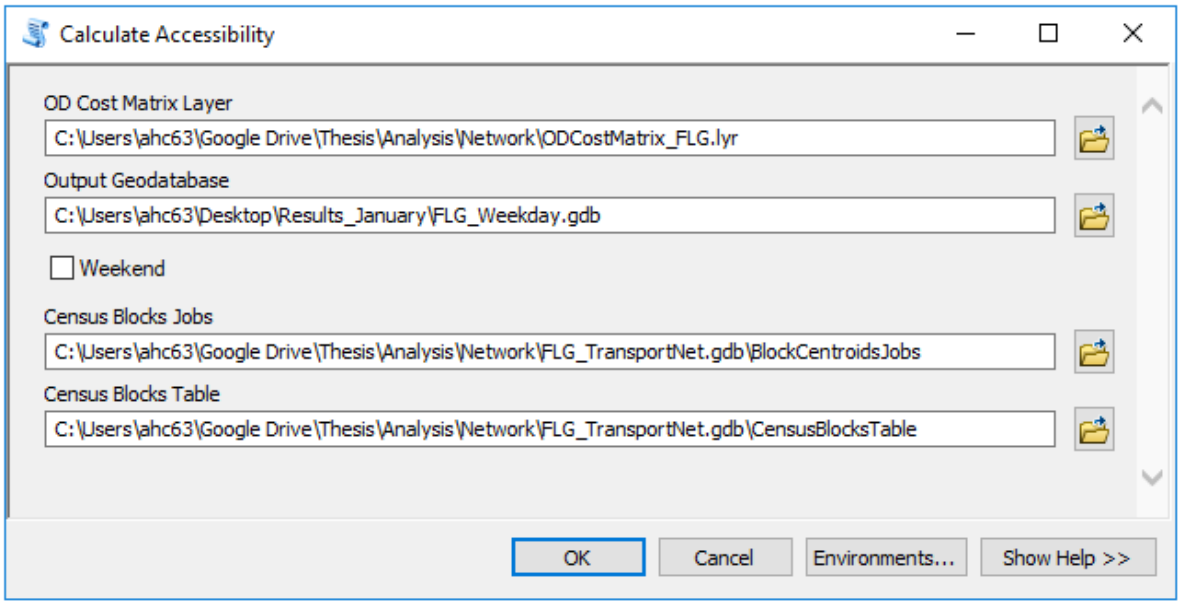

Figure 17. *CalculateAccessibility* tool's dialog box and required input parameters.

The tool starts by defining the day of the week that the analysis will be performed for (Monday for weekday, and Saturday for the weekend), depending on the Boolean value used as an input. In either case, the model runs from 3:20 AM to 11:55 PM in 5-minute intervals. This is to make sure that it captures both the time when the first transit trip is possible, and the time when the last transit trip is available. This time frame was decided based on the transit service's start and end times, obtained from the transit timetable.

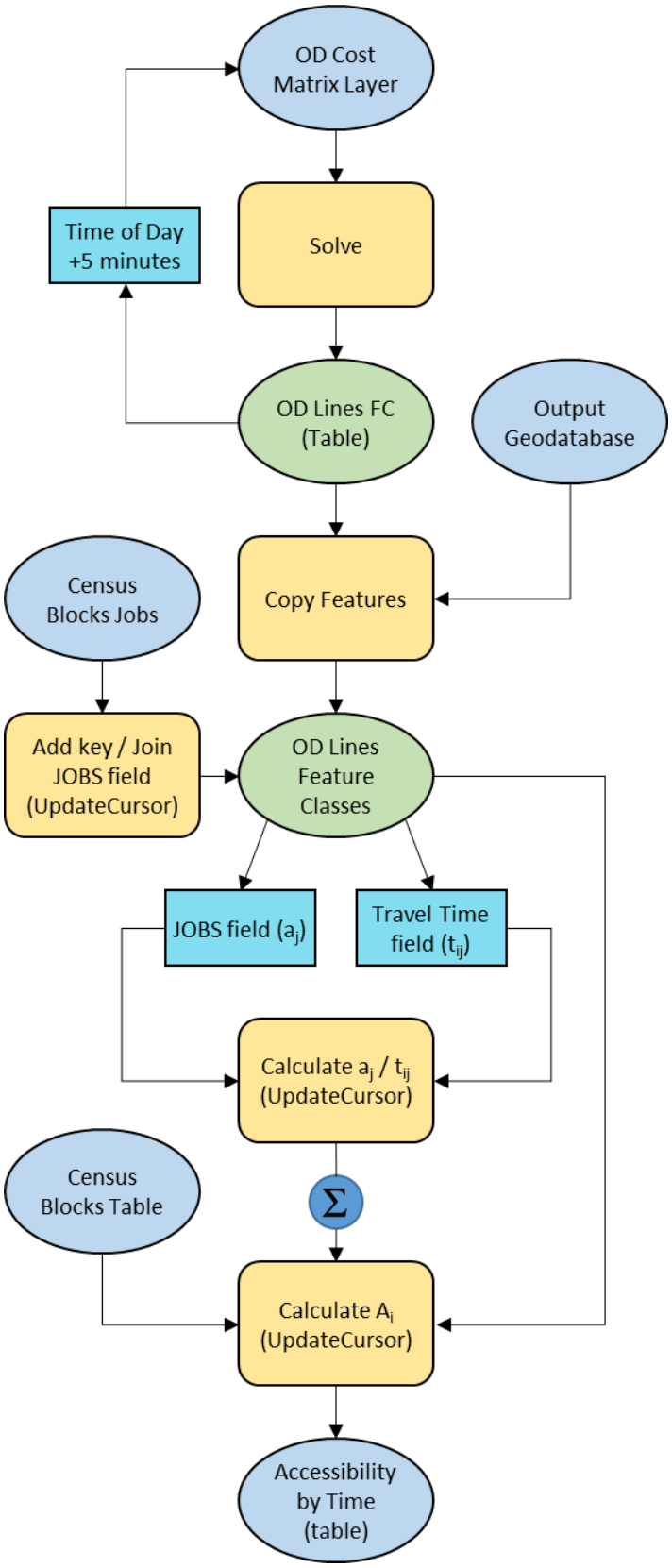

Figure 18. Flowchart describing the execution path of the *CalculateAccessibility* Python script.

Then, the OD cost matrix layer is solved, the resulting line feature class is copied to the output geodatabase, and 5 minutes are added to the parameter that defines the time of the day for the analysis. This process is repeated again and again in a loop, until the end time previously defined is reached. At the end of this step, the output geodatabase contains a line feature class for each time that the network analysis problem was solved. A total of 247 feature classes are generated each time that the tool runs.

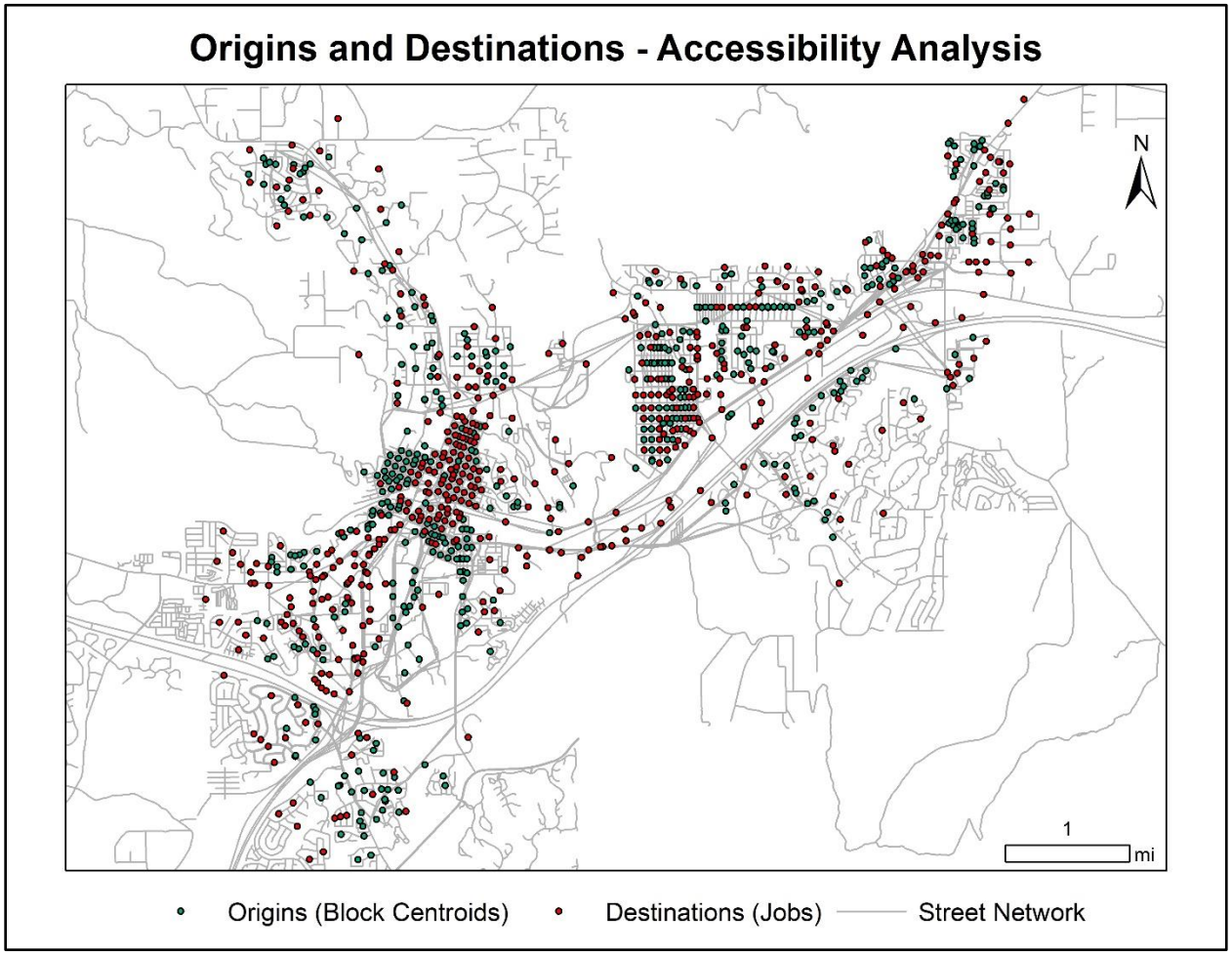

Figure 19. Origins and destinations used in the accessibility analysis.

The next step was to separate the fields with the unique IDs that distinguish the specific origins and destinations for each feature class in the output geodatabase. Once the destination IDs are obtained, the field with the number of jobs can be joined to the feature classes. At this point,

each record in each table (feature class) represents a trip from an origin to a destination and contains attributes for the travel time with transit (t_{ij}) and the number of jobs (attractiveness) in the destination (a_j). The number of jobs is then divided by the travel impedance and the result (a_j / t_{ij}) is recorded in a new field (Figure 20).

The final step involved calculating the accessibility value for each one of the origins. Each feature class corresponds to a time of the day and includes trips from multiple origins to multiple destinations. The procedure for calculating accessibility involved organizing the trips according to the individual origins, and then adding the (a_j / t_{ij}) values together. The sum of these quotients is the accessibility measurement at a given time. The results are recorded in a table, with the rows representing the origin locations, and the columns the different times of the day (Figure 21).

Matrix0800

OBJECTID	FromTo	OriginID	DestinationID	TravelTime	Jobs	aj_tij
2	040050001001000 - 040050001001014	40050001001000	40050001001014	9.83	5	0.51
361	040050001001006 - 040050002001016	40050001001006	40050002001016	38.27	137	3.58
1119	040050001001012 - 040050012001050	40050001001012	40050012001050	57.49	3	0.05
1483	040050001001015 - 040050002002016	40050001001015	40050002002016	59.24	152	2.57
1522	040050001001016 - 040050002002030	40050001001016	40050002002030	47.4	974	20.55
1809	040050001001018 - 040050012001050	40050001001018	40050012001050	57.49	3	0.05
2289	040050001002000 - 040050011023010	40050001002000	40050011023010	31.52	5	0.16
2724	040050001002001 - 040050012002040	40050001002001	40050012002040	31.39	82	2.61
3379	040050001002002 - 040050008003056	40050001002002	40050008003056	54.19	162	2.99
3520	040050001002003 - 040050012001050	40050001002003	40050012001050	31.9	3	0.09

Figure 20. Attribute table for the feature classes that result from solving the OD cost matrix layer.

CensusBlocksWeekday

OBJECTID *	GEOID10	Ai0800	Ai0805	Ai0810	Ai0815	Ai0820	Ai0825
1	040050009003039	589.159565	421.146542	555.372053	655.307735	804.547489	397.590916
2	040050005002015	643.423008	326.445758	375.777812	400.661116	476.992925	553.669757
3	040050005004010	642.427612	619.384339	414.382135	490.983323	566.498799	635.312374
4	040050005004017	659.868245	640.977501	730.759332	509.599698	588.712111	657.191364
5	040050005004013	657.230025	699.669673	711.455317	511.419776	584.826712	652.527873
6	040050004005012	630.257855	680.976918	703.185682	812.538075	572.582897	637.039856
7	040050008003105	797.96934	856.418342	945.782262	807.568123	835.352047	898.303691
8	040050001002015	714.099758	713.074687	728.098235	769.187239	886.331787	699.514837
9	040050012001050	1148.445315	1227.359733	1087.281738	1155.933303	1073.558199	1098.442167
10	040050012002068	1237.922426	1310.034164	1143.117614	1131.325896	1153.20694	1292.8916

Figure 21. Table that summarizes the accessibility measurements for origin locations at different times of the day.

4.2.2 - Temporal Variability Analysis

The temporal variability of accessibility can be analyzed once measurements are obtained for all origin locations at different times of a day. As already mentioned, the accessibility levels fluctuate over the course of a day due to transit service frequency. To illustrate that, consider the 20-minute service areas obtained for the SBS West building at NAU (Figure 22). Service area is the extent that encloses all the locations that are within a specific travel impedance (time or distance) from the origin location.

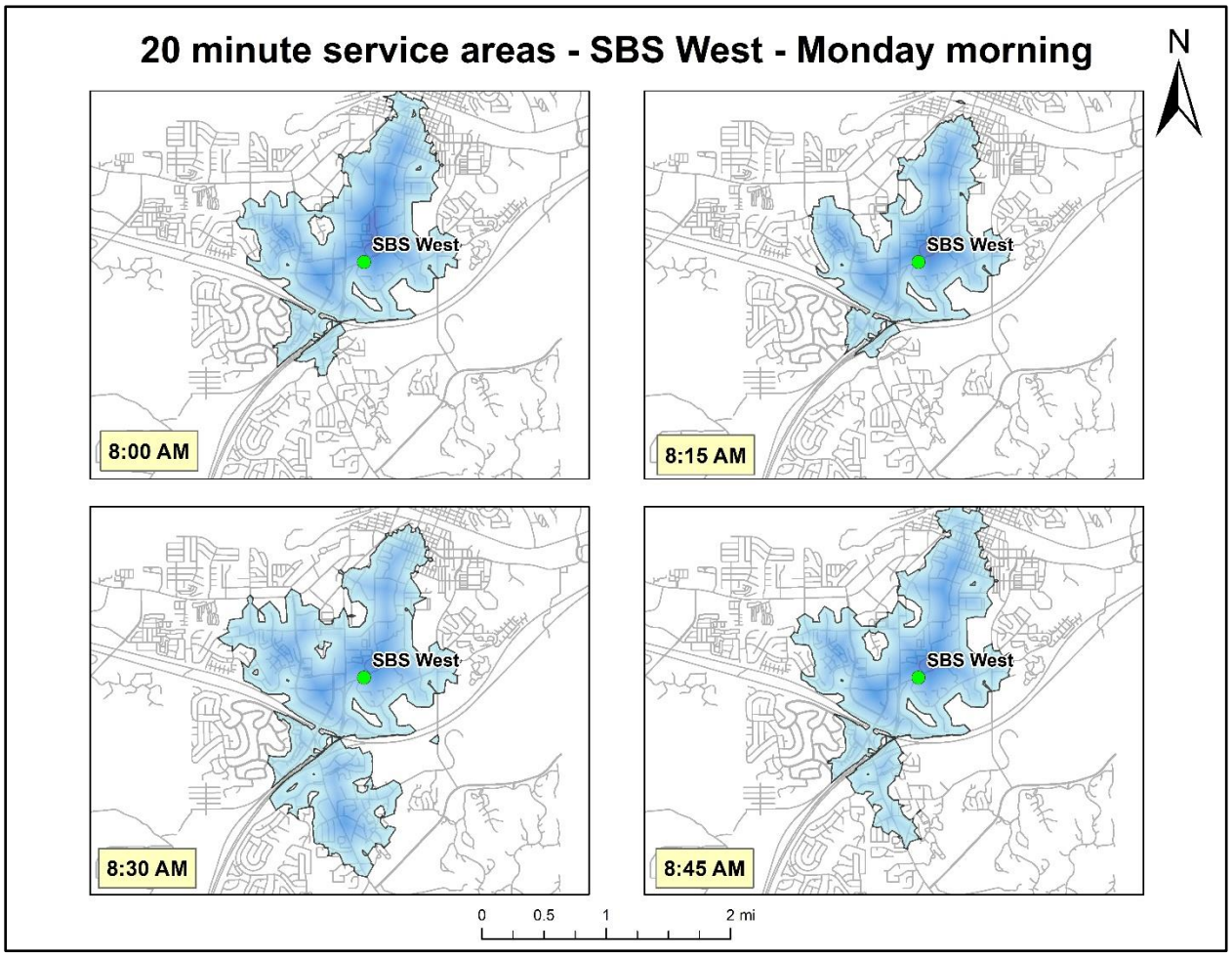

Figure 22. Map showing the 20-minute service areas for the SBS West building for four different times in a Monday morning.

As demonstrated in the map above (Figure 22), the areal extent in the city that can be reached in 20 minutes or less by transit depends on the time of the day. At 8:30 AM the service area is significantly larger compared to the service area at 8:15 AM, for instance. Consequently, the accessibility measurement at 8:30 AM is also significantly higher because it is easier to travel to a higher amount of destinations at this time. Hence, this temporal variability is an important factor to be considered in accessibility measurements.

The first step in analyzing the temporal variability of accessibility was to look at how the overall accessibility level fluctuates over the course of the day. For that reason, the accessibility measurements were summarized for each time of the day. In other words, the mean of each column in the output tables (Figure 21) was obtained and the results were plotted in time series graphs (Figure 23). The data analysis and the construction of graphs were carried out in R programming.

One of the objectives of this step was to find out when the first and last trip with transit is possible. An issue is that the NAU Campus Shuttle has buses running until 11:30 PM during the week, and a late night service (SafeRide) from 9:00 PM to 2:00 AM during the weekend. These services are still operating after the last Mountain Line bus stops running. However, it is unlikely that these late night services within the NAU community are used for work trips. Besides, these services are highly restricted to the NAU campus boundaries and most frequently used by students at these times. Therefore, the NAU transit services were excluded from the network dataset used in this step, and the time series graphs are solely based on NAIPTA's GTFS feed.

A great deal of relevant information can be obtained from the graphs below (Figure 23). The blue curve represents the overall accessibility level and shows how accessibility varies over time. By visual inspection, it is possible to notice that the accessibility levels are generally higher during the week (mean = 1118.95) compared to the levels experienced on the weekends (mean =

990.43). At around 6:00 PM, accessibility starts to gradually decrease in both situations as a result of the decline in transit service frequencies. In addition, as demonstrated by the difference between peaks and valleys in the curve, the oscillations are more abrupt over the weekend as a result of lower service frequencies.

The blue curve in Figure 23 becomes horizontal at the extremes (early morning and late night) of the graphs, showing that accessibility is constant for those periods. Those regions in the graph correspond to periods where trips with transit are not available. The accessibility measurements have a value of approximately 825 at these intervals because the model computes accessibility values based on walking times when transit trips are not available. These baseline features allow the detection of the time window where transit is available.

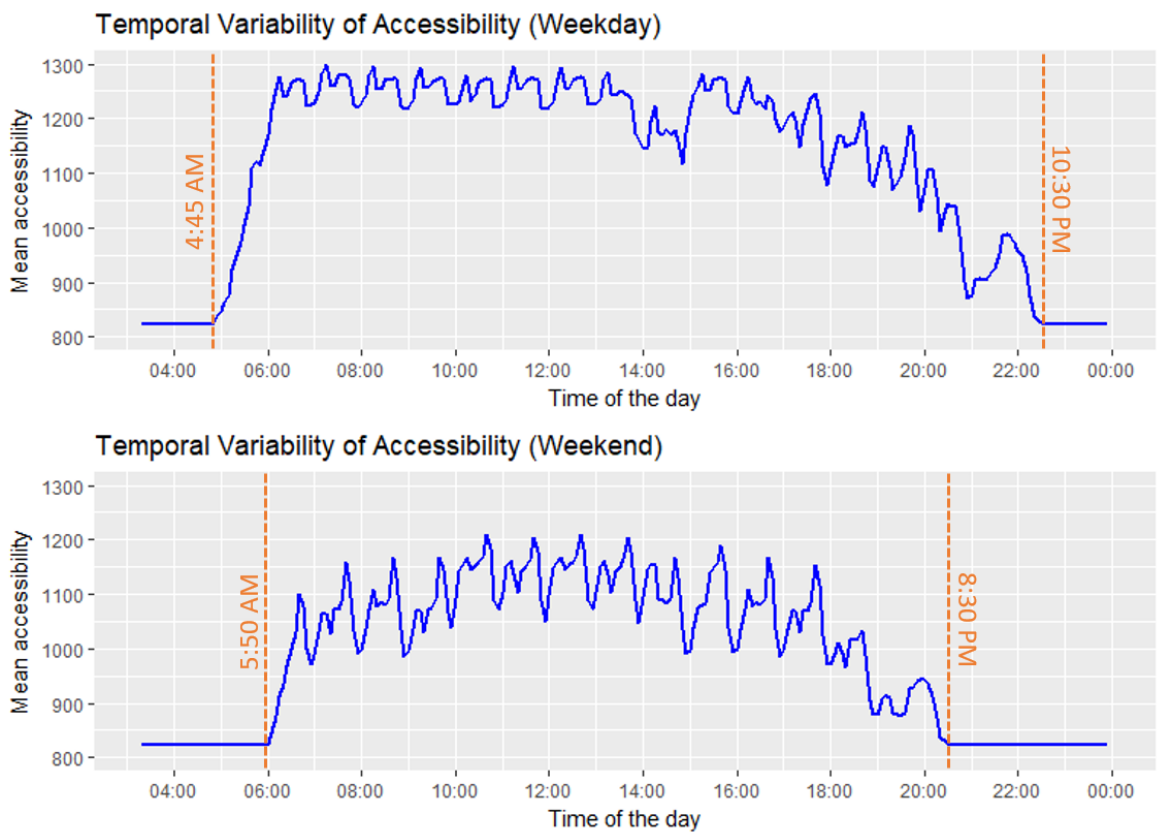

Figure 23. Time series graphs showing the temporal variability of accessibility over the course of a typical weekday (above) and a weekend day (below).

The vertical orange dashed lines define the time periods where it is possible to travel with transit. Transit trips are only available if an individual leaves home between 4:45 AM and 10:30 PM during the week, and between 5:50 AM and 8:30 PM on the weekends. The FMPO transportation plan mentions the existence of a temporal gap in the current service, where many workers need extended hours to access their jobs (FMPO 2017). The 2016 ACS data at the block group level contains information about the time individuals leave home to go to work. That data can be combined with the transit time window information so that the number of workers who leave home to go to work when transit is not available can be estimated and mapped (Figure 24).

Two attribute fields are relevant for this step: the number of workers who leave home between 4:00 PM and 11:59PM, and the number of workers who leave between 12:00 AM and 4:59 AM. Although it is possible to travel with transit to work and back if someone leaves home around 4:00 PM (assuming a maximum of 5-6 hours spent at work), it is unlikely that individuals working late and early hours are going to use the bus to go to work and back home. Therefore, despite the margin for error, the individuals leaving home to work between 4:00 PM and 4:59 AM were assumed to be the workers who would potentially need extended transit service hours for work trips. A spatial join was performed to summarize the block group data into the neighborhood polygons.

The maps in Figure 24 show a concentration of workers with a potential need for extended transit hours in the neighborhoods that surround the NAU campus, possibly due to the amount of students living in the area with work schedules after the school hours. These maps and the associated attribute information are again valuable information that can be used guide the decision making process that relates to the public transit system. In an eventual extension of the current

service hours, neighborhoods can be prioritized according to the amount of individuals that are assumed to have a higher need for the extended schedule.

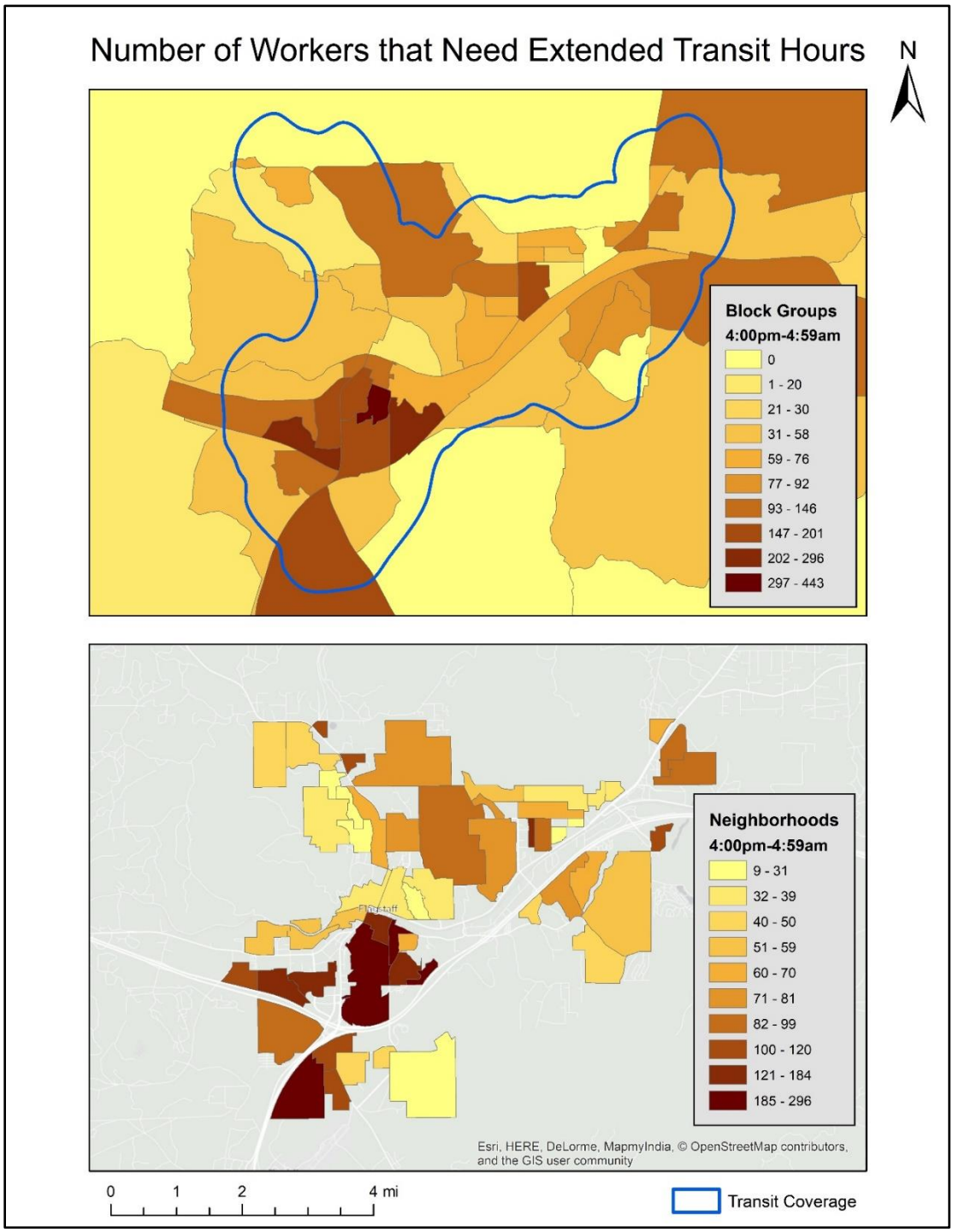

Figure 24. Map with the number of workers in each census block group that leave home to work from 4:00 PM to 4:59 AM (top), and the average for each neighborhood (bottom).

4.2.3 - Spatio-Temporal Variability Analysis

The previous procedure evaluated the temporal variability of accessibility in an overall manner, and represents the average for the entire area covered by the transit system. The following process summarizes the accessibility measurements for each origin location (census block), with the results representing a temporally-enabled measurement of accessibility over space. This time, the mean and the coefficient of variation (CV) of each row (census block) in the output table (Figure 21) were obtained, and the results were displayed in a few maps (Figures 25, 26, and 27).

As previously mentioned in the methodology chapter, the mean provides an assessment of the overall accessibility level experienced, while the CV measures the relative variability of accessibility at each location. The CV used here is defined as the standard deviation divided by the mean and multiplied by 100. Therefore, it is the standard deviation measured in percentages of the mean. The maps below show the mean and the variability of the accessibility measurements in three different situations: during weekdays (Figure 25), during weekends (Figure 26), and the weighted average that represents overall accessibility (Figure 27).

Although the weekday accessibility levels are higher and show less variability due to higher service frequency, the spatial pattern observed for a weekend day is similar. Consequently, this accessibility pattern is also observed in the overall accessibility map. Higher accessibility levels are experienced in the downtown area and surroundings, along Milton Road and the NAU campus, and also in the east side of Flagstaff (green areas in Figure 27). As you move away from these areas and approach the outer edge of the transit system coverage, accessibility decreases (areas in yellow, transitioning into orange and red).

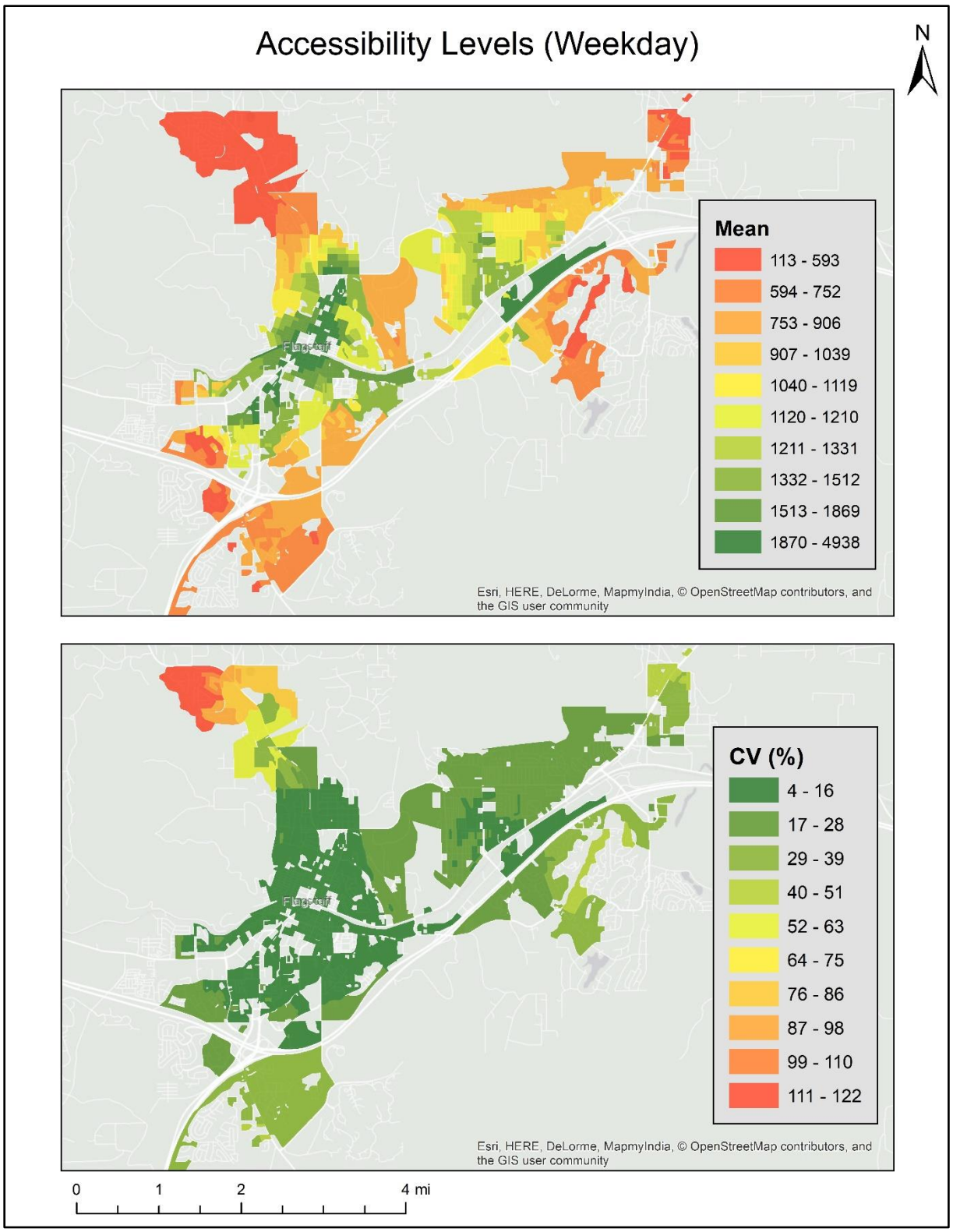

Figure 25. Maps showing the weekday mean accessibility levels (top) and the variability of accessibility (bottom) for each census block in the city.

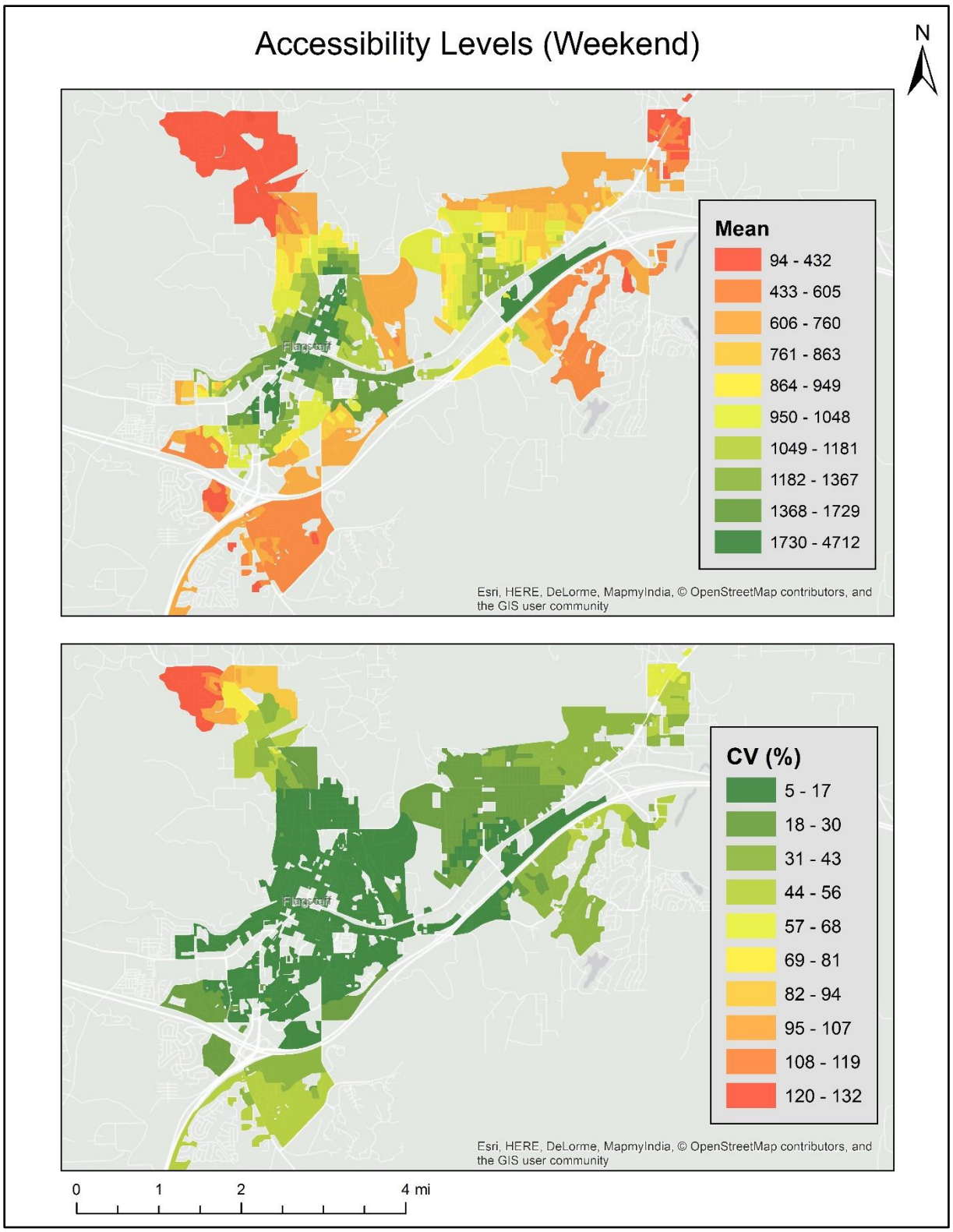

Figure 26. Maps showing the weekend mean accessibility levels (top) and the variability of accessibility (bottom) for each census block in the city.

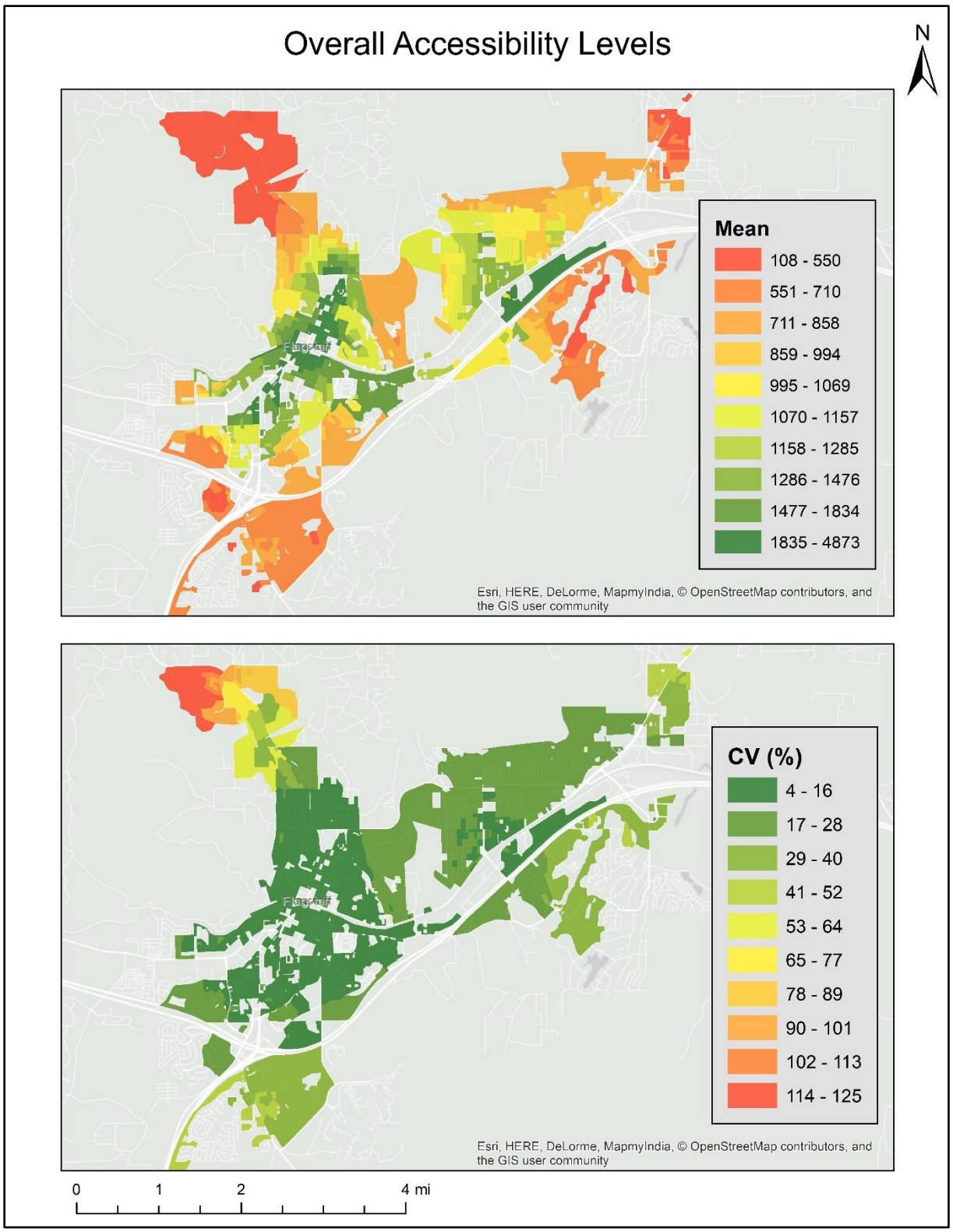

Figure 27. Maps showing the overall (weighted average) mean accessibility levels (top) and the variability of accessibility (bottom) for each census block in the city.

Two elements control the accessibility pattern observed in the origin locations: the spatio-temporal configuration of the transit network, and the spatial distribution of jobs (Figure 28). The transit network is centralized in the Downtown Connection Center (green dot in Figure 28), therefore the maximum access to the different branches of the transit network is experienced in this location. The distribution of jobs is less centralized, but most of the employment opportunities are concentrated in the downtown area, along Milton Road, within NAU, and along Route 66. The interaction between the origin census block locations, the transit network, and the spatial distribution of jobs determines the spatial distribution of accessibility observed.

All the bus routes operated by NAIPTA have a stop in the Downtown Connection Center. An individual departing from the downtown area has the advantage of easy access to the entire Mountain Line network without any transfers. The transit network is also relatively denser along Milton Road and within the NAU campus, with several choices of routes. The NAU Campus Shuttle operates within the university campus, providing enhanced mobility in the area. In addition, a significant amount of jobs exist in those areas, including NAU as the largest employer in the city. The combined effect of transit and job availability explain the high accessibility values measured in these areas.

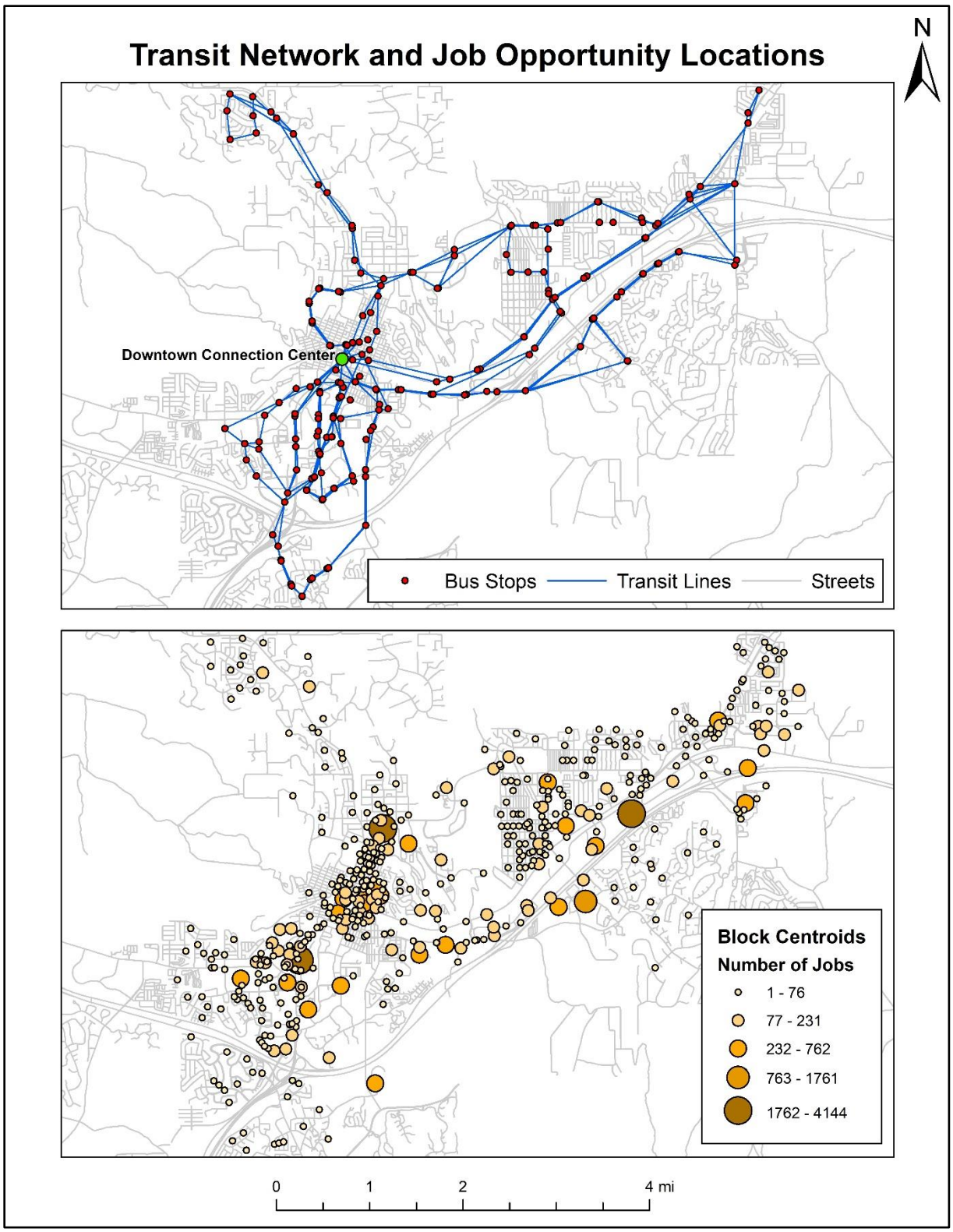

Figure 28. Flagstaff's transit network, including bus stops, transit lines and the street network (top map), and the number of jobs associated with each census block centroid (bottom map).

On the other hand, the relatively higher accessibility experienced in the east side of Flagstaff is less explained by the transit system itself and mostly results from job availability in the area. As shown in Figure 28, the transit network in the area is rather sparse compared to the downtown area. However, several census blocks are associated with an expressive number of jobs (Figure 11). Several major employers are located in the east side of Flagstaff, such as the Walgreens Distribution Center, Gore, the Walmart Supercenter, Nestle Purina, and the Flagstaff Mall businesses. This demonstrates the significant role played by the land-use component, where the proximity to the opportunity locations determines increased accessibility to jobs, rather than the mobility offered by the transit system.

The temporal variability of accessibility, measured by the coefficient of variation, is controlled by transit service frequency. In other words, the CV will be smaller for places where the bus stops very often, and higher for locations where the interval between two buses is long. The maps above show that variability increases as you move away from the areas characterized by high overall accessibility. However, it is possible to notice that the CVs are considerably higher for the census blocks in some of the neighborhoods located north of downtown Flagstaff (red and orange areas in the bottom part of Figure 27).

The variability is especially high for Cheshire, Lynwood, and Valley Crest, where the variability can reach over 100% of the mean measurements. These values contrast with the variability measurements obtained for other peripheral areas, where the highest values in the outer edge of the transit coverage area are in the order of 50%, but generally lower than that. This can be explained by the fact that these areas are only served by a single route (Route 5) with very low service frequency. Route 5, or the orange route, is the only route with a constant 60-minute service frequency both during the week and also on weekends (NAIPTA 2018). Although a few routes

operate on a 60-minute frequency during off-peak times or weekends, most of them operate on a higher frequency at most times. This is the only location in Flagstaff where the bus only stops once every hour.

As discussed in section 4.1, the creation of new routes to connect more distant neighborhoods to Flagstaff is a major change in the transit system that would require more careful considerations. However, in the specific case of these neighborhoods located north of downtown Flagstaff, even minor adjustments in the transit system to increase service frequency in the area would improve the accessibility situation by increasing the mean levels and decreasing the variability experienced. This would make the spatial distribution of accessibility more homogeneous across the study area, and the accessibility levels at these neighborhoods would be similar to the measurements at the other peripheral areas.

As a final step in this part of the analysis, the difference between the weekday and weekend accessibility measurements was analyzed. The difference in means was measured by the percent decrease in accessibility in the weekends, while the difference in variability was calculated by simply subtracting the weekday CVs from the weekend CVs. The result is presented in the map below (Figure 29). The general pattern is that the difference in means and variability increases as the distance from the downtown connection center also increases.

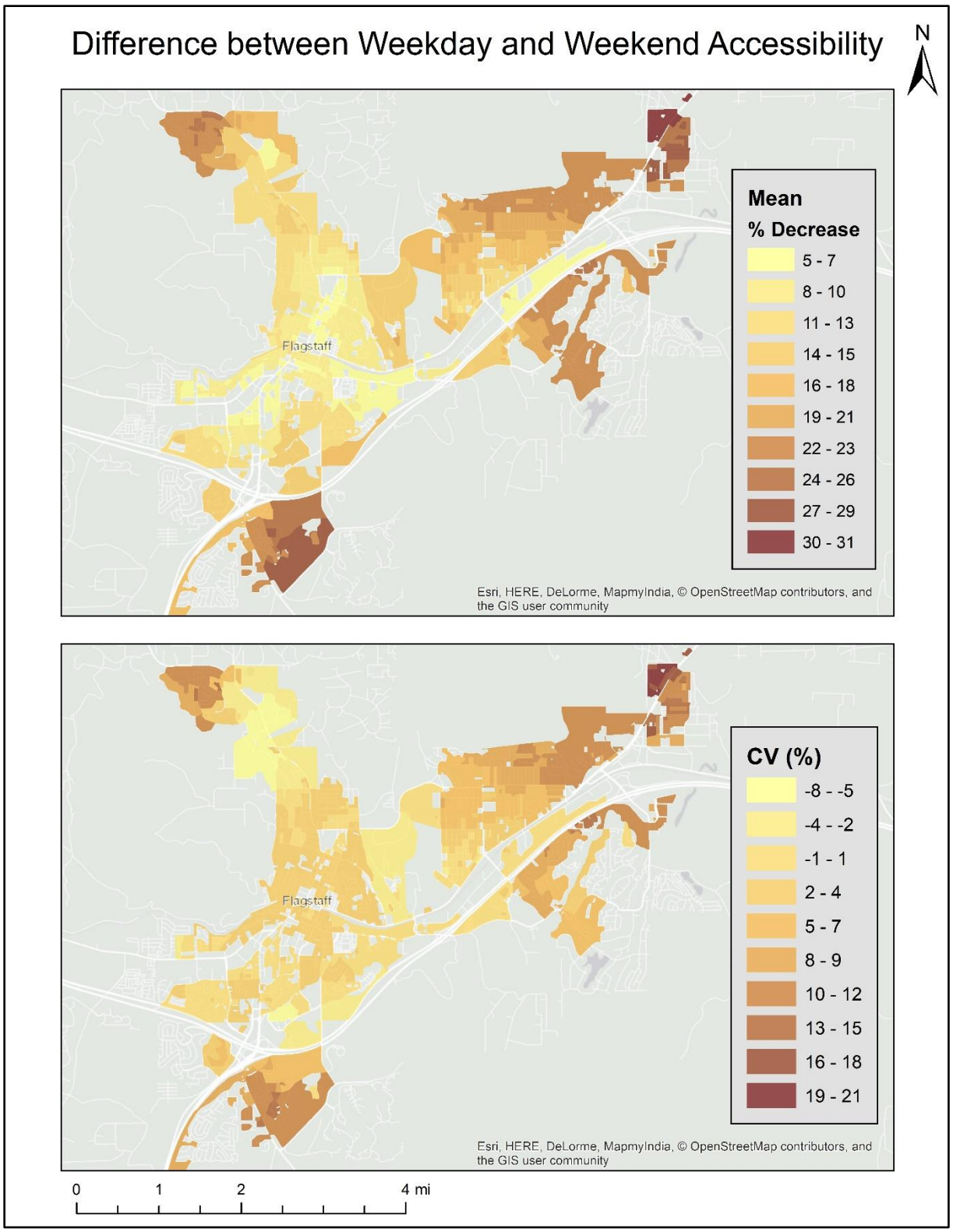

Figure 29. Map showing the difference in means (top) and in variability (bottom) between weekday and weekend accessibility.

The difference in accessibility and in the variability of accessibility between weekdays and weekends is again a result of the different schedules that apply to different routes. The map above (Figure 29) shows that the difference in mean accessibility levels is higher for the east side of Flagstaff. This is explained by the fact that the bus lines (routes 2, 3, 7, and 66) that serve this area operate on a reduced service frequency on weekends (NAIPTA 2018). With the exception of route 2 (blue) that operates on a 30-minute frequency between 10:00 AM and 4:00 PM, all the other lines operate on a reduced 60-minute frequency on weekends. The neighborhoods in the southern area along Lake Mary Road also experience decreased weekend accessibility for similar reasons.

The neighborhoods in the northern part of Flagstaff also experience decreased weekend accessibility. As already mentioned, that area is only covered by a single bus line (route 5) with low service frequency both for weekdays and weekends. Given that service frequency does not change, the difference can be attributed to the decrease in frequencies of the other routes with which the orange line connects.

Regarding the variability in accessibility, the general pattern is that the peripheral locations experience a greater difference in variability, meaning that accessibility fluctuates more during the weekend for those locations. The small number of routes serving those locations may explain this difference. Only one or two routes can be accessed from those areas, meaning that any decrease in transit service frequency has greater effects on the accessibility levels experienced. The more central neighborhoods have more choices of routes, and the decrease in frequency in one of them can be compensated by the existence of other alternatives. Therefore, the neighborhoods located near the center of the transit network are less affected by the decrease in transit service frequency.

In addition, the presence of negative values is intriguing at a first glance. Negative values indicate greater variability during the week compared to the weekend, and that goes against the

fact that the transit service frequency is higher during the week. However, all census blocks experience higher accessibility during the week. Therefore, this can be attributed to the fact that some origin locations have access to a greater amount of opportunities during the week, increasing both the overall accessibility and its variability.

The FMPO states that one of the unmet transportation needs in Flagstaff is that some neighborhoods are not well served by the transit service (FMPO 2017). The maps displayed in this section both map and quantify this spatial gap for work trips. The overall accessibility results not only account for the temporal variations over the course of a day, but they also represent the average transit-based accessibility experienced most of the time in the city. The results are not only useful to identify the underserved areas, but also to measure how underserved they are.

As demonstrated by Jin and Paulsen (2018) in their study, increased job accessibility can potentially decrease unemployment rates and raise household income, improving the lives of disadvantaged population groups. Therefore, the information produced here is relevant for the coordinated transportation plan and should be considered in the current public transportation decision-making process.

Lei and Church (2010) mention that simplifying assumptions are usually made in studies that attempt to model travel times with transit. The accessibility calculations exposed in this section are based on a robust door-to-door approach, where none of the steps of a typical trip with transit are overly simplified. Therefore, this transit-enabled schedule-aware transportation network can be effectively used by transportation agencies as a tool to simulate different situations so that different aspects of their specific transit networks can be evaluated. If necessary, this workflow is flexible enough to allow the use of different origins, destinations, and measurements. It is important to mention that this network dataset not only supports OD cost matrices and accessibility

calculations, but it is also capable of solving any other type of network analysis problem that could be relevant for public transportation planning (Figure 5).

This section focused on providing an assessment of the current transit-based accessibility scenario and its variations over space. The results obtained emerge from the interaction between the land-use (spatial distribution of job opportunities and origin locations), transportation (characteristics of the transit system and street networks), and temporal (variations in transit service frequency over the course of a day) components. The fourth fundamental component of accessibility (Geurs and Ritsema van Eck 2001; Geurs and van Wee 2004), the individual component that distinguishes the level of demand for transit at the origin locations, will be incorporated in the next section.

The following section addresses the correlation between accessibility and the socioeconomic attributes that determine the demand for transit. The idea is to analyze the social aspects of the transit-based job accessibility scenario that was measured.

4.3 - PLOTTING ACCESSIBILITY AGAINST SOCIO-DEMOGRAPHICS

The relationships between the accessibility measurements and the individual socio-demographic variables were measured, and the results are presented in this section. The need for transit and the Title VI requirements will be discussed separately. First, the individual variables will be ranked according to the correlation coefficients obtained, and the relationships will be described in a more general way based on this ranking. Then, the relationships between accessibility and the individual variables will be addressed individually so that some specific features can be explored in more detail.

It is statistically important for comparison purposes that the different variables considered are measured in the same spatial scale and at the same time. Therefore, to ensure that the analysis

was consistent spatially and temporally, the correlations were performed using the 2016 ACS estimates for the census block groups (USCB 2017a). A spatial join was used to summarize the accessibility values at the block level. The accessibility value for each block group corresponds to the mean of the accessibility measurements of the census blocks it contains.

4.3.1 - Accessibility and the Need for Transit

Table 3 contains the Spearman correlation coefficients measured for the spatial relationships between accessibility and the individual variables that determine the need for public transit. The coefficients can be understood as a measurement of the spatial match between accessibility and the need for transit. If the locations with higher number of individuals with a need for transit also experience higher accessibility, the correlation coefficient will assume higher positive values. A negative correlation coefficient represents a situation where the opposite occurs: increasing numbers of individuals with a need for transit experience decreasing accessibility levels.

Table 3. Spearman correlation coefficients for the relationships between accessibility and the individual variables that determine the need for public transit service.

Spearman correlation coefficient (r_s)	
Number of low income households	0.33
Number of unemployed individuals	0.22
Number of households with no vehicle available	0.20
Number of workers who did not work at home	0.11
Number of individuals with a disability	0.05
Number of senior individuals (over 65)	-0.22
Total Population	**0.13**

Figure 30 shows a visual representation of those relationships (bar charts). Mean accessibility values are laid out in the y-axis, while the x-axis contains increasing value ranges for the socioeconomic variables. An approximately increasing pattern is more characteristic of higher

positive correlation coefficients, while decreasing patterns are associated with negative relationships. Weaker relationships show no apparent pattern. The interpretation is similar: increasing patterns reveal a better spatial match between accessibility and the specific socioeconomic variable.

The results suggest that the variable that has the strongest positive relationship with accessibility is the number of low income households (r_s = 0.33), followed by the number of unemployed individuals (r_s = 0.22) and the number of households with no vehicle available (r_s = 0.20), and finally the number of individuals with a disability (r_s = 0.05). The number of senior individuals is negatively correlated (r_s = -0.22) with the accessibility measurements. Accessibility also has a positive relationship (r_s = 0.13) with the total population variable.

As already mentioned in the introductory chapter, a transportation system aligned with horizontal equity would only consider the number of individuals in each location, regardless of their individual socioeconomic characteristics. Therefore, the relationship between accessibility and total population compared to the other relationships can be used to evaluate horizontal and vertical equity.

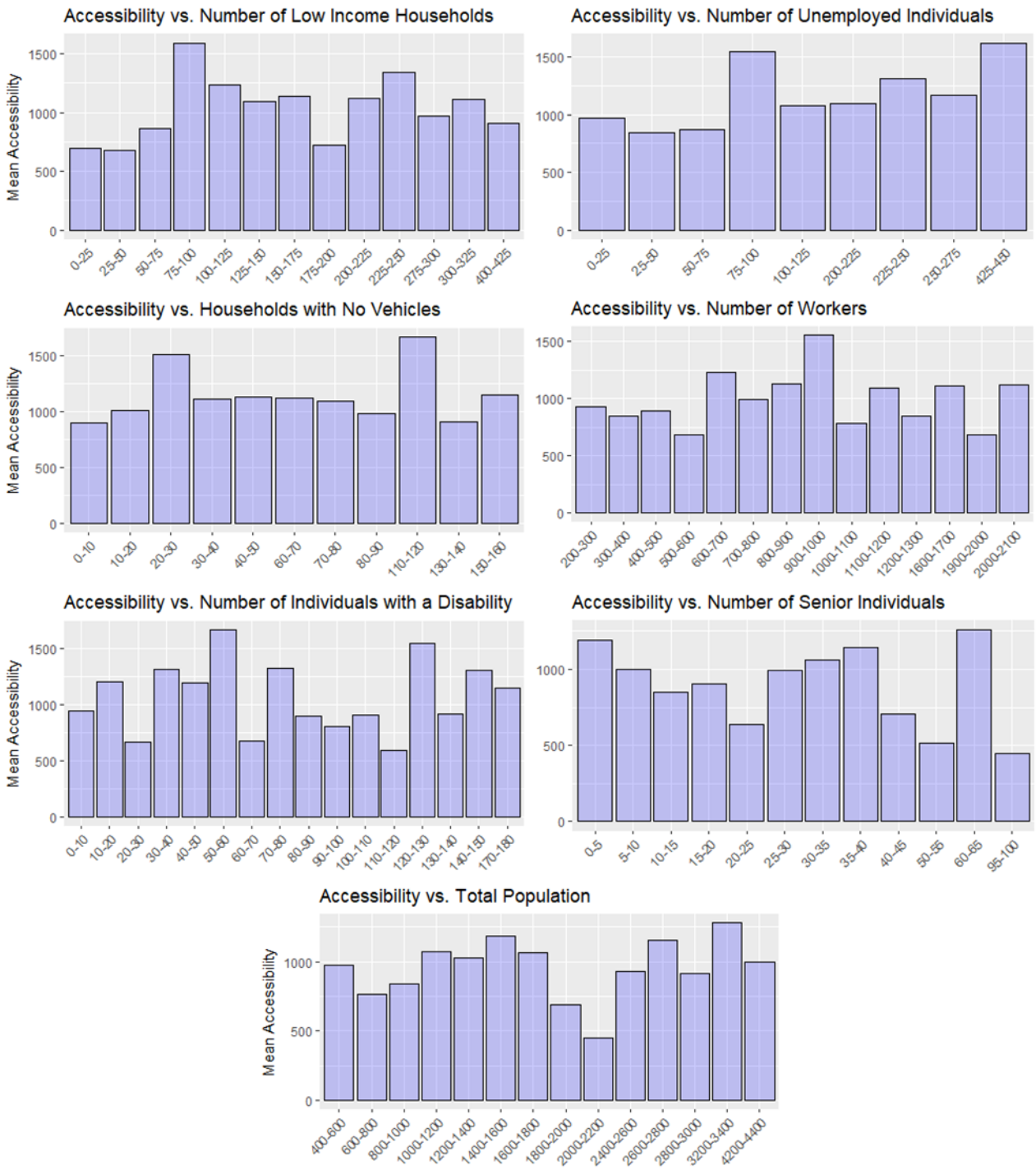

Figure 30. Bar charts showing the relationship between accessibility and the individual variables that determine the need for public transit service.

It is possible to notice the existence of connections between the correlation coefficients and some of the information included in the transportation plan (FMPO 2017). The FMPO lists the

origin locations with a special need for transportation services (Figure 31). These locations include low income housing developments and neighborhoods, and assisted living facilities for seniors. The low income neighborhoods are mostly located in areas that experience high transit-based accessibility, explaining the high correlation coefficient obtained for the income variable.

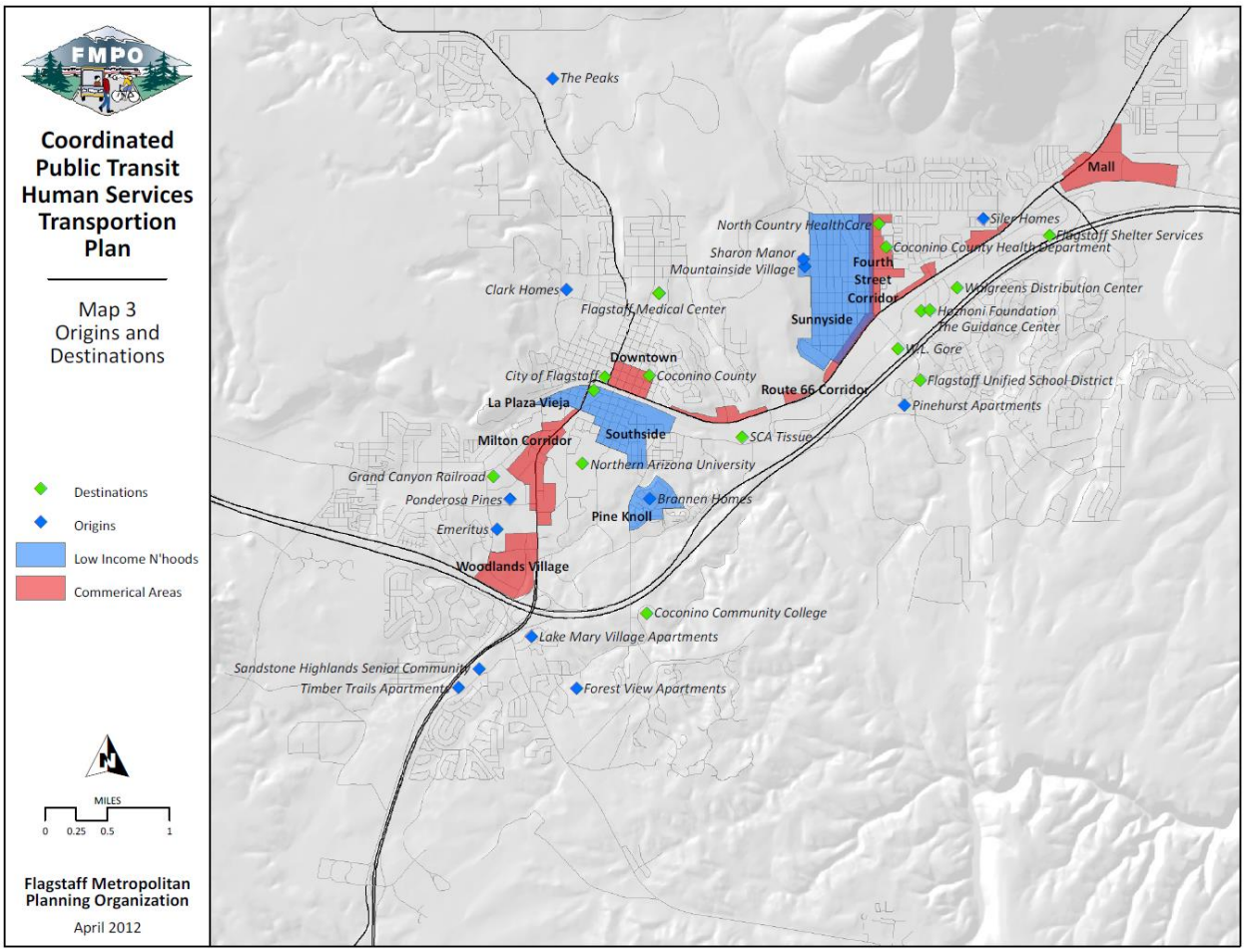

Figure 31. Map with the origins and destinations with a special need for transportation services (FMPO 2017).

The unemployment and car ownership variables may be related to the income variable. The number of unemployed individuals in a household can decrease its income. Low income households are also less likely to be able to afford private vehicles. This may explain why these two variables are also associated with high correlation coefficients. In Figure 32, it is possible to

see that the areas with higher numbers (darker red tones) of low income households, unemployed individuals, and households with no vehicles available better match the high accessibility areas (larger blue circles) compared to the number of senior individuals and the number of individuals with a disability (Figure 33). On the other hand, the individuals with a disability and the senior individuals are less concentrated in the high accessibility areas (Figure 33). The disability variable yields a weak positive relationship, while a negative relationship exists for the number of seniors.

The results collectively suggest that the spatial configuration of the transit system is significantly influenced by the low income neighborhood locations. By comparing the correlation coefficients obtained for the income, unemployment, and car ownership variables with the coefficients obtained for the total population attribute, it is also clear that the transit-based job accessibility scenario is oriented more toward vertical equity than horizontal equity. The origins prioritized by the transit system were defined based on the social need for transit (mainly based on income), meaning that the individuals were weighted based on a socioeconomic characteristic that determines the high need for transit service. If the transit service was solely designed based on the population distribution (horizontal equity), the correlation coefficient obtained for the total population variable would be higher than the coefficient for any other socioeconomic attribute.

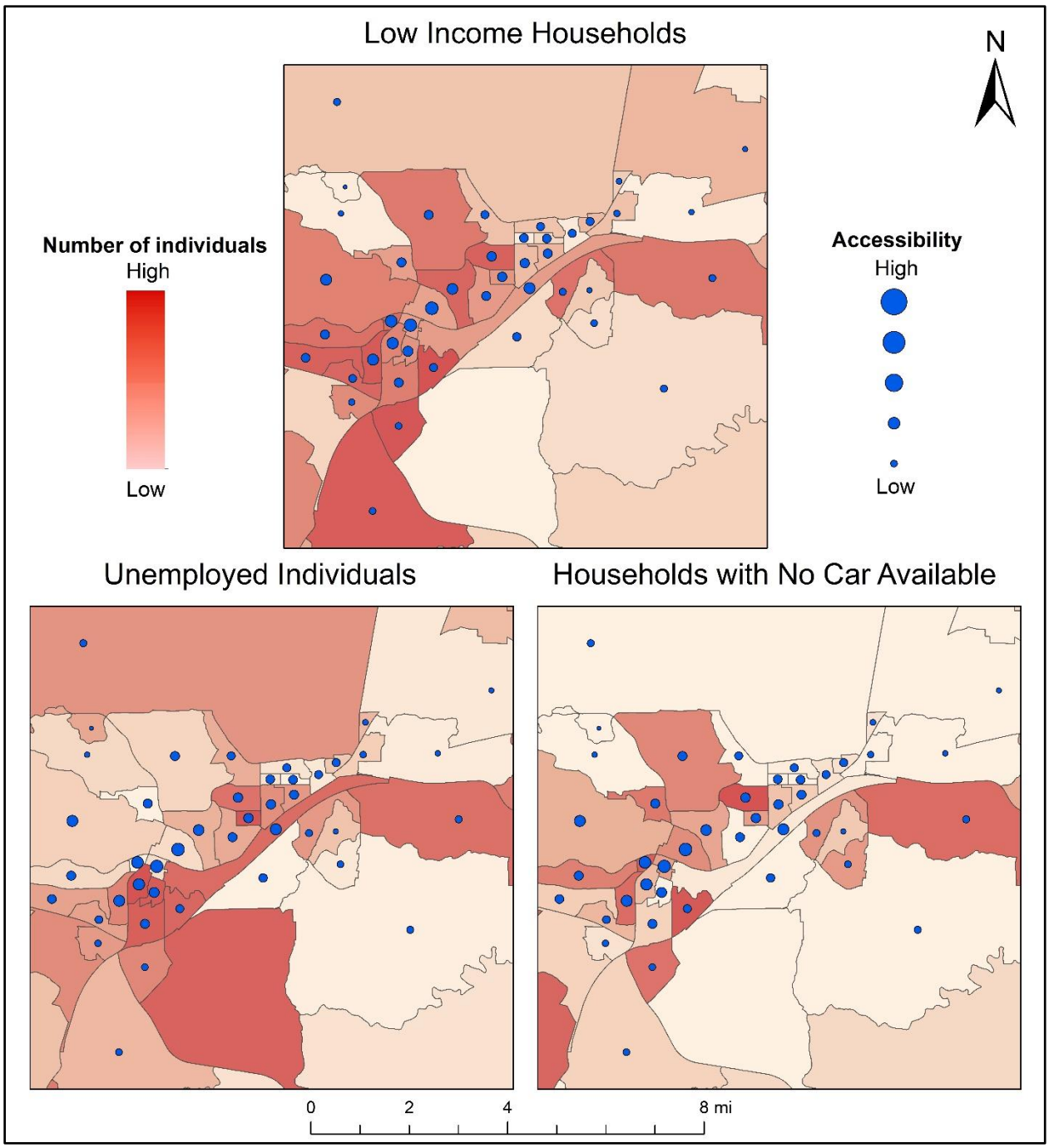

Figure 32. Maps showing the accessibility values relative to the individual socioeconomic attributes (income, unemployment, and car ownership).

As already mentioned, Jin and Paulsen (2018) provided evidence that increased job accessibility can potentially decrease unemployment rates and raise household income. In addition,

low income households spend a higher proportion of income on transportation (Haas et al. 2006). Therefore, it might be beneficial that the low income and unemployment variables are positively correlated with the accessibility levels. The high transit-based job accessibility levels currently experienced by these individuals can potentially aid in decreasing the costs with transportation, increasing household income, and decreasing unemployment rates in the future.

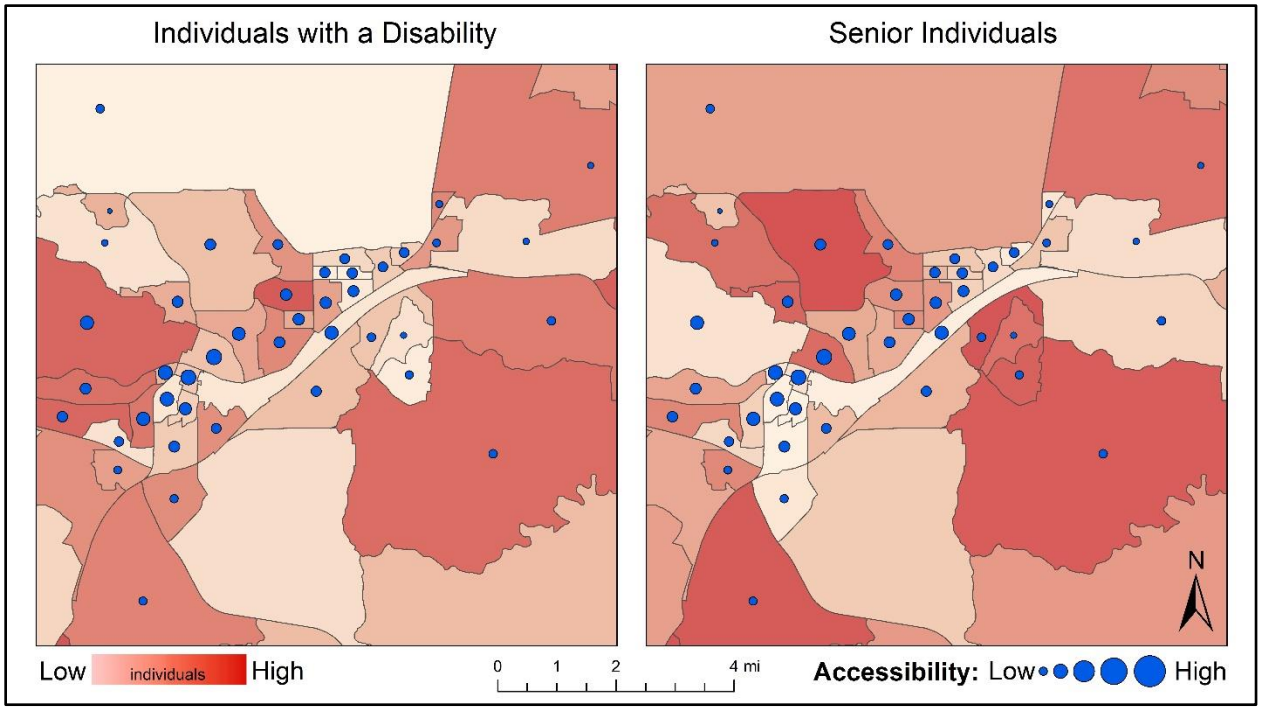

Figure 33. Maps showing the accessibility values relative to the individual socioeconomic attributes (individuals with a disability and senior individuals).

Although the correlation coefficient measured is significantly higher for the low income variable, the FMPO mentions that older adults and individuals with disabilities also have a special need for transit service (FMPO 2017). Therefore, these individuals could be prioritized in eventual expansions or service changes. However, this does not necessarily mean that these individuals are poorly served by the transit system. As already mentioned, NAIPTA operates the Mountain Lift, a demand-based paratransit service aimed at individuals with disabilities. This service is not

incorporated in the calculations in this study. Therefore, individuals with disabilities can take advantage of this additional service and may experience increased accessibility for that reason.

In addition, the ridership versus coverage problem also plays a role here. The problem would be easily solved if transportation agencies could operate on an unlimited budget. But again, these agencies have to make decisions to try balancing ridership with spatial coverage. The accessibility budget is limited and although an adjustment in the transit system configuration could benefit a certain population group, it would most likely compromise the service levels experienced by other population groups that might also have a special need for the service.

The transportation plan by the FMPO (2017) includes information about the origin locations that are prioritized by the transit system (Figure 31). Origin locations with a special need for transit include low income neighborhoods. Low income neighborhoods are defined based on the proportion of low income households. However, not all low income households are located in low income neighborhoods. As shown in Figure 32, a significant amount of low income households is located outside the low income neighborhoods delimited in Figure 31. The results presented in this section account for a significant portion of the low income population that live outside the low income neighborhoods. Therefore, a more comprehensive way of defining origins that should be prioritized is to develop some kind of transit users' index that does not exclude individuals with the definition of boundaries based on the proportions of a single specific socio-demographic characteristic.

4.3.2 - Accessibility and the Title VI Requirements

The Title VI requirements exist to ensure that the level and quality of public transportation service are provided in a way that individuals are protected from being discriminated based on race, color, or national origin (FTA 2017). All transportation agencies that receive financial

assistance from the Federal Transit Administration are required to comply with those requirements. NAIPTA maintains a Title VI Program and Implementation Plan to ensure that the requirements are being met (NAIPTA 2017). Therefore, measuring the spatial correlation between accessibility levels and socio-demographic data on race and national origin can potentially provide important information for transportation agencies with respect to these federal requirements.

Table 4 contains the Spearman correlation coefficients measured for the spatial relationships between accessibility and the individual variables on race and national origin. Two types of variables are included in the table, based on the most representative races and ethnicities that make up Flagstaff's population. The first type contains two variables that are concerned with the number of individuals of Hispanic or Latino origin (Hispanic or Latino and Not Hispanic or Latino). The second type includes variables for the number of individuals of each race (African American, Asian, Native American, white, two or more races, and other race alone). Figure 34 contains the respective bar charts that graphically illustrate the spatial relationships.

Table 4. Spearman correlation coefficients for the relationships between accessibility and the individual variables that are relevant to the Title VI requirements (race and national origin).

Spearman correlation coefficient (r_s)	
Hispanic or Latino	0.358
Not Hispanic or Latino	0.062
Two or More Races	0.270
African American	0.119
Asian	0.115
Native American	0.089
White	0.086
Other	0.038

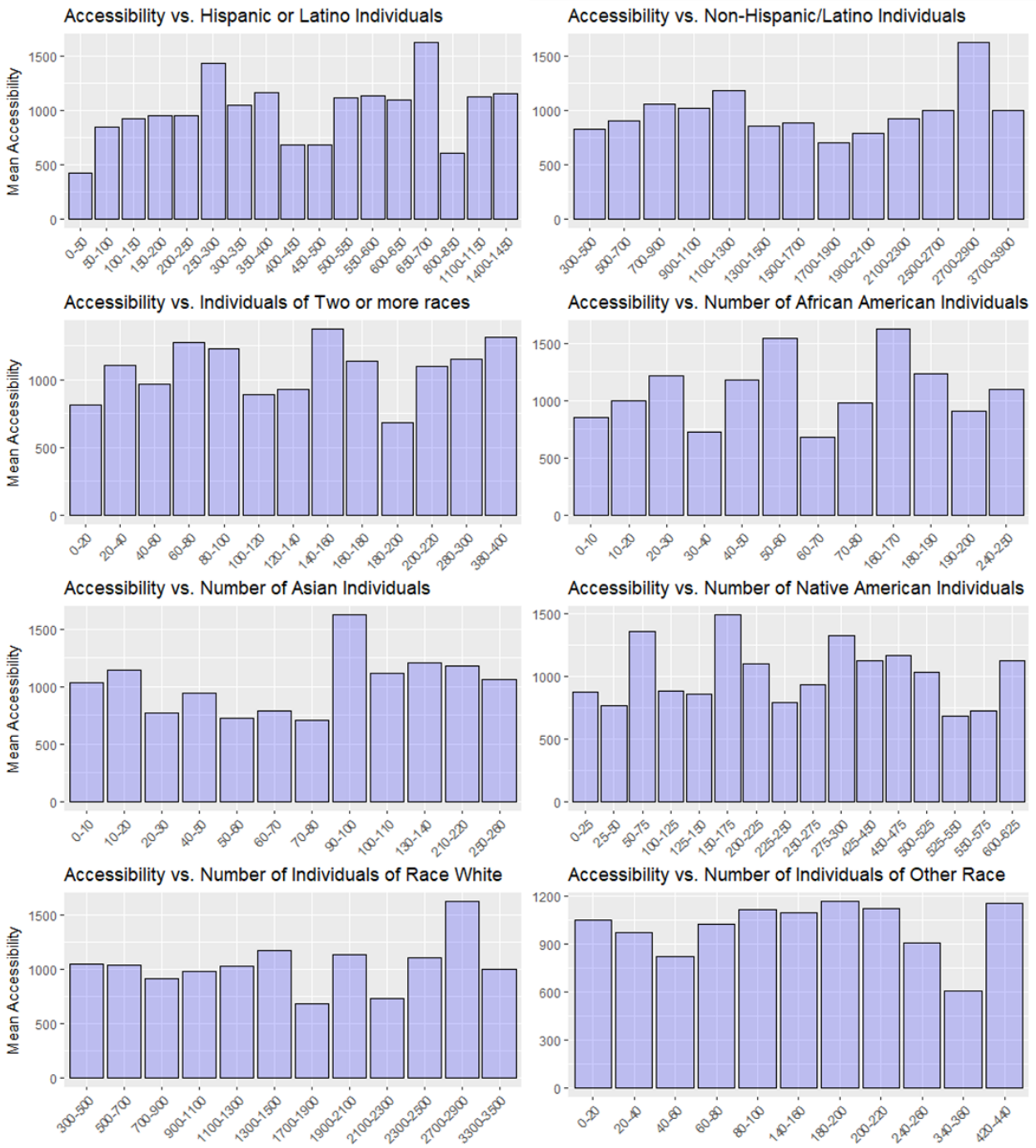

Figure 34. Bar charts showing the relationship between accessibility and the individual variables that determine race and national origin.

There is a significant difference between the coefficients obtained for the Hispanic-Latino population compared to the coefficients obtained for the non-Hispanic-Latino population. That can

be explained by the Hispanic-Latino population distribution better matching locations where higher accessibility levels are experienced. Figure 35 shows the difference between the spatial distributions of these two population groups, where it is clear that the Hispanic-Latino population is significantly more concentrated around the high accessibility areas. The Hispanic population is especially more significant at the east side of Flagstaff (green circle in Figure 35), where job accessibility is high and the non-Hispanic population is less representative.

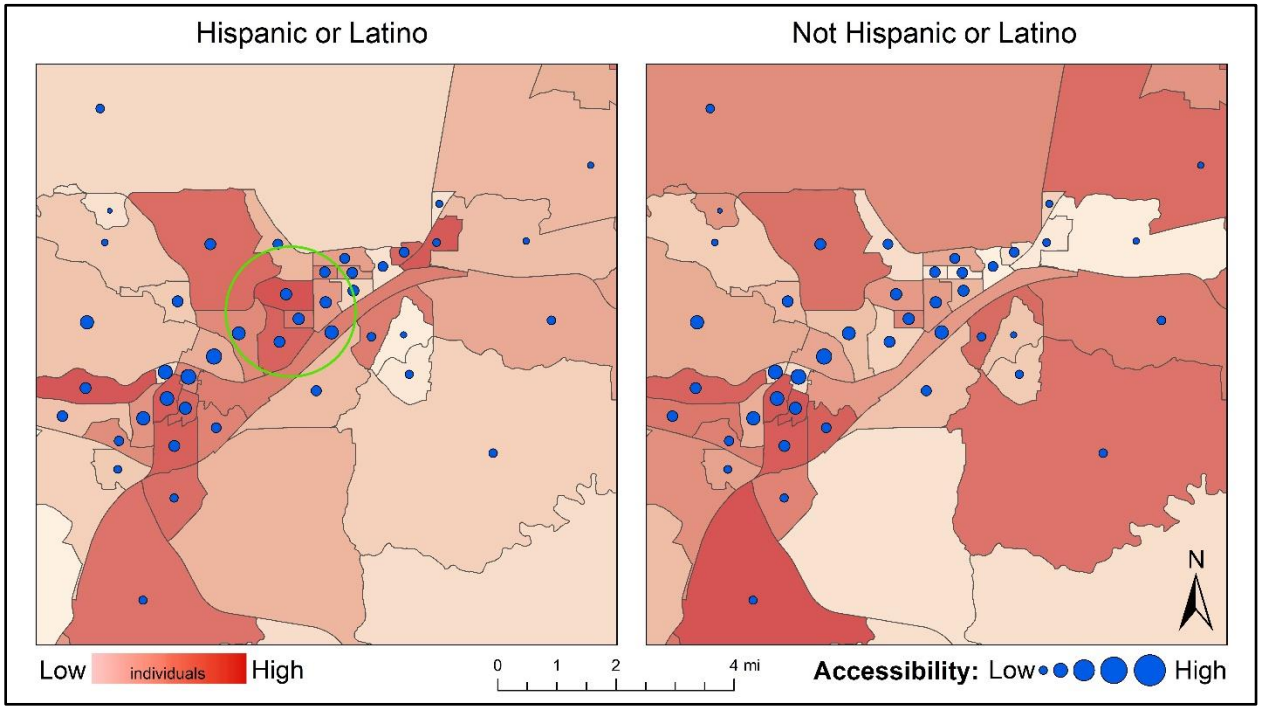

Figure 35. Maps showing the accessibility values relative to the individual ethnicity attributes (Hispanic or Latino and not Hispanic or Latino).

By looking at the different correlation coefficients obtained for the Hispanic-Latino population compared to the non-Hispanic-Latino population separately from the other results, one might get the wrong impression that the transit system is intentionally more beneficial to the Hispanic population. However, this might simply be due to the fact that the spatial distribution of the Hispanic population (Figure 35) well matches the distribution of the low income

neighborhoods (Figure 32). In fact, the Spearman correlation coefficient between these two variables is equal to 0.51. Therefore, it may be that the Hispanic-Latino population represents a significant part of the low income households in Flagstaff. Given that the transportation system has low income areas as one of its priorities (Figure 31), accessibility also happens to be high for the Hispanic population.

In addition, a transportation system is not only designed based on origin locations, but it also accounts for the destinations that are important for the individuals with a special need for transit service. In this specific case, destinations include large employers, commercial areas, medical facilities, and human service agencies (Figure 31; FMPO 2017). The Hispanic-Latino population is also significantly concentrated around these destinations.

The correlation coefficient between accessibility and the race variables is approximately the same (around 0.10), with the exception of the number of individuals who identify themselves with a single other race (0.04) or two or more races (0.27). The difference can be explained by the individuals identifying themselves with some other race alone being more representative in low transit-based accessibility areas, compared to the individuals who identify themselves with two or more races. This is especially clearer in the upper half of the maps in Figure 36. For the other races, although the relationships with accessibility have distinct spatial patterns (Figure 37), the differences are evened out and the overall spatial match with accessibility is approximately the same, as demonstrated by the correlation coefficients in Table 4.

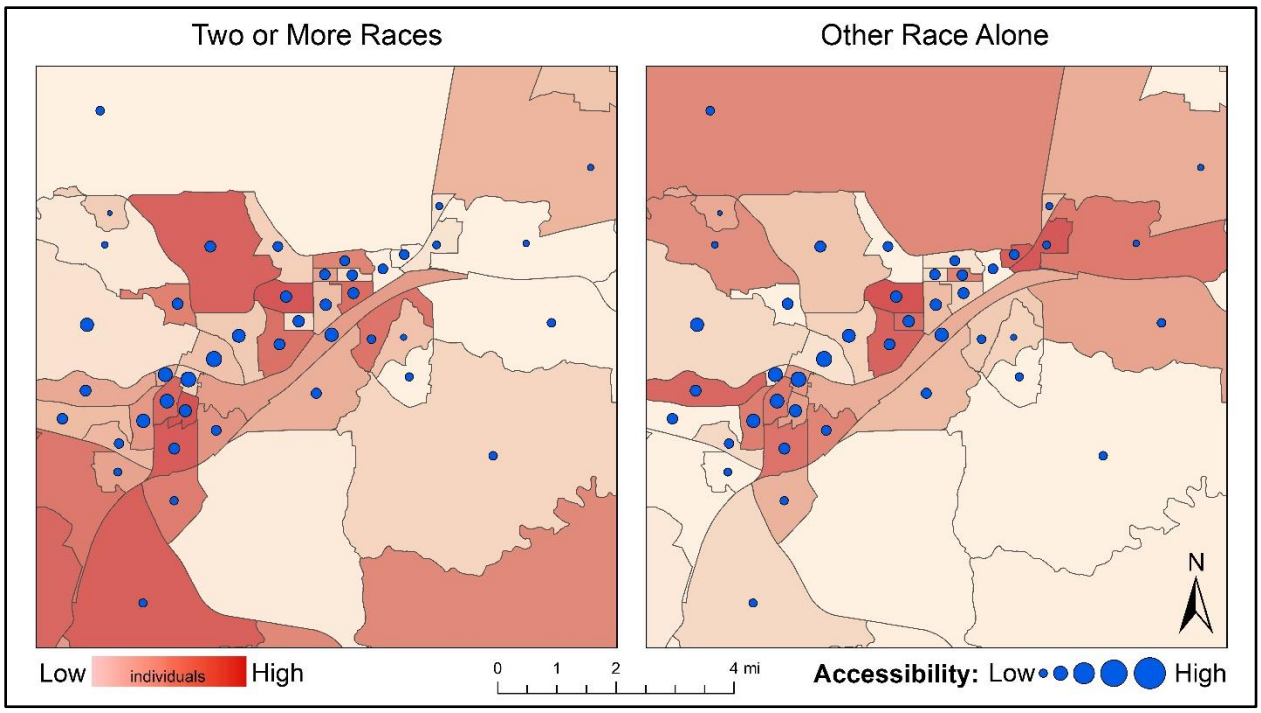

Figure 36. Maps showing the accessibility values relative to the individual race variables (two or more races and other race alone).

These results suggest that the spatial distribution of transit-based job accessibility is mostly impartial for the race variables. The coefficients are only divergent for the variables that do not correspond to a specific race and may well represent a variety of races. This means that the significantly higher or lower coefficients measured are not associated with any spatial bias toward any specific population group based on race. Given that these correlation coefficients measure the strength and direction of the spatial relationship between transit-based job accessibility and the number of individuals identifying themselves with different races, the outcomes are positive regarding the compliance with the Title VI requirements.

The main limitation of this type of analysis is that the variables that were correlated with accessibility are not independent. This was demonstrated, for instance, with the high correlation coefficients obtained for the low income and Hispanic-Latino variables. Those variables are

dependent because an individual can be Hispanic or Latino and live in a low income household at the same time. Therefore, unconditional observations cannot be made by analyzing these relationships individually. Information about the policies and practices employed by transportation agencies, such as the map in Figure 31 and the transportation plan (FMPO 2017), are generally helpful in uncovering the real meaning of these spatial relationships.

This analysis still significantly suffers from the MAUP due to the use of aggregated data at the block group level. The intra-zone variability within each block group is ignored. Although some peripheral block groups are associated with an accessibility measurement, in reality they are only partially served by the transit system. This introduces error in the analysis due to the inclusion of individuals who do not use the transit service. However, the census blocks from which these accessibility measurements were taken generally represent the majority of the population within these peripheral block groups. Therefore, the errors were not considered to be relevant enough to significantly influence the results.

In addition, the ACS datasets are only estimates that might deviate from the true value for these demographic attributes at varying levels. This is another source of inaccuracy that may influence the results by introducing error. However, the ACS estimates are the only open datasets that contain a more comprehensive set of socioeconomic attributes at the block group level. Nevertheless, the compatibilities between the results and the information provided by the transportation plan (FMPO 2017) indicate that the results are still appropriate.

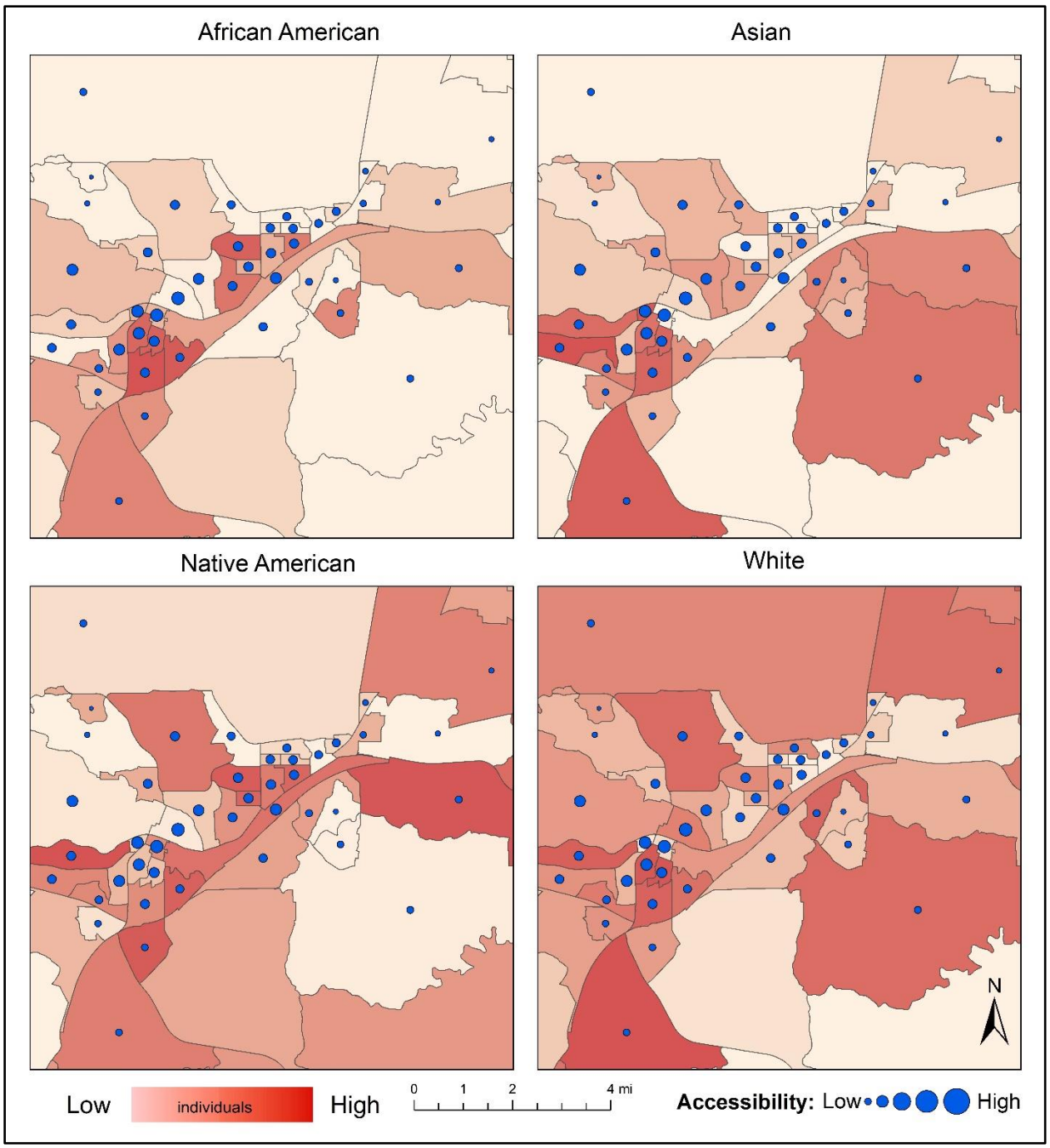

Figure 37. Maps showing the accessibility values relative to the individual race variables (African American, Asian, Native American, and White).

Despite the limitations of an analysis based on aggregated census counts, the block groups are detailed enough to capture variations within different neighborhoods in the city. In addition,

the results are presented in a simple and intuitive format that can be easily interpreted by policy makers and the general public. In short, the higher the correlation coefficient, the stronger the spatial relationship between accessibility and the specific socio-demographic variable. Hence, the methodology and the associated results presented in this section are still effective in producing valuable information that can be incorporated by public transportation agencies in their decision-making process.

CHAPTER 5: CONCLUSIONS

The purpose of this research project was to calculate the spatio-temporal variability of transit-based job accessibility, correlate the measurements obtained with the socioeconomic characteristics assumed to determine the need for public transit service, and discuss the social aspects of the accessibility scenario in Flagstaff. This was achieved by:

1) Constructing a temporally-enabled and schedule-aware transit-based multimodal network dataset that accounts for every step of a transit trip;
2) Using a gravity-based measurement to calculate accessibility from all populated census block locations to all the census block locations that are associated with at least one job;
3) Collecting and processing the relevant census-based data that was used in the socioeconomic characterization of the city with respect to the social need for transit;
4) Analyzing and interpreting the spatial patterns of accessibility and socioeconomics individually and simultaneously with maps, graphs, and correlation coefficients;
5) Using the measurements and correlations to evaluate the social aspects of the transit-based job accessibility scenario in the city of Flagstaff.

In addition, the study's research questions were addressed in the previous chapter (results and discussions) and can be answered as follows:

1) What is the spatial and temporal variability of the transit-based accessibility levels experienced by the city's residents to employment opportunities?

   The spatio-temporal variability of accessibility was mapped and measured, and the results are presented in sections 4.2.2 and 4.2.3. The most important results are the accessibility time series graphs (Figure 23) and the maps that summarize the accessibility measurements (Figures 25, 26, 27, and 29).

2) What is the spatial relationship between the accessibility measurements and the socio-demographic parameters that determine the need for public transit?

   The spatial relationship between accessibility and the socioeconomic attributes assumed to determine the need for public transit was measured with correlation coefficients (Tables 3 and 4) and displayed in bar charts (Figures 30 and 24) and maps (Figures 32, 33, 35, 36, and 37). These results are presented in section 4.3.

3) Does the accessibility scenario in Flagstaff reflect horizontal or vertical equity? In other words, are the individuals equally weighted or the resources are being distributed according to socioeconomic condition and the need for transit service?

   As discussed in section 4.3.1, the results show that the distribution of transit-based job accessibility has a better spatial match with the distribution of low income households compared to the population distribution, indicating that the accessibility scenario is more aligned with vertical equity. This means that the locations assumed to have a higher need for transit actually experience better transit-based job accessibility. However, given that the configuration of the transit system is not the only variable determining the accessibility levels experienced, and the fact that the socioeconomic variables are not independent, the distribution of transportation resources only partially explains this spatial relationship.

4) Are there any locations with a relatively high need for public transit that are not currently covered by the transit system?

   This question was addressed in section 4.1 in the analysis of the need for transit service. A transit users' index was calculated and maps displaying the need for transit in the study area were produced (Figures 14 and 15). Table 2 lists the transit users' index for neighborhoods currently not covered by the transit system. Some of these neighborhoods

have indexes greater than or equal to the average index for the neighborhoods served by the transit system.

Important findings concerning the transit-based job accessibility scenario were presented in this study. By using a transit users' index to estimate the general need for transit service, it was possible to identify four neighborhoods that currently do not have access to transit with an assumed relatively high need for transit: Fernwood Estates, Fort Valley Estates, Walnut Meadows, and Bellemont. These locations could be prioritized in an eventual spatial coverage expansion. The temporal variability of accessibility was analyzed, and the workers with schedules outside the current transit service hours were found to be concentrated in the NAU campus and in the surrounding neighborhoods. Therefore, these locations are assumed to have a higher need for an expansion of the current temporal coverage.

In the spatio-temporal variability analysis, downtown Flagstaff, the NAU campus, Milton Road, and the east side of the city along Route 66 were found to experience the highest transit-based job accessibility levels in the city due to either the density of the transit network or the proximity to a high number of jobs. On the other hand, some neighborhoods located north of downtown Flagstaff, specifically Cheshire, Lynwood, and Valley Crest experience significantly higher variability in the average accessibility levels. The average accessibility at these locations is also lower compared to most of the area covered by the transit system. An adjustment in the service frequencies could potentially improve and stabilize the accessibility levels experienced in that area.

The correlation coefficients obtained show that the transit-based job accessibility levels have the best spatial match with the number of low income households. By comparing this coefficient with the coefficient obtained for the total population, and considering other variables that are also associated with high coefficients (number of unemployed individuals and number of

households with no vehicle available), it can be suggested that the accessibility scenario is better for the individuals that are assumed to have a higher need for work trips with transit. On the other hand, the correlation with the number of individuals with a disability is weak and the correlation with the number of seniors is negative. However, the individuals with disabilities might still experience better accessibility by using the paratransit services that were not included in this study. In addition, although senior individuals are more likely to be unable to drive, they are also less likely to need transit service for work trips compared to other age groups.

In the analysis relevant for the Title VI requirements, it was found that the transit-based job accessibility scenario is mostly impartial considering the variables associated with specific races. However, the correlation between accessibility and the Hispanic-Latino population is significantly stronger compared to the correlation for the non-Hispanic population. The Hispanic-Latino population has a reasonable spatial match with the distribution of the low income households and the distribution of job opportunities. In addition, the non-Hispanic population is more representative in peripheral areas where job accessibility is lower. These two facts together explain the difference in the correlation coefficients obtained.

Not only were the research objectives accomplished and the research questions answered, but the workflow developed for the calculation of accessibility was compiled into a custom user-friendly ArcGIS geoprocessing tool. The main practical contribution of this study is that this tool is flexible enough to be reused with different settings and in other locations, as long as the input data requirements are met. Generally, that is not a problem given that street network data and census datasets are available for virtually every city in the United States. Transportation agencies are also increasingly more interested in publicly sharing their transit data through GTFS feeds.

The *CalculateAccessibility* tool together with the transit-based network dataset can be used by transportation agencies in simulations that can assist with the solution of many transit-related problems. The tool allows the use of different sets of origins and destinations, and can be adapted to perform different types of measurements. For instance, other types of accessibility calculations or travel time measurements can also be performed with small adjustments in the Python script. In addition, this network dataset not only solves OD cost matrices, but can also solve any other type of network problem that could be relevant for public transportation planning (Figure 5).

The Add GTFS to a Network Dataset tool (Morang 2016) allows for the integration between GTFS feeds of different agencies, being extremely helpful in a coordinated effort setting, such as the one in Flagstaff. This has been demonstrated with the network dataset created in this study, which includes the transit systems operated by NAIPTA and NAU. Other fixed-route services, such as the regional Navajo and Hopi transit systems (FMPO 2017) could be incorporated in the network dataset in the future. The accessibility effects of service changes can also be simulated with modifications in the GTFS feeds. Service frequencies can be updated, routes can be added and excluded, and the accessibility tool can be executed again with the new settings.

Geurs and van Wee (2004) highlight the importance of the development of sophisticated accessibility measurements that can still be easily interpreted by policy makers. Although Python programming skills are required for understanding and manipulating the source code, the custom ArcGIS geoprocessing tool can be accessed through the standard user-friendly interface (Figure 17) that is characteristic of any other ArcGIS tool. Therefore, even the more basic GIS users can have access to the tool's functionality without any programming experience. The transit-based network dataset can also be easily used within the ArcGIS user interface. In addition, the maps, graphs, and tables presented are simple and relatively easy to be interpreted.

The calculations performed in this study use a door-to-door approach based on transit timetable data and, unlike many studies previously developed (Lei and Church 2010), it does not overly simplify any of the steps of a trip with transit. In addition, the study also contributes to the research agenda by statistically exploring the spatial distribution of accessibility across different population groups, something that has not yet been fully explored in the current accessibility studies (van Wee and Geurs 2011). The information produced is relevant for transportation policies, and the outcomes of using this type of information to guide transportation efforts could be later assessed. Although this was not addressed in this study, this is a potential avenue for future work (van Wee 2016), and the methodology introduced offers opportunities in this direction.

This study also overcomes the limitations associated with the use of proportion-based socioeconomic indicators, such as the transit users' index in the study by Tribby and Zandbergen (2012), and the correlations based on percentages of population groups in the study by Farber, Morang, and Widener (2014). As already explained, this is especially relevant when the ridership versus coverage dilemma is considered (Walker 2015). Proportion-based analyses are deficient because different weights are attributed to individuals depending on the total population living in a census block or block group. When the calculations are based on percentages or proportions, an individual living in an underpopulated area will have a greater effect on the results than an individual living in an overpopulated region (Figure 16).

On the other hand, the analyses carried out in this study still significantly suffer from the MAUP because they are still based on aggregated areal units, such as census blocks and block groups. Given the unavailability of spatial data in higher resolution, this problem is unlikely to be easily solved with the use of open data. A possibility for future work is the use spatial microsimulation techniques to reconstruct microdata by combining non-spatial survey data, e.g.

the American Community Survey Public Use Microdata Sample (PUMS; USCB 2017b), with aggregate spatial data (Lovelace, Ballas, and Watson 2014). PUMS files include untabulated records of surveys about individuals and housing units. Another alternative would be requesting access to restricted-use census-based microdata under the supervision of the Census Bureau in a Federal Statistical Research Data Center (USCB 2015). Qualified researchers can request access to census-based microdata to address important research questions.

It is also possible that different types of datasets that could be useful for this type of analysis will be available in the near future. It has never been easier to collect, store, process and produce data, and data is being continuously generated from people's interaction with other people, machines, the environment, cities, and virtual codes and architectures (Graham and Shelton 2013). Currently, much more data are produced within two days than it has previously been produced in years, and a great deal of these datasets are spatially and temporally enabled, offering many opportunities for improving the understanding of geographical phenomena (Kitchin 2013). This large scale automatic data generation through sensors is currently relevant for movement and transport in places that are now referred to as smart cities (Batty 2013). Some examples include the travel card based automatic detection systems (ADCs) used in Chicago (Zhao, Rahbee, and Wilson 2007), London (Batty 2013), and Brisbane, Australia (Tao et al. 2014).

Ideally, an accessibility measurement should not ignore or overly simplify any of the four basic components defined in the reviews by Geurs and Ritsema van Eck (2001) and Geurs and van Wee (2004). The accessibility calculation employed accounts for the land-use, transportation, and temporal components in a comprehensive way. The individual component is addressed separately in the socioeconomic characterization of the study area with regard to the need for transit, and in the correlations performed to estimate the relationships between accessibility and socioeconomics.

Although the level of generalization is relatively higher in the individual component, mainly due to the unavailability of socio-demographic microdata, the other components are properly addressed without any kind of excessive generalization that could significantly compromise the results.

Despite the fact that the Python script developed (Appendix A) successfully automated the accessibility calculations, it can still be further improved. Datasets with large numbers of origins and destinations may take a long time to be processed and require large amounts of storage space. The storage space requirements could be minimized if the individual origin-destination line feature classes were processed in-memory instead of being processed after being permanently saved in a geodatabase. This would also eliminate some of the operations that are repeated in each one of the individual custom functions that make up the tool, resulting in processing time savings. On the other hand, since this operation would end up merging some of the individual functions into a continuous process, the flexibility of changing the script to perform specific steps separately instead of the entire workflow would be compromised. This is especially useful when small adjustments are necessary. Therefore, the ideal solution that enables both flexibility and efficiency would be to break down the tool into five separate tools for the individual functions, and an additional tool that executes the complete workflow.

Finally, it is also important to highlight that, although the results produced aim to benefit transportation planning efforts, the accessibility levels experienced are not only the result of the spatial configuration of the transit system. As demonstrated in this study, high accessibility can be experienced even when the transit network is not as dense as in other areas. This proves that the land-use patterns have a great influence on accessibility. Therefore, the decentralization of facilities and mixed land-use planning strategies can also aid in improving the accessibility scenario (Cheng and Bertolini 2013). In addition, as addressed by Levine et al. (2012), the

transportation network density also plays a role in accessibility. For that reason, the construction of new roads or even the improvement of pedestrian access to transit stops may be more effective and less costly than improving travel speeds or service frequencies. Consequently, transit systems are only one of the several variables that make up the accessibility equation.

The findings of this study and the methodology developed open up a range of possibilities that can play an important role in transit planning processes. As demonstrated in this study, geographic information systems and spatial analysis techniques can be used to translate complex relationships between transit systems, transportation networks, land-use parameters, and socioeconomics into simple user-friendly networks with problem-solving capabilities. This transit-based network dataset can greatly benefit transit planning processes with realistic simulations that can be used not only to analyze the current transit configuration, but also to forecast potential future scenarios. The GTFS-based system seamlessly combines transit information from different providers, making it significantly easier for transit agencies to coordinate their efforts. Transit planners that employ this methodology will be able to draw more accurate conclusions about the advantages and disadvantages of different alternatives for the configuration of transit networks in the future. Therefore, this study is significantly valuable for the improvement of the processes that determine the way that public transit resources are delivered, ultimately increasing accessibility and improving the quality of service for transit users.

BIBLIOGRAPHY

Batty, M. 2013. Big data, smart cities and city planning. *Dialogues in Human Geography* 3 (3): 274-279.

Bell, M. G. H., and Y. Iida. 1997. *Transportation Network Analysis*. Chichester, UK: John Wiley.

Beyazit, E. 2011. Evaluating Social Justice in Transport: Lessons to be Learned from the Capability Approach. *Transport Reviews* 31 (1): 117-134.

Bhat, C., S. Handy, K. Kockelman, H. Mahmassani, Q. Chen, and L. Weston. 2000. *Development of an Urban Accessibility Index: Literature Review.* Austin, TX: Center of Transportation Research, The University of Texas at Austin.

Chen, J., J. Ni, C. Xi, S. Li, and J. Wang. 2017. Determining intra-urban spatial accessibility disparities in multimodal public transport networks. *Journal of Transport Geography* 65: 123-133.

Cheng, J., and L. Bertolini. 2013. Measuring urban job accessibility with distance decay, competition and diversity. *Journal of Transport Geography* 30: 100-109.

COF (City of Flagstaff). 2016. City of Flagstaff Official Website - Geographic Information Systems (GIS), http://www.flagstaff.az.gov/index.aspx?NID=1720, accessed 14 November 2016.

COF (City of Flagstaff). 2018. Economic Information | City of Flagstaff Official Website, https://flagstaff.az.gov/1099/Economic-Information, accessed 2 January 2018.

Delmelle, E. C., and I. Casas. 2012. Evaluating the spatial equity of bus rapid transit-based accessibility patterns in a developing country: The case of Cali, Colombia. *Transport Policy* 20: 36-46.

Djurhuus, S., H. S. Hansen, M. Aadahl, and C. Glümer. 2016. Building a multimodal network and determining individual accessibility by public transportation. *Environment and Planning B: Planning and Design* 43 (1): 210-227.

El-Geneidy, A., R. Buliung, E. Diab, D. v. Lierop, M. Langlois, and A. Legrain. 2016a. Non-stop equity: Assessing daily intersections between transit accessibility and social disparity across the Greater Toronto and Hamilton Area (GTHA). *Environment and Planning B: Planning and Design* 43 (3): 540-560.

El-Geneidy, A., D. Levinson, E. Diab, G. Boisjoly, D. Verbich, and C. Loong. 2016b. The cost of equity: Assessing transit accessibility and social disparity using total travel cost. *Transportation Research Part A: Policy and Practice* 91: 302-316.

ESRI (Environmental Systems Research Institute). 2006. *Linear Modeling with Networks.* Redlands, CA: ESRI Press.

———. 2016. OD cost matrix analysis—Help | ArcGIS for Desktop, http://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/od-cost-matrix.htm, accessed 2 December 2016.

———. 2017a. Esri Support GIS Dictionary, https://support.esri.com/en/other-resources/gis-dictionary, accessed 17 November 2017.

———. 2017b. Types of network analysis layers—Help | ArcGIS Desktop, http://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/types-of-network-analyses.htm, accessed 20 December 2017.

———. 2017c. What is the ArcGIS Network Analyst extension?—Help | ArcGIS Desktop, http://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/what-is-network-analyst-.htm, accessed 19 December 2017.

Farber, S., K. Bartholomew, X. Li, A. Páez, and K. M. Nurul Habib. 2014. Assessing social equity in distance based transit fares using a model of travel behavior. *Transportation Research Part A: Policy and Practice* 67: 291-303.

Farber, S., M. Z. Morang, and M. J. Widener. 2014. Temporal variability in transit-based accessibility to supermarkets. *Applied Geography* 53: 149-159.

FMPO (Flagstaff Metropolitan Planning Organization). 2017. *Coordinated Public Transit – Human Services Transportation Plan.* Flagstaff, AZ: Flagstaff Metropolitan Planning Organization.

FTA (Federal Transit Administration). 2017. Title VI of the Civil Rights Act of 1964, https://www.transit.dot.gov/regulations-and-guidance/civil-rights-ada/title-vi-civil-rights-act-1964, accessed 16 March 2017.

Galati, S. R. 2006. *Geographic Information Systems Demystified.* Boston, MA: Artech House.

Geurs, K. T., and Ritsema van Eck, Jan R. 2001. *Accessibility measures: review and applications. Evaluation of accessibility impacts of land-use transportation scenarios, and related social and economic impact.* Bilthoven, the Netherlands: National Institute of Public Health and the Environment (RIVM).

Geurs, K. T., and B. van Wee. 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography* 12 (2): 127-140.

Goodchild, M. F. 1998. Geographic Information Systems and Disaggregate Transportation Modeling. *Geographical Systems* 5: 19-44.

Google Developers. 2016. GTFS Static Overview | Static Transit, https://developers.google.com/transit/gtfs/, accessed 14 November 2016.

Graham, M. R., M. J. Kutzbach, and B. McKenzie. 2014. *Design Comparison of LODES and ACS Commuting Data Products.* Washington, DC: U.S. Census Bureau.

Graham, M., and T. Shelton. 2013. Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography* 3 (3): 255-261.

Grengs, J. 2010. Job accessibility and the modal mismatch in Detroit. *Journal of Transport Geography* 18 (1): 42-54.

Haas, P. M., C. Makarewicz, A. Benedict, T. W. Sanchez, and C. J. Dawkins. 2006. *Housing & transportation cost trade-offs and burdens of working households in 28 metros.* Chicago, IL: Center for Neighborhood Technology.

Hägerstraand, T. 1970. What about people in regional science? *Papers and Proceedings of the Regional Science Association* 24 (1): 7-24.

Hansen, W. G. 1959. How Accessibility Shapes Land Use. *Journal of the American Institute of Planners* 25 (2): 73-76.

Ikhrata, H., and P. Michell. 1997. Technical Report of Southern California Association of Governments' Transportation Performance Indicators. *Transportation Research Record* 1606: 103-114.

Jin, J., and K. Paulsen. 2018. Does accessibility matter? Understanding the effect of job accessibility on labour market outcomes. *Urban Studies* 55 (1): 91-115.

Kitchin, R. 2013. Big data and human geography Opportunities, challenges and risks. *Dialogues in Human Geography* 3 (3): 262-267.

Koenig, J. G. 1980. Indicators of urban accessibility: Theory and application. *Transportation* 9 (2): 145-172.

Kwan, M. 2012. The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers* 102 (5): 958-968.

Lei, T. L., and R. L. Church. 2010. Mapping transit-based access: integrating GIS, routes and schedules. *International Journal of Geographical Information Science* 24 (2): 283-304.

Levine, J., J. Grengs, Q. Shen, and Q. Shen. 2012. Does Accessibility Require Density or Speed? *Journal of the American Planning Association* 78 (2): 157-172.

Levitas, R., C. Pantazis, E. Fahmy, D. Gordon, E. Lloyd, and D. Patsios. 2007. *The Multi-dimensional Analysis of Social Exclusion*. Bristol, UK: Department for Communities and Local Government (DCLG).

Litman, T., and M. Brenman. 2012. *A New Social Equity Agenda For Sustainable Transportation.* Victoria, BC: Victoria Transport Policy Institute.

Litman, T., and D. Burwell. 2006. Issues in sustainable transportation. *International Journal of Global Environmental Issues* 6 (4): 331-347.

Longley, P. A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind. 2011. *Geographic Information Systems and Science*. Hoboken, NJ: John Wiley & Sons.

Lovelace, R., D. Ballas, and M. Watson. 2014. A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. *Journal of Transport Geography* 34: 282-296.

Lucas, K. 2004. *Running on empty: Transport, social exclusion and environmental justice*. Bristol, UK: The Policy Press.

———. 2012. Transport and social exclusion: Where are we now? *Transport Policy* 20 (Supplement C): 105-113.

Massey, D. S., J. Rothwell, and T. Domina. 2009. The Changing Bases of Segregation in the United States. *The Annals of the American Academy of Political and Social Science* 626: 74-90.

Milan, B. F., and F. Creutzig. 2017. Lifting peripheral fortunes: Upgrading transit improves spatial, income and gender equity in Medellin. *Cities* 70: 122-134.

Miller, H. J. 1999a. Measuring Space-Time Accessibility Benefits within Transportation Networks: Basic Theory and Computational Procedures. *Geographical Analysis* 31 (1): 1-26.

———. 1999b. Potential Contributions of Spatial Analysis to Geographic Information Systems for Transportation (GIS-T). *Geographical Analysis* 31 (4): 373-399.

———. 2004. Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers* 94 (2): 284-289.

Morang, M. 2016. Yay, transit! - Using GTFS Data in ArcGIS Network Analyst, http://www.transit.melindamorang.com/, accessed 14 November 2016.

Morris, J. M., P. L. Dumble, and M. R. Wigan. 1979. Accessibility indicators for transport planning. *Transportation Research Part A: General* 13 (2): 91-109.

NAIPTA (Northern Arizona Intergovernmental Public Transportation Authority). 2017. *Title VI Program and Implementation Plan.* Flagstaff, AZ: Northern Arizona Intergovernmental Public Transportation Authority.

NAIPTA (Northern Arizona Intergovernmental Public Transportation Authority). 2018. Routes & Schedules| Fixed Route Service in Flagstaff, https://mountainline.az.gov/routes/, accessed 12 January 2018.

Neutens, T. 2015. Accessibility, equity and health care: review and research directions for transport geographers. *Journal of Transport Geography* 43: 14-27.

Neutens, T., T. Schwanen, F. Witlox, and P. De Maeyer. 2010. Equity of Urban Service Delivery: A Comparison of Different Accessibility Measures. *Environment and Planning A* 42 (7): 1613-1635.

Omer, I. 2006. Evaluating accessibility using house-level data: A spatial equity perspective. *Computers, Environment and Urban Systems* 30 (3): 254-274.

Openshaw, S. 1984. *The modifiable areal unit problem.* Norwich, UK: Geo Books.

Páez, A., R. G. Mercado, S. Farber, C. Morency, and M. Roorda. 2010. Relative Accessibility Deprivation Indicators for Urban Settings: Definitions and Application to Food Deserts in Montreal. *Urban Studies* 47 (7): 1415-1438.

Páez, A., D. M. Scott, and C. Morency. 2012. Measuring accessibility: positive and normative implementations of various accessibility indicators. *Journal of Transport Geography* 25: 141-153.

Preston, J., and F. Rajé. 2007. Accessibility, mobility and transport-related social exclusion. *Journal of Transport Geography* 15 (3): 151-160.

Shahandashti, S. K. F., X. C. Liu, and G. Zhang. 2017. An efficient General Transit Feed Specification (GTFS) enabled algorithm for dynamic transit accessibility analysis. *PLoS ONE* 12 (10): e0185333.

Tao, S., J. Corcoran, I. Mateo-Babiano, and D. Rohde. 2014. Exploring Bus Rapid Transit passenger travel behaviour using big data. *Applied Geography* 53: 90-104.

Thill, J. C., and J. L. Horowitz. 1997. Travel-time constraints on destination-choice sets. *Geographical Analysis* 29 (2): 108-123.

Tribby, C. P., and P. A. Zandbergen. 2012. High-resolution spatio-temporal modeling of public transit accessibility. *Applied Geography* 34: 345-355.

USCB (United States Census Bureau). 2015. Federal Statistical Research Data Centers, https://www.census.gov/about/adrm/fsrdc/about.html, accessed 28 February 2018.

———. 2017a. American FactFinder, https://factfinder.census.gov/, accessed 11 December 2017.

———. 2017b. PUMS Documentation, https://www.census.gov/programs-surveys/acs/technical-documentation/pums.html, accessed 28 February 2018.

———. 2018. LEHD Origin-Destination Employment Statistics (LODES), https://lehd.ces.census.gov/data/, accessed 5 January 2018.

van Wee, B. 2016. Accessible accessibility research challenges. *Journal of Transport Geography* 51: 9-16.

van Wee, B., and K. T. Geurs. 2011. Discussing equity and social exclusion in accessibility evaluations. *European Journal of Transport and Infrastructure Research* 11 (4): 350-367.

Viegas, J. M., L. M. Martinez, and E. A. Silva. 2009. Effects of the Modifiable Areal Unit Problem on the Delineation of Traffic Analysis Zones. *Environment and Planning B: Planning and Design* 36 (4): 625-643.

Walker, J. 2015. Explainer: The Transit Ridership Recipe, http://humantransit.org/2015/07/mega-explainer-the-ridership-recipe.html, accessed 24 January 2018.

Wang, C., and N. Chen. 2015. A GIS-based spatial statistical approach to modeling job accessibility by transportation mode: case study of Columbus, Ohio. *Journal of Transport Geography* 45: 1-11.

Werner, C. 1985. *Spatial Transportation Modeling.* Beverly Hills, CA: Sage Publications.

Zeiler, M. 1999. *Modeling Our World: The ESRI Guide to Geodatabase Design*. Redlands, CA: ESRI Press.

Zhao, J., A. Rahbee, and N. H. M. Wilson. 2007. Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. *Computer-Aided Civil and Infrastructure Engineering* 22 (5): 376-387.

APPENDICES

Appendix A: *CalculateAccessibility* Geoprocessing Tool Source Code (Python Script)

```
#---------------------------------------------------------------------------
# Tool name: CalculateAccessibility
#
# Description: Automates the calculation of accessibility from a
#              transit-enabled network dataset for several times of a day.
#              Results are recorded in a table.
# Input Parameters:
#                  ---REQUIRED---
#                  OD Cost Matrix Layer (layer file)
#                  Output Geodatabase (file geodatabase)
#                  Weekend (Boolean)
#                  Census Blocks Jobs (point feature class)
#                  Census Blocks Table (table)
#
# Output: Census Blocks Table (table) -- contains the accessibility for each
#         location at each calculated time.
#---------------------------------------------------------------------------

# import required modules
import arcpy, datetime, math

###########################################################################

def SolveMatrix(matrix_lyr, outputGDB, weekend):
    """
    Iteratively solves the OD Cost Matrix Layer and saves the individual
    feature classes in the output geodatabase.
    """
    arcpy.AddMessage("Solving OD Cost Matrix loop...")

    # checks the network analyst extension
    arcpy.CheckOutExtension("Network")

    # accesses the network analysis layer properties
    matrix_solver = arcpy.na.GetSolverProperties(matrix_lyr)

    # specifies whether the analysis will be performed for weekday or weekend
    if weekend: # True for weekend
        # start time
        matrix_solver.timeOfDay = datetime.datetime(1900, 1, 6, 3,20,0)
        # end time
        end = datetime.datetime(1900, 1, 6, 23,55,0)
    else: # False for weekday
        # start time
        matrix_solver.timeOfDay = datetime.datetime(1900, 1, 1, 3,20,0)
        # end time
        end = datetime.datetime(1900, 1, 1, 23,55,0)
```

```
    # establishes time interval between two consecutive calculations
    increment = datetime.timedelta(minutes=5)
    # stores the network analysis sublayer names
    subLayerNames = arcpy.na.GetNAClassNames(matrix_lyr)

    # while loop that iteratively solves the network analysis layer
    while True:
        arcpy.na.Solve(matrix_lyr) # solves OD cost matrix layer
        # holds the layer with the result
        outputMatrix = arcpy.mapping.ListLayers(matrix_lyr,
                                                subLayerNames["ODLines"])[0]

        # standardizes the names for the output feature classes: MatrixHHMM
        if matrix_solver.timeOfDay.hour < 10:
            hour = '0' + str(matrix_solver.timeOfDay.hour)
        else:
            hour = str(matrix_solver.timeOfDay.hour)
        if matrix_solver.timeOfDay.minute < 10:
            minute = '0' + str(matrix_solver.timeOfDay.minute)
        else:
            minute = str(matrix_solver.timeOfDay.minute)

        # holds the name of the output FC: MatrixHHMM
        outputName = "Matrix" + hour + minute
        # copies the feature class to the output geodatabase
        arcpy.CopyFeatures_management(outputMatrix,
                                      outputGDB + "/" + outputName)
        arcpy.AddMessage("Matrix solved: " + outputName[-4:])
        # checks if the end time for the calculation was reached
        if matrix_solver.timeOfDay == end:
            break # exits the loop
        # adds the time increment for the next calculation
        matrix_solver.timeOfDay += increment

    # clean up
    del matrix_lyr
    del matrix_solver

    # returns the network analyst extension
    arcpy.CheckInExtension("Network")

##########################################################################

def AddIDfields(outputGDB):
    """
    Creates separate attribute fields for the origin and destination unique
    IDs for each feature class in the output geodatabase.
    """

    arcpy.AddMessage("Creating ID fields for outputs...")

    # sets the workspace environment to the output geodatabase
```

```
    arcpy.env.workspace = outputGDB

    # creates a list of all feature classes in the workspace
    FClist = arcpy.ListFeatureClasses()

    # for loop that iterates through the feature classes in FClist
    for fc in FClist:

        arcpy.AddMessage("Adding ID fields for: " + fc)

        # creates a new attribute field for the origin IDs
        arcpy.AddField_management(fc, "GEOID10", "TEXT", field_length=15)
        # creates a new attribute field for the destination IDs
        arcpy.AddField_management(fc, "JOBID", "TEXT", field_length=15)

        # creates an update cursor based on the current FC
        cursor = arcpy.da.UpdateCursor(fc, ["Name", "GEOID10", "JOBID"])
        # for loop that iterates over the rows in the update cursor
        for row in cursor:
            # writes the origin ID in the GEOID10 field
            row[1] = row[0].split(" - ")[0]
            # writes the destination ID in the JOBID field
            row[2] = row[0].split(" - ")[1]
            # updates the current row
            cursor.updateRow(row)

    # clean up
    del row
    del cursor

##########################################################################

def JoinField(outputGDB, BlockJobs):
    """
    Joins the values for the number of jobs from the census blocks feature
    class to each feature class in the output geodatabase.
    """
    arcpy.AddMessage("Joining JOBS field to outputs...")

    # sets the workspace environment to the output geodatabase
    arcpy.env.workspace = outputGDB

    # creates an empty dictionary to hold the {Job ID : Number of Jobs} pairs
    JobDict = {}

    # creates a search cursor based on the census blocks FC
    cursor = arcpy.da.SearchCursor(BlockJobs, ["JOBID", "JOBS"])
    # for loop that iterates over the rows in the search cursor
    for row in cursor:
        # adds an entry in the dictionary -> JOBID:JOBS
        JobDict[row[0]] = row[1]

    # clean up
```

```
    del row
    del cursor

    # creates a list of all feature classes in the workspace
    FClist = arcpy.ListFeatureClasses()

    # for loop that iterates through the feature classes in FClist
    for fc in FClist:

        arcpy.AddMessage("Joining JOBS fields to: " + fc)

        # adds an attribute field for the number of jobs in the destinations
        arcpy.AddField_management(fc, "JOBS", "DOUBLE")

        # creates an update cursor based on the current FC
        cursor = arcpy.da.UpdateCursor(fc, ["JOBID", "JOBS"])
        # for loop that iterates over the rows in the update cursor
        for row in cursor:
            # writes the number of jobs in the JOBS field
            row[1] = JobDict[row[0]]
            # updates the current row
            cursor.updateRow(row)

    # clean up
    del row
    del cursor

##############################################################################

def Calculate_ajti(outputGDB):
    """
    Calculates the individual aj/ti values and records the results in a new
    attribute field.
    """
    arcpy.AddMessage("Calculating individual (aj/ti) values...")

    # sets the workspace environment to the output geodatabase
    arcpy.env.workspace = outputGDB

    # creates a list of all feature classes in the workspace
    FClist = arcpy.ListFeatureClasses()

    # for loop that iterates through the feature classes in FClist
    for fc in FClist:

        arcpy.AddMessage("Calculating aj/ti for: " + fc)

        # adds a new attribute field for the aj/ti values
        arcpy.AddField_management(fc, "aj_ti", "DOUBLE")

        # creates an update cursor based on the current FC
        cursor = arcpy.da.UpdateCursor(fc, ["aj_ti", "JOBS",
                                            "Total_TransitTravelTime"])
```

```
        # for loop that iterates over the rows in the update cursor
        for row in cursor:
            # if travel time is less than 1 minute: aj/ti = JOBS
            if row[2] <= 1.0:
                row[0] = row[1]
            # else: aj/ti = JOBS/Total_TransitTravelTime
            else:
                row[0] = row[1]/row[2]
            # updates the current row
            cursor.updateRow(row)

    # clean up
    del row
    del cursor

##############################################################################

def Calculate_Ai(outputGDB, CensusBlocksTable):
    """
    Sums up the aj/ti values and calculates the accessibility (Ai) value for
    each origin location (census block centroid) at each calculated time.
    Results are recorded in the output table (rows with census block ID and
    columns with the time of the day).
    """
    arcpy.AddMessage("Calculating accessibility (Ai) by time...")

    # sets the workspace environment to the output geodatabase
    arcpy.env.workspace = outputGDB

    # creates a list of all feature classes in the workspace
    FClist = arcpy.ListFeatureClasses()

    # for loop that iterates through the feature classes in FClist
    for fc in FClist:

        arcpy.AddMessage("Calculating Ai for: " + fc)

        # creates an empty dictionary to hold the
        # {Census Block ID : Accessibility} pairs
        Ai_Dict = {}

        # creates a search cursor based on the current FC
        cursor = arcpy.da.SearchCursor(fc, ["GEOID10", "aj_ti"])
        # for loop that iterates over the rows in the search cursor
        for row in cursor:
            # if ID already exists in the dictionary, aj/ti value is added to
            # the current value
            if Ai_Dict.has_key(row[0]):
                Ai_Dict[row[0]] += row[1]
            # if ID is encountered for the first time, the current aj/ti
            # value is assigned to this ID
            else:
                Ai_Dict[row[0]] = row[1]
```

```
        # clean up
        del row
        del cursor

        # adds a new field to the output table: AiHHMM (based on the name of
        # the current feature class)
        arcpy.AddField_management(CensusBlocksTable,
                                  "Ai" + fc[-4:], "DOUBLE")

        # creates an update cursor based on the output table
        cursor = arcpy.da.UpdateCursor(CensusBlocksTable,
                                       ["GEOID10", "Ai" + fc[-4:]])
        # for loop that iterates over the rows in the update cursor
        for row in cursor:
            # updates the row by retrieving the corresponding Ai value from
            # the dictionary
            row[1] = Ai_Dict[row[0]]
            # updates the current row
            cursor.updateRow(row)

        # clean up
        del row
        del cursor

##########################################################################

# gets the input parameters from the ArcGIS script tool
matrix_lyr = arcpy.GetParameter(0)
outputGDB = arcpy.GetParameterAsText(1)
weekend = arcpy.GetParameter(2)
BlockJobs = arcpy.GetParameter(3)
CensusBlocksTable = arcpy.GetParameter(4)

# executes the individual parts (functions) of the analysis using the input
# parameters provided
SolveMatrix(matrix_lyr, outputGDB, weekend)
AddIDfields(outputGDB)
JoinField(outputGDB, BlockJobs)
Calculate_ajti(outputGDB)
Calculate_Ai(outputGDB, CensusBlocksTable)
```