SPATIAL ANALYSIS OF TRAFFIC CONGESTION AND TRANSIT

ACCESSIBILITY IN AUSTIN, TEXAS

by

Adam Bakiera, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Geography
August 2016

Committee Members:

Russell Weaver, Chair

Colleen Hiner

Yongmei Lu

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

ACKNOWLEDGEMENTS

I would first like to thank the Graduate College at Texas State University for giving me the opportunity to attend and complete my Master’s Degree. I would also like to thank all of my committee members, Dr. Rusty Weaver, Dr. Colleen Hiner, and Dr. Yongmei Lu. All of you helped me tremendously in this project. Lastly, I would like to thank my family for always supporting me.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

I. INTRODUCTION

In the period following World War II, processes of urbanization in the United States began to take on "parasitic" forms (Beauregard, 2006), whereby growth and prosperity in some places generated stagnation, shrinkage, and decline in others (Beauregard, 2011). For instance, many American central cities became intimately tied to notions of poverty, aging infrastructure, crime, and related "urban problems" during this era (Weaver et al., 2016). Emerging suburban communities beyond city borders therefore allowed some types of households to leave these problems behind, by making a crucial trade-off: quick and convenient access to central city employment and consumption opportunities were given up in favor of newer homes on larger lots farther from urban centers. Hence, households that were (are) able to economically absorb longer commutes and higher transportation costs relocate(d) from cities to suburbs in droves (Teaford, 2007).

Historically, large-scale patterns of city-to-suburb migration were facilitated by American federal programs that promoted homeownership (Hayden, 2004), invested heavily in the creation of transportation infrastructure (Leyden and Goldberg, 2015), especially highways and interstates (Bullard, 2000), and then eventually devolved almost all urban policymaking activities to lower levels of government (Kantor, 2010). The result has been a decades-long exercise in *sprawl*. While the term *urban sprawl* lacks a universal definition, most researchers agree that it refers to:

> "the low-density outward expansion of metropolitan areas often characterized by leapfrog development. Rather than laying out compact,

> contiguous, interconnected subdivisions, developers skip over vacant land to create housing tracts or commercial strips ever more distant from the metropolitan core. [This situation] implies automobile-dependent development where everyone must drive to every destination because homes are isolated in homogenous residential subdivisions rather than within walking distance of stores and offices" (Teaford, 2007: 188).

In the United States, sprawl was a novel process of urban development for most pre-War cities. Older cities were generally constructed to be "walking cities," in which settlements were dense and heterogeneous, and where the primary mode of transport was walking (Weaver et al., 2016). By contrast, American cities that experienced their major population booms after World War II were largely built on top of a "culture" of sprawl (Briggs, 2009), such that their current urban forms privilege automobility over virtually all other forms of movement (e.g., Bullard, 2000; Agyeman, 2013). Indeed, to participate in the processes of sprawl, one is effectively required to "have access to an automobile because public transit is usually inadequate or nonexistent" in most parts of sprawling metropolitan regions (Bullard, 2000: 1).

Ample research has shown that sprawl has "huge" social and environmental costs (Briggs, 2009: 52; also see Agyeman [2005] and Ewing [2008]). Foremost among these costs are traffic congestion (Ewing, 2008),

> "inefficient use of water and other natural resources, overburdened infrastructure, pollution, disinvestment in older communities, a spatial mismatch between where many disadvantaged job seekers live and where jobs are growing, and more" (Briggs, 2009: 52).

For these reasons, it is widely claimed that sprawl is counter to contemporary notions of "sustainability" and "sustainable development" (e.g., Agyeman, 2005; Register, 2006; Briggs, 2009). One idea that is often put forward to curb or reverse patterns of sprawl is to better organize urban spaces around mass public transit systems (Bernick and Cervero, 1997; Register, 2006), or at least to provide better and more equitable access to public transit and other forms of mobility in existing urban spaces (e.g., Agyeman, 2005, 2013).

The argument is simply that better access to and greater utilization of public transit in U.S. metropolitan regions might help to mitigate some of the social and environmental costs of sprawl listed above (e.g., Register, 2006). Clearly, then, empirical and applied geographic research on the relationships between public transit and the consequences of sprawl can add considerable value to policymaking and planning efforts in American urban areas. This thesis seeks to contribute to the analytical toolbox and the body of quantitative evidence related to these efforts. In that sense, it is beyond the scope of this study to engage further with the critical literature on sprawl, sustainability, and the many meanings and intersections of these two concepts (see Bullard et al., 2000; Weaver, 2015). Rather, to facilitate its analytical and empirical contributions, the thesis takes as given three well-developed and testable propositions that are implied in the preceding paragraphs. Namely, (1) the roadways of sprawling metropolises are frequently characterized by substantial traffic congestion (e.g., Ewing, 2008); (2) accessibility to and usage of public transit has the capacity to alleviate some of the costs of sprawl, including traffic congestion (e.g., Bernick and Cervero, 1997); and (3) access to and usage of public transit tend to be unevenly distributed in many cities (e.g., Bullard et al., 2000; Agyeman, 2005).

The thesis tests these three general propositions for the specific case of Austin, Texas and its surrounding metropolitan region. Austin is one of the fastest-growing cities in one of the fastest growing metropolitan regions in the United States, with most of its growth occurring after World War II (Weissmann, 2015). Accordingly, Austin is said to be an "exemplar of urban sprawl" (Torrens, 2008: 5) that is located in a state where the "sprawling development patterns that require so much driving" are a "primary cause of congestion" (Surface Transportation Policy Project, 1999: 5). Austin is therefore assumed to be a meaningful case for analyzing relationships between congestion and public transit in a sprawling metropolis.

With these points in mind, the study will test the three above-mentioned propositions in the context of the following research questions and subquestions:

1. What is the geographic distribution of traffic congestion in the greater Austin area?
    a. What is the relationship between traffic congestion and public transit *access*?
    b. What is the relationship between traffic congestion and public transit *usage*?
2. Are traffic problems and access to public transit distributed equitably between socioeconomic groups in the city of Austin?

In addressing these questions, a core objective of the study is to develop an analytical framework and adopt a methodology for operationalizing the concepts of “traffic congestion” and “public transit access,” using readily available state and federal administrative data sources. Researchers argue that quantifying these concepts is typically necessary in transportation planning projects and applied urban geographic investigations; yet both are difficult to measure. With respect to traffic congestion, traffic data tend to be sparse and are usually only available for highways and other major roadways (e.g., Lowry and Dixon, 2012). As a result, mapping congestion across an entire metropolitan road network can be extremely difficult. With respect to public transit access, public transportation data tend to be maintained by individual transit operators. Because transit operators take on various forms—including public/private partnerships, public corporations, and private corporations—public transit data necessarily vary from region to region. In that sense, it is a challenge to measure “access” consistently across space.

To push back against these challenges, this thesis leverages data from the state of Texas Department of Transportation (TxDOT) and the federal Environmental Protection Agency (EPA). Following existing regression-based techniques from the literature (e.g., Anderson et al., 2006), the TxDOT data are used to spatially extrapolate known Average Daily Traffic (ADT) counts, as well as Designed ADT counts, for major roadways to every road segment in the greater Austin area. These results allow for an operational definition of “traffic congestion.” From there, spatial analytical tools and techniques within a

Geographic Information Systems (GIS) environment enable that measure of congestion to be mapped across the entire study area, and to be summarized for small subareas (U.S. census block groups) in the Austin city limits. The EPA data are then used to compute a composite index of "public transit access" in the city limits of Austin. The index is based on a neighborhood's average distance to a transit stop, as well as the aggregate frequency of transit trips that run in the given neighborhood. Ultimately, these two variables are used to evaluate the relationships, if any, that exist between traffic congestion and public transit access in the city of Austin. The latter (access) measure can further be evaluated for unevenness in its distribution between different types of socioeconomic neighborhoods. To the extent that the study's methodology relies on publicly available datasets that are relatively consistent across study areas, the analyses can easily be replicated in other metropolitan or micropolitan regions, cities, and states for purposes of transportation planning and related applied geographic research.

II. LITERATURE ON TRAFFIC CONGESTION MODELING AND PUBLIC TRANSIT ACCESS

Modeling traffic congestion and public transit accessibility allows researchers to pinpoint problem areas in transportation systems. Among other uses, traffic modeling helps transportation planners to identify roads or road segments whose existing flows are near, at, or over capacity (e.g., Rodrigue et al., 2006). Put differently, modeling assists researchers in locating those parts of the road network where existing traffic volumes exceed their intended or designed volumes. Concerning transit accessibility, many individuals and families, particularly those in low income groups, lack access to automobiles (Sanchez and Brenman, 2007). For such individuals, inadequate provision of public transportation negatively affects one's ability to reach employment opportunities or other amenities (Bullard et al. 2000). In some cases, urban public transportation networks either do not physically connect to certain neighborhoods, or, in cities such as Austin, Texas, high traffic congestion can decrease the functionality and viability of available public transit options. As the literature surveyed below will show, both traffic congestion and public transit access have received significant attention from scholars and practitioners. In many cases, though, the two concepts are studied somewhat independently. Synthesizing the two lines of literature here is therefore crucial for moving forward with the project outlined in Chapter I, which will consider the interrelationships between congestion and public transit in Austin, Texas.

Traffic Congestion Modeling

One of the most widely used approaches for estimating traffic congestion begins with Average Daily Traffic (ADT) count data (Anderson et al., 2006). ADT is a measure used primarily in transportation planning and transportation engineering. It is the total volume of vehicle traffic on a road segment for one year, divided by 365 days. Importantly, however, due to limited resources and manpower, ADT cannot practically be measured on every single road segment—not even within a single city. Consequently, researchers have focused on spatially extrapolating ADT counts from a given location (often major roadways and highways) to other locations within a road network (often minor and residential streets).

Regression Based Modeling

In a widely cited study, Anderson et al. (2006) developed a multiple linear regression model to estimate demand for various roadways using ADT data for a small town in Alabama. The authors note that their study was specifically designed just for this particular town. However, their methods have a wider range of applicability (Anderson et al., 2006; Lowry and Dixon, 2012). In brief, Anderson et al. (2006) sampled 96 roadways around the town of Anniston, Alabama, and collected data on the following independent variables: roadway functional classification, number of lanes, population within a half mile radius of the roadway, employment within a half mile radius, and a variable indicating the roadway as a through street or designation street based on side friction. The dependent variable for the model was ADT. The authors reported that the model's R-squared value of 0.819 suggested that it was a good fit for the data.

Another approach is based on spatial regression methods. Eom et al (2006) used a spatial regression model in their study of predicting Average Annual Daily Traffic (AADT), where averages were taken over a multi-year period. The authors propose that AADT at one monitoring station is correlated with AADT counts at neighboring monitoring stations. Their method provides reliable AADT estimates for coverage counts which they argued can improve the predictive capability of the aspatial regression model. The spatial model takes into account both spatial trend (mean) and spatial correlation, which is modeled using a kriging technique. The study area was Wake County, North Carolina. The data used for the regression is divided into three components: AADT, roadway characteristics, and census data. Roadway characteristics included urban, suburban, or rural area type classification, number of lanes, posted speed limit, functional classification, signal density, and presence of a median. The signal density is the number of signals within 1 mile in a uniform roadway segment. Census data at the block group level included total population, number of households, household size, number of households with young children, median income, and employment characteristics.

The study produced some insightful findings. First, the effect of the Highway Functional Class 3—rural arterial—is much higher than the effects of the other functional classes. Also, urban area relative to rural area (Area Type 1) and suburban area relative to rural area (Area Type 2) have significant higher AADT, as expected. The city of Raleigh has much higher predictions with lower standard deviations than the rest of the country because the average distance between any two monitoring stations in the city is much shorter. As a result, on average, each station in the city has many more neighbors, which is crucial in kriging. Additionally, if a spatial correlation exists between AADT at one station and locations of its neighbors, then the overall predictive capability of the regression model is much better than that of an ordinary regression model.

A very similar study by Wang and Kockelman (2009) also used spatial interpolation and kriging. Their study used linear extrapolation for temporal predicting and kriging for spatial interpolation. They used data from the Texas Department of Transportation (TxDOT) for AADT, which includes almost 28,000 sites across the state. They predicted the next year's traffic first, and then circled back and used interpolation to predict traffic on roadways without AADT monitors.

Temporal Extrapolation was done using the ordinary least square regression (OLS) based on the seven years of collected traffic records (1999-2005), while the kriging method was used for interpolation. Their extrapolated counts are shown in Figure 1.

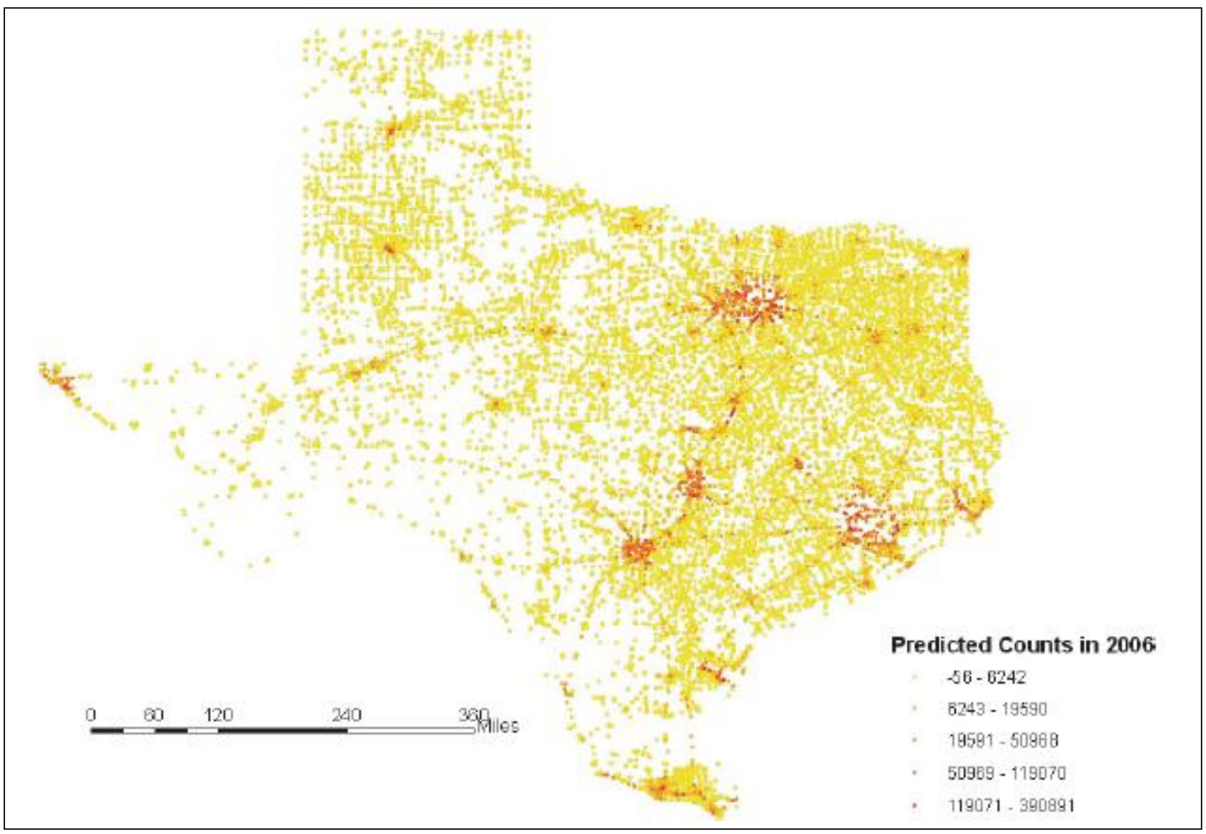

Figure 1. Predicted counts for sites in 2006 (source: Wang and Kockelman, 2009).

The results of the spatial interpolation led Wang and Kockelman to conclude that variance increases with distance but levels out at a certain point. Most of the data from TxDOT is for Class 1 roads (interstate highways) and Class 2 roads (principal arterials). The study found that spatial autocorrelation of Class 1 roads is more distance-dependent as well as distance sensitive. The authors argue that this phenomenon occurs because Class 1 roads are better connected; access points along Class 1 roads lead to fluctuations while Class 2 roads appear more continuous.

While both Eom et al. (2006) and Wang and Kockelman (2006) find that spatially explicit modeling can lead to higher predictive power relative to spatial extrapolation from linear regression models alone (recall the study by Anderson et al. [2006]), Lowry and Dixon (2012: 5) observe that spatially explicit models can be computationally complex and rely on data that may not be available in all types of places. In addition, they point to the high explanatory power and predictive capabilities of multiple regression models that use only a handful of variables (e.g., Mohamad et al. 1998; Anderson et al. 2006). Ultimately, then, Lowry and Dixon (2012) recommend and illustrate the value of using regression-based extrapolation techniques for estimating ADT in practical transportation planning applications. They do so by developing a custom toolbox for Esri ArcGIS software.

Lowry and Dixon's (2012) Custom GIS Toolbox

Following earlier regression-based methods (e.g., Mohamad et al. 1998; Anderson et al. 2006), Lowry and Dixon (2012) demonstrate how characteristics of a road segment and its surrounding area can be used to extrapolate known ADT data to other segments in a given road network. In particular, Lowry and Dixon (2012), developed a Geographic Information Systems (GIS) toolbox for estimating traffic counts on roads that did not have readings in a small town in Idaho. Their toolbox included five related tools. Four of the tools involved preparing spatial data for the linear regression, and exploring the connectivity of the streets. The fifth, and arguably the most powerful, is then used to carry out a linear regression in which ADT is modeled as a function of user-specified independent variables. After the authors provided background details on the toolbox, Lowry and Dixon (2012) proceeded to use the tools to estimate ADT for minor roadways in their study area. In doing so, Lowry and Dixon (2012) highlight the value of parsimonious models and methods for engaging in applied transportation planning and urban geographic research.

GIS-Based Modeling

As an applied project, this thesis follows the general regression-based modeling strategy used by Lowry and Dixon (2012). However, it is important to note that GIS offers transportation researchers much more than the ability to create customized analytical tools. Indeed, another way in which traffic congestion can be forecasted and simulated is through the use of GIS-based modelling. Much of the existing literature in this area focuses on European cities. However, a growing body of scholarship is now applying and developing similar techniques for American cities.

Zhong and Hanson (2009) used GIS for travel demand modeling (TDM) to estimate traffic on low functional class (i.e., non-primary) roads. The authors point out that using travel demand models is currently rare in the extant literature. To address this gap, their study used TDM to estimate traffic volumes throughout New Brunswick, Canada. They divided the study area into two parts: York County, which includes the capital of the province (Fredericton), and the Beresford Census Consolidated Subdivision, which is a popular tourist area in the northern part of the province. The TDM was developed using TransCAD software, and it relies on a regression equation that estimates trips to a zone based on the number of housing units in the zone and the zone's work activity. The results show a clear overestimation of traffic counts. In all cases, the volumes were overestimated by a range of 11% to 700%. The authors note that only 65% of the roads in the network were associated with traffic flows; so, by default, these roads would be more congested in the model. They then used regression to modify the TDM and manually add traffic flows to local roads, which improved the overall accuracy. However, Zhong and Hanson (2009) identified two specific areas where the study could be improved. First, lower class road traffic can be modelled more effectively by reducing the size of the study area. However, this would limit the model's effectiveness, especially if it would require connecting multiple study areas through complex network algorithms. Second, a more comprehensive model could be made by increasing focus of the trip generating sites. This would require data at household levels, which the authors note is sometimes difficult to acquire, and almost always complex to include in the overall model.

Salonen and Toivonen (2013) took a different approach by including three different ways to model accessibility to various destinations in Helsinki, Finland. Their "simple" approach ignores traffic, congestion, and related factors, while their "intermediate" model includes traffic but ignores parking. Finally, their "advanced model", which also had the best results, is known as the "door to door" approach. In this model, walking to the car/bus stop, traffic, finding parking and walking from the bus/car to work is all factored in. They found that direct comparisons can be made between public and private transportation when using the same model, for example, simple private and simple public transportation. It is not possible to compare simple private and intermediated private. Overall, the authors concluded that it is more efficient to commute in Helsinki by private car—a surprising result considering the public transit infrastructure and relatively high cost of anything automobile related (the price of the car itself, insurance, gasoline, etc.) The biggest downside or gap in this study is the availability of data. Salonen and Toivonen were fortunate to be provided data by the Helsinki Regional Transport, but they note that their particular model is "data hungry," and that it may be overwhelming for someone without moderate GIS skills.

Tang et al (2003) used four different methods to develop four different models for traffic forecasting in Hong Kong. As in previous studies, these models were compared to real data to check for accuracy. The four methods include time series, neural network, nonparametric regression, and Gaussian maximum likelihood (GML). For the time series, the authors used the Box-Jenkins method, a commonly used technique for forecasting either discrete or continuous data. The Neural Network Method (NN) is the idea of writing computer software based on the build of the human brain. The structure has three parts: the input later, the hidden layers, and the output layer. The Nonparametric Regression (NPR) creates predictions based on a group of similar past cases defined around the current state at the time of the prediction. This is essentially a nearest neighbor analysis. The last method is the GML. The GML was proposed in the early 21st century, and it integrates historical traffic information and real-time information.

In predicting AADT with the four models, the authors note that the GML method performed the best, while the ARIMA (time series) performed the worst. NPR was also notably accurate in predicting values for known data points. Nevertheless, the authors note that while NPR and GML are valuable modeling techniques, they are not perfect. NPR is complex because identification of neighbors can be difficult. The GML model, on the other hand, makes two restrictive assumptions: the variables must be normally distributed and time dependent.

Summary of Traffic Congestion Modeling Literature

The preceding subsections communicate some of the popular, though diverse, strategies used by researchers and practitioners to model traffic and traffic congestion in various types of road networks. Two key themes seem to emerge from this literature. First, the data needed to answer some of the most important traffic-related research and planning questions are sparse. For that reason, statistical methods and/or geospatial modeling techniques are needed to estimate otherwise "unknown" or "unobserved" information. Second, there is no universal "best" method or modeling technique for performing such analyses. While complex spatial methods can increase predictive accuracy in some cases (e.g., Eom et al., 2006; Wang and Kockelman, 2006), comparatively simpler multiple regression techniques have noteworthy—if not equal or greater—explanatory power (Anderson et al., 2006; Lowry and Dixon, 2012), and may be more accessible to practitioners and other applied researchers (e.g., Salonen and Toivonen 2013). Given the scope and intended contributions of this thesis, the present study adopts a regression-based spatial extrapolation method going forward. Notwithstanding this methodological choice, it should be clear from the above subsections that the continued methodological developments in traffic modeling—including the growing number of studies that employ GIS-based modeling techniques—are valuable and necessary contributions to scholarship in transportation planning and urban geography.

Selected Studies of Access

Access is a multidimensional concept that deals with the capacity of an entity to reach, or to be reached by, other entities. From a public planning perspective, "access" therefore plausibly centers on the question of whether individuals can safely travel to various types of land uses, such as a mix of [quality] residential, employment, and recreation opportunities (e.g., Sanchez and Brenman, 2007). For example, in a study of the San Francisco Bay Area by Kawabata and Shen (2006), *access* to jobs was measured as a function of: (1) the travel times between two locations by car and public transit; (2) a threshold travel time; (3) the total number of jobs in a given location; (4) the total number of workers (consisting of both the employed and unemployed) living in given location; and (5) the proportion of households with cars in that location (Kawabata and Shen, 2007). Regression analyses showed that in both 1990 and 2000, job accessibility was associated with shorter commutes for driving than for public transit. In other words, commuting by public transit took workers much longer. The model also found a significant negative association between population density and public transit commuting, meaning that transport service is better for high density locations. Mean income was inversely statistically significantly related to access, suggesting that working families with higher incomes accept longer commutes (via automobile) to afford better housing (refer to Chapter I). Notably, though, Kawabata and Shen caution that San Francisco has an abnormally high public transit usage relative to the national average (approximately two times), and that disparities in areas with higher personal vehicle use will be different in the Bay Area than in the U.S. as a whole (Kawabata and Shen, 2007).

In another study, Hess (2005) examined *access* to employment for low-income populations in Erie and Niagara Counties in western New York State. This study reached several interesting conclusions with respect to the social aspect of connectivity. In effect, the author found that access to employment in the city via public transportation is better than access to employment in the suburbs (Hess, 2005). While at first this may seem like a positive thing, the study also found that there are many more jobs in the suburbs. Furthermore, African Americans in the study area had considerably less access to automobiles relative to whites. As such, the areas that African Americans might search for jobs is limited to where public transit is accessible, which is predominantly within the city. Because more African Americans live in the city than in the suburbs—specifically, Buffalo is 37% African American whereas the suburbs are only 2% African American—lack of access to both automobiles *and* public transit is a severe limiting factor in finding employment.

Another article, by Handy (2005), explores how certain propositions from proponents of "smart growth" are affecting the country. The propositions include: (1) building more highways will contribute to more sprawl; (2) building more highways will lead to more driving; (3) investing in light rail systems will increase densities; and (4) adopting New Urbanism design strategies will reduce automobile use (Handy, 2005). In other words, the propositions seem to argue that, as summarized in Chapter I above, *access* to highways facilitates sprawl; while *access* to public transit can mitigate the costs of sprawl. The evidence produced by Handy suggests these claims are likely to be true, but to what extent they are true remains to be seen.

As Handy (2005) observes, "[n]ew roads fuel the already explosive growth in the amount we drive. New and wider roads bring short-term relief, at great expense". The author also found that increase in lane miles is associated with a 3-11% increase in vehicle miles traveled (Handy, 2005). Handy then argues that light rail systems potentially impact development and growth in two ways. First, if these systems reduce travel time, they may actually encourage residents to live farther out because of decreased travel time. However, she notes that most light rails are built in areas that are already developed, in which case the density may increase. Next, if there is not development along a proposed new light rail, development might spring up around the line once it is built. Overall, the study showed that investing in light rail—increasing *access*—will increase population density; but only if the conditions are right. Areas of success will likely need economic growth, station locations in areas that can grow, and maybe most important of all, public sector involvement (Handy, 2005).

Finally, Handy explores the degree to which strategies recommended by the Congress for the New Urbanism (CNU) can reduce automobile dependency. A general claim associated with the CNU is that if designed correctly, a "New Urban" city will make it easier to walk and not depend so much on the automobile. The study found that trip frequencies are mostly affected by socioeconomic characteristics but trip lengths are affected more by the built environment. Furthermore, the mode of transportation depends on the combination of socioeconomics and built environment (Handy, 2005). Handy cites a previous study that found residents of an Austin neighborhood chose to live there because it was within walking distance to the grocery store. Generally, New Urbanist strategies do make it easier to live without so much reliance on the automobile, but like the previous points, it is uncertain just how much driving it really saves.

In his book *Growing Stronger*, Robert Bullard (2008) discusses issues related to inequitable *access* to a variety of resources between minority and non-minority population subgroups. He observes that tax subsidies help create suburban sprawl because money was allocated towards building more roads, rather than expanding public transportation. In turn, this sprawl leaves the most disadvantaged populations "behind" in the depopulating areas. Bullard (2008) further observes that traffic congestion, long commutes, and lost time and efficiency for business are results of sprawl. Crippled central cities and declining suburbs are a drag on the entire regional economy (Bullard, 2008: 54). However, social justice advocates are skeptical of anti-sprawl movements because they have a regional focus. People are unaware that what is happening outside of the inner city affects them. They are also afraid to lose cultural and neighborhood identities by placing their trust in regional entities (Bullard, 2008: 60). Finally, Bullard writes that it is possible for certain regions to be doing so well that concentrated poverty is not seen as an issue (Bullard, 2008: 61).

In a case study of the Atlanta metropolitan area, Bullard (2008) found that the majority of entry level jobs were not within a quarter mile of (i.e., *accessible* to) public transportation. In addition, 39% of all Black households in Atlanta do not have access to cars, and in 2000, only 34% of the region's jobs were within a one-hour public transit ride of low-income neighborhoods (Bullard, 219). The author concludes that "[w]hether highway or airport sprawl is "good" or "bad" will almost always depend on where you live, and whether or not you own a car" (Bullard, 2008: 241). According to Bullard, 40% of public transit riders are low income individuals. The problem is jobs are located in areas that are not *accessible*. Bullard advocates for construction of affordable housing near rail transit lines. He references a study done in 2004 where it was discovered that housing demand near rail transit stations is high (Bullard, 2008: 305).

Summary of Selected Access(ibility) Literature

The foregoing, non-exhaustive survey of accessibility studies that are relevant to this thesis suggests that *access* in a transportation planning context depends, at minimum, on spatial distance (Bullard 2008) and quality of the entity being accessed (e.g., Sanchez and Brenman, 2007; Agyeman, 2005, 2013). In what follows, these insights will guide this study's definition of *access* to public transit in the city of Austin.

III. DATA, ANALYTICAL FRAMEWORK, AND METHODS

Data Sources

The data for this study came from three different sources: The Texas Department of Transportation (TxDOT), the Smart Location Database is published by the U.S. Environmental Protection Agency (EPA), and the most recent U.S. Census Five-Year American Community Survey (ACS) for the period 2010-2014. Data obtained from TxDOT contain spatial and nonspatial characteristics for every road segment in the greater Austin region (Figure 2). Following existing studies (Mohamad et al., 1998; Anderson et al., 2006; Lowry and Dixon, 2012), the main road segment attributes of interest are: known Average Daily Traffic (ADT) counts, year ADT was recorded, the ADT volume for which the road segment was designed ("Designed ADT"), functional classification of the road, and number of lanes. In addition, GIS-based zoning data from the city of Austin were used to calculate the spatial distance (in kilometers) from each road segment to Austin's "central business district" zoning area.

Next, the EPA's Smart Location Database (SLD) was developed to address the growing demand for data products and tools that consistently compare the location efficiency of various places. The SLD summarizes several demographic, employment, public transportation, and built environment variables for every census block group (CBG) in the United States (Ramsey and Bell, 2014). Once again with existing traffic modeling studies in mind (e.g., Mohamad et al., 1998; Anderson et al., 2006), data on the total level of employment (number of jobs) and total population in each CBG in the metro area were extracted from the SLD. The SLD further provides data on the distance from a CBG's population-weighted center to the nearest transit stop, in meters, and the aggregate frequency of transit service (number of trips) per square mile in each CBG. Data on the locations transit stops and frequency of services were obtained by the EPA from Capital Metro, the city of Austin's public transportation operator. The employment and population totals contained in the SLD are for the Decennial Census year 2010, while the public transit data correspond to 2012 (Ramsey and Bell, 2014). Because the Capital Metro services are not provided throughout the entire greater Austin region, these latter public transit variables are only obtained for CBGs that fall within Austin's city limits.

Finally, data were extracted from the most recent U.S. Census ACS in order to classify CBGs in the city of Austin based on a variety of relatively current socioeconomic status (SES) and demographic characteristics. Doing so allows for an assessment of the second research question articulated earlier in this thesis—namely, whether access to public transit is distributed equally between different types of socioeconomic neighborhoods (or, in this case, CBGs). ACS data were also obtained on the percentage of workers (16 years and older) who commute to their jobs via public transportation.

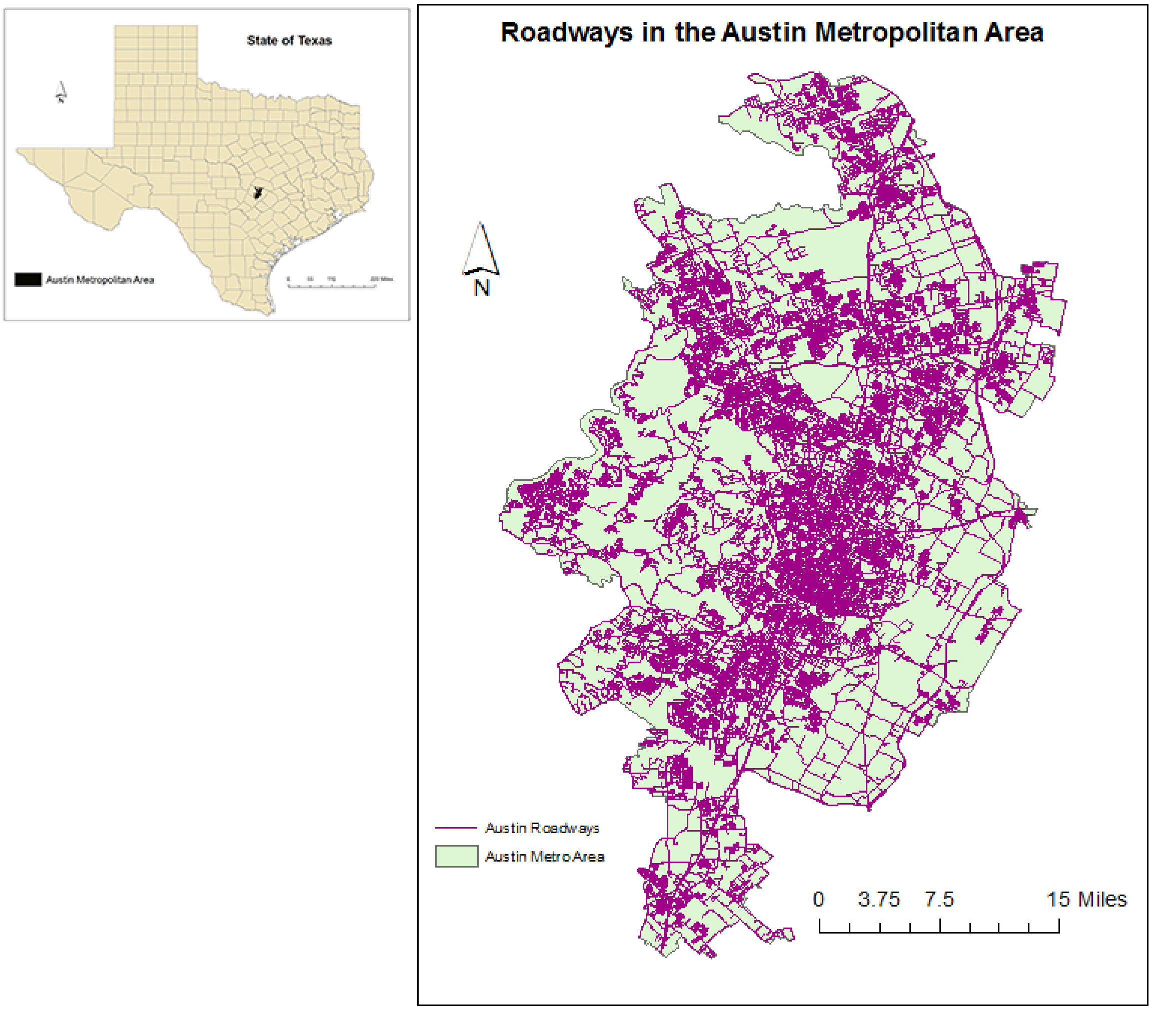

Figure 2. The state of Texas and the Austin Metropolitan Area with roadways.

<u>Approaching Research Question #1</u>

Descriptive statistics and additional details on the above-mentioned data are provided later in this section. For now, it is possible to present the general analytical framework/workflow adopted herein to answer the parts of research question #1 that were laid out in Chapter I. Recall that research question #1 was stated as follows:

1. What is the geographic distribution of traffic congestion in the greater Austin area?
    a. What is the relationship between traffic congestion and access to public transit?
    b. What is the relationship between traffic congestion and public transit usage?

The workflow for addressing these questions is illustrated in Figure 3. In summary, the analysis begins with a multiple regression model of ADT on several explanatory variables that are suggested by the literature (e.g., Mohamad et al., 1998; Anderson et al., 2006; Lowry and Dixon, 2012) and practical expectations. Because ADT is only available for a subset of road segments in the Austin study area (see below), the parameter estimates from that model are used to estimate ADT for *all* road segments for the most recent year of observed ADT data (i.e., 2014). Next, the same progression is followed for a second regression model, in which "Designed ADT" is the dependent variable. The TxDOT dataset provides information on Designed ADT for a large subset of road segments. As such, the parameters estimated from the second regression model are used to estimated Designed ADT for *all* road segments in the greater Austin area.

Following these two regression and extrapolation exercises, kernel density estimation (KDE) is used to generate two density surfaces with equal resolutions—one that corresponds to estimated ADT, and one that corresponds to estimated Designed ADT. KDE is a geospatial analytical technique that computes the density of linear features within a given "neighborhood" (Silverman, 1986). That is, (Designed) ADT for a given road segment is considered in the context of (Designed) ADT in its surrounding “neighborhood,” in order to generate a continuous surface of (Designed) ADT across the study area. Following the creation of these surfaces, map algebra is used to divide the cell values in the estimated ADT surface by the corresponding cell values in the estimated Designed ADT surface. This operation cancels out the originals units, leaving only a ratio of estimated ADT to estimated Designed ADT for the full extent of the study area. In other words, the result is a capacity ratio that can be mapped for the whole road network. Values greater than 1.0 indicate that, collectively and with respect to the surrounding street grid, the road segments within a given cell in the network are over intended capacity. Values less than 1.0 indicate a collection of road segments that are operating below capacity. Hence, values greater than 1.0 reasonably represent areas of the road network that are *congested*.

With this operational definition of congestion in hand, the ADT-to-capacity raster surface is used for two purposes (Figure 3). First, it can be reclassified into easier-to-interpret categories. For simplicity, five practical categories are chosen in this analysis: (1) below 50% of capacity; (2) between 50% and 100% of capacity; (3) between 100% and 150% of capacity; (4) between 150% and 200% of capacity; and (5) over 200% of capacity. The first two categories are therefore “under” and “approaching” capacity, while the final three correspond to increasingly problematic situations of congestion (or “over” capacity scenarios). From there, the five data classes can be vectorized, or converted into polygons. The purpose of vectorizing the raster data is to facilitate a spatial join between the TxDOT road segment spatial data layer and the congestion classification. Within a GIS environment, a spatial join takes two or more input layers that share a data model (e.g., vector) and geographic projection and coordinate system, and it combines them into a new layer that takes on attributes of the initial input layers (e.g., Bolstad, 2012). The resulting layer in this case is the set of all road segments in the greater Austin area, where each segment features an attribute with its respective traffic congestion category. The road segments can then be mapped to present a simple visual depiction of the geographic distribution of congestion in the Greater Austin area (research question #1).

Second, the ADT-to-capacity raster surface can be summarized by CBGs, where CBGs are the unit of analysis that feature in the EPA SLD and the U.S. Census ACS data. Using the Zonal Statistics tool in Esri's ArcGIS software, it is straightforward to aggregate ADT and Designed ADT to CBG boundaries. From there, the former aggregate measure is divided by the latter to create a CBG-level estimate of congestion. That measure can then be compared to measures of public transit access and usage to answer the two subquestions associated with research question 1—namely, (a) what is the relationship between traffic congestion and access to public transit; and (b) what is the relationship between traffic congestion and public transit usage?

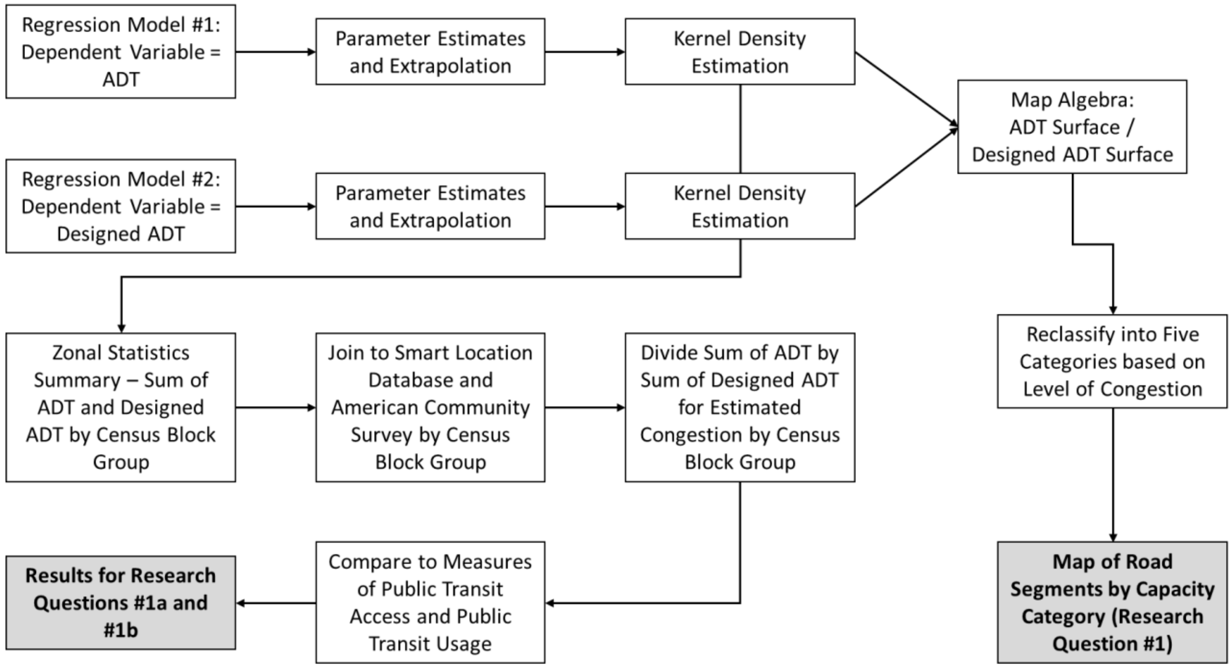

Figure 3. Workflow/analytical framework for answering research question #1.

Table 1 lists and defines all of the variables employed in the forthcoming analyses according to the source from which they were collected. Three tables found in the remainder of this section—Table 2 to 4—provide descriptive statistics on those variables for the different analytical samples that are used herein. The following subsections discuss each of these samples in detail.

Table 1. A list of all variables for this analysis and their sources.

Variable	Description	Source
Average Daily Traffic (ADT)*†	Average daily traffic (ADT) represents the total traffic for a year divided by 365, or the average traffic volume per day.	TxDOT
Number of lanes	The amount of lanes in a particular roadway segment	TxDOT
ADT year‡	The year in which Average Daily Traffic (ADT) was last collected in a particular roadway segment	TxDOT

Table 1-Continued		
Designed ADT††	Designed average daily traffic represents the total traffic for a year divided by 365, or the average traffic volume per day that a roadway was designed to handle	TxDOT
Road functional class	The process by which streets and highways are grouped into classes, or systems, according to the character of traffic service that they are intended to provide. This data set contains the following: local road, major collector, minor arterial, principal arterial, urban freeway, and interstate.	TxDOT
Distance from CBD	Distance from a road segment to the Austin central business zoning district, in kilometers	Computed in GIS
Surrounding population*	Sum of total population in all census block groups (CBGs) that are intersected by the given road segment	EPA SLD
Surrounding employment*	Sum of total employment (jobs) in all CBGs that are intersected by the given road segment	EPA SLD
Distance from population-weighted centroid to nearest transit stop	The minimum walk distance between the population weighted CBG centroid and the nearest transit stop.	EPA SLD
Aggregate frequency of transit trips per square mile	Measures transit frequency per square mile of land area	EPA SLD
% Minority population	Fraction of total population in a CBG classified as non-white	Census ACS
% Adults without a high school diploma	Fraction of total population 25 years or over in a CBG without a high school education	Census ACS
Unemployment rate	Fraction of unemployed civilian workers relative to the civilian labor force in a CBG	Census ACS
% Households on public assistance	Fraction of all households in a CBG that receive public assistance income	Census ACS
% Renters	Fraction of occupied housing units that are renter-occupied in a CBG	Census ACS

Table 1-Continued		
% House burdened	Fraction of all occupied housing units in a CBG (renter and owner combined) for which householders spend more than 30% of their gross monthly income on housing	Census ACS
% Poor or struggling	Fraction of all persons for whom poverty status is determined with an income-to-poverty ratio below 2.0	Census ACS

*Indicates that variable was log-transformed for regression analysis. For total population and total employment, the precise transformation was log(x+1), where x is the original value of the variable. This choice was made because a handful of road segments took on 0 values for these variables (see Quinn and Keough, 2002); †indicates dependent variable in regression model 1; ††indicates dependent variable in regression model 2; ‡There were only two road segments with an ADT Year of 2013. In the analysis, these two observations were combined with the ADT data for the prior year, 2012.

Note: shaded cells refer to variables used for research question #2

TxDOT Data and Regression Model Samples

A request made to the Texas Department of Transportation (TxDOT) yielded a geographic dataset, in Esri shapefile format, containing 23,939 road segments in the greater Austin area (Figure 2). As mentioned above, among the road segment attributes included in the TxDOT dataset are ADT, ADT Year, and Designed ADT. Recall that ADT is a count variable that measures average traffic volume on a given road segment over the course of a calendar year. The ADT Year variable, then, provides information on the precise calendar year for which the ADT variable was measured. For the majority of road segments (n=15,701, or 66% of all segments), the reported ADT Year is 1978.

Within this set of observations, virtually all segments that share the same Functional Class (Table 1) take on identical ADT values. Thus, the 1978 measures are not likely to be reliable, particularly insofar as Austin has experienced tremendous population growth since that time. Of the remaining 8,238 road segments, then, fewer than one percent (n=70) are associated with ADT Year values between 2000 and 2007. On the other hand, more than 99% of those segments (n=8,168) have recent ADT Year values, between 2010 and 2014. For these reasons, the road segments with ADT values for this more recent five-year period—i.e., 2010 to 2014—were selected as the sample for designing the first regression model described above in Figure 3. In a related fashion, only 34% (n=8,232) of the 23,939 road segments in the dataset have nonzero values for the Designed ADT attribute. As such, this sample of road segments is used to design the second regression model mentioned in Figure 3.

With those caveats in mind, for any given road segment, both ADT and Designed ADT are hypothesized to vary as functions of: (1) its number of lanes, (2) its functional classification (see Table 1), (3) its distance from Austin's CBD, (4) an interaction between its functional classification and its distance from Austin's CBD, (5) the size of the surrounding population, and (6) the total number of jobs in the surrounding area. With respect to item (4), the [multiplicative] interaction term is simply an acknowledgement that certain types/classes of roads—e.g., freeways—will have different relationships with traffic volume depending on how near they are to Austin's urban core. Roadways that are nearer to the CBD may have different ADT volumes compared to the same classes of roadways farther from the CBD. Finally, for the ADT model, but not the Designed ADT model, we add (7) time dummy variables, which take on a value of 1 if ADT was recorded in a given year (e.g., 2011) and a value of 0 otherwise. Descriptive statistics for the ADT and Designed ADT regression samples are presented in Table 2 and Table 3, respectively. Whereas variables (1) through (4), as well as (7) can all be derived from the TxDOT data, variables (5), (6), and several other variables come from the U.S. Environmental Protection (EPA) Smart Location Database (SLD).

Table 2. Descriptive Statistics for the ADT Model.

Variable	Mean	Standard Deviation
ADT†	6.982	2.251
Number of Lanes	2.576	1.175
Distance from CBD (km)	19.167	11.076
Surrounding Population†	7.925	0.906
Surrounding Employment†	6.334	1.613
ADT Year = 2010*	0.571	n/a
ADT Year = 2011*	0.183	n/a
ADT Year = 2012/13*‡	0.003	n/a
ADT Year = 2014*	0.243	n/a
Road Functional Class = Local Road*	0.520	n/a
Road Functional Class = Major Collector*	0.245	n/a
Road Functional Class = Minor Arterial*	0.083	n/a
Road Functional Class = Principal Arterial*	0.089	n/a
Road Functional Class = Urban Freeway*	0.044	n/a
Road Functional Class = Interstate*	0.018	n/a
n=8,168		

†Indicates variable was log-transformed for regression analysis (see Quinn and Keough, 2002); ‡There were only two road segments with an ADT Year of 2013. In the analysis, these two observations were combined with the ADT data for the prior year, 2012; *indicates dichotomous variable (mean = proportion of sample)

Table 3. Descriptive Statistics for the designed ADT Model.

Variable	Mean	Standard Deviation
Designed ADT†	6.977	2.253
Number of Lanes	2.585	1.179
Distance from CBD	19.219	11.102
Surrounding Population†	7.930	0.905

Table 3-Continued		
Surrounding Employment†	6.344	1.618
Road Functional Class = Local Road*	0.506	n/a
Road Functional Class = Major Collector*	0.256	n/a
Road Functional Class = Minor Arterial*	0.085	n/a
Road Functional Class = Principal Arterial*	0.091	n/a
Road Functional Class = Urban Freeway*	0.044	n/a
Road Functional Class = Interstate*	0.017	n/a
n=8,232		

†Indicates variable was log-transformed for regression analysis (see Quinn and Keough, 2002);

*indicates dichotomous variable (mean = proportion of sample)

EPA SLD Data: Public Transit Access

The U.S. EPA SLD "is a nationwide geographic data resource for measuring location efficiency" (EPA, n.d.). SLD data are collected and provided at the census block group (CBG) level of analysis. The CBG-level variables that were extracted from the SLD for this thesis (refer to Table 1) include: (1) total population, (2) total employment/number of jobs, (3) distance from the CBG's population-weighted centroid to the nearest public transit stop, and (4) the aggregate frequency of public transit service (i.e., number of transit trips) per square mile within the CBG. Variables (1) and (2) feature in the ADT and Designed ADT regression models, as discussed above. On the other hand, variables (3) and (4) are used to derive a measure of *access* to public transit in Austin. That being said, the Capital Metro public transit service from which variables (3) and (4) are derived (see Ramsey and Bell, 2014) is not available throughout the greater Austin area. Specifically, CBGs outside of the Austin city limits do not receive the vast majority of Capital Metro services. Consequently, our measurement and analysis of public transit access will be limited to only those CBGs with (a) centroids that fall within the Austin city limits, and (b) non-missing data for both variables (3) and (4) listed above.

Having said that, there are a total of 814 CBGs that intersect with the road network pictured that is in Figure 1. Of these, 399 have valid, non-missing data entries for both of the transit-related variables mentioned in the preceding paragraph. In the interest of comparing proverbial “apples to apples,” note that of these 399 CBGs, 385 have centroids that lie within Austin’s city limits, while the remaining 14 fall mostly or entirely outside of the city. Accordingly, it is the 385 city CBGs that form the sample for analyzing the relationship between congestion and public transit *access* (research question #1a). The next task, then, is to operationalize the concept of *access* using the two aforementioned SLD variables. To do so, we begin from the proposition that *access* is a function of both proximity and functionality (e.g., Bullard 2008). With respect to the former, the SLD variable *distance from a CBG’s population*-*weighted center to the nearest transit stop* (hereafter the *distance* variable) is a useful proxy for proximity, insofar as it captures the spatial distance between where people live and the nearest transit access point (Ramsey and Bell, 2014). Nevertheless, proximity alone cannot be a surrogate for access. For instance, it could be the case that a transit stop is within walking distance to most people in a given CBG; but that there are very few trips which depart from or stop at that location. In other words, the transit stop might not be very functional. Along those lines, the SLD variable *aggregate frequency of transit trips in a CBG per square mile* (hereafter the *frequency* variable) is proposed here as a proxy for functionality.

Thus, only where transit stops are simultaneously proximate to the people within a CBG and functional are they considered veritably *accessible*. Because both dimensions—nearness and functionality—are therefore [relatively] equally important in this conception of accessibility, they ought to be combined in a way that gives them equal weight in a composite measure/index of *public transit access*. Following the method used by the United Nations (UN) to compute its Human Development Index (United Nations, 2015), Weaver et al. (2016) argue that collapsing multiple, equally important variables into a single composite index can be achieved through the calculation of a geometric mean. A geometric mean is a product-based average, as opposed to the conventional arithmetic mean, which is additive. In that sense, such a mean captures compounding, or interrelationships among its constituent parts (Weaver et al., 2016). In the present case, the *frequency* (i.e., *functionality*) variable relates positively to our conceptualization of *access*. By contrast, the *distance* (i.e., *proximity*) variable is inversely related to *access*—as distance increases, transit stops become less accessible. Therefore, the proximity variable must be reverse-coded prior to its inclusion in the geometric mean. With these points in mine, we define *public transit access* here as:

$$Access = \sqrt{P \times F}\ , \qquad \text{(Eq. 1)}$$

where *P* and *F* are, respectively, indices of *proximity* and *functionality* that range from 0 to 1 in the limit, where values near 0 indicate low accessibility and values near 1 indicate high accessibility. *P* is computed as:

$$P = 1 - \frac{d_i - \min(d)}{\max(d) - \min(d)}, \qquad \text{(Eq. 2)}$$

where *d_i* is the value of the *distance* variable for CBG *i*, and min(*d*) and max(*d*) are the minimum and maximum values of the *distance* variable in the sample. Note that the final term on the right hand side of the equation is subtracted from 1, so that higher distances lead to smaller values of *P*. And *F* is computed as:

$$F = \frac{f_i - \min(f)}{\max(f) - \min(f)}, \qquad \text{(Eq. 3)}$$

where *f_i* is the value of the *frequency* variable for CBG *i*, and min(*f*) and max(*f*) are the minimum and maximum values of the *frequency* variable in the sample.

Table 4 presents descriptive statistics for the SLD variables described in this subsection, as well as for our adopted index of *access*.

Table 4. Descriptive Statistics for access variables.		
Variable	**Mean**	**Standard Deviation**
Distance from population-weighted centroid to nearest transit stop (km)	0.434	0.279
Aggregate frequency of transit trips per square mile	747.585	1,225.606
P index (see Eq. 2)	0.637	0.233
F index (see Eq. 3)	0.069	0.113
Access	0.180	0.141
n=385		

U.S. Census American Community Survey Data: Public Transit Usage

To round out the data collection for all parts of research question #1, commuting data were collected from the most recent five-year (2010-2014) vintage of the U.S. Census American Community Survey (ACS). The ACS is a rolling survey that takes place each year. However, data are provided at the CBG level only for five-year period increments rather than annually in order to increase the reliability of the estimates. Within the data reported by the ACS, Table B08301 contains data on “means of transportation to work.” Among the modes of transportation for which data are reported is “public transportation (excluding taxicab)". The universe to which the reported data applies is the population of all workers 16 years of age or older.[1] Hence, by dividing the number of workers in each block group who commute using "public transportation (excluding taxicab)" by the total number of workers 16 years or older, one can obtain the fraction of workers in any given CBG who use public transit. For the 385 CBG sample described in the preceding subsection, the mean fraction of workers who use public transit is 0.054 with a standard deviation of 0.063.

Correlation Analysis for Research Question #1

While the road network, population, and employment data can described above be used to map the geography of congestion across the entire greater Austin region pictured in Figure 1, the SLD data limit our investigation of the relationships between congestion, public transit access, and public transit usage to CBGs within the Austin city limits exclusively. For the sample of 385 CBGs within Austin that contain all relevant public transit data, correlation analysis is used to study these relationships. First, bivariate correlations are derived for each pairwise relationship between these three variables. Second, the partial correlation is found between public transit usage and traffic congestion, *controlling for public transit access*. The latter method is employed to test the hypothesis that for a given level of access, public transit usage might contribute to a reduction in traffic congestion.

[1] http://censusreporter.org/tables/B08301/

Approaching Research Question #2

Recall from Chapter I that research question #2 was stated as:

2. Are traffic problems and access to public transit distributed equitably between socioeconomic groups in the city of Austin?

Two analytical procedures are essential for addressing this question. First, it is necessary to classify the census block groups (CBGs) in the city of Austin into socioeconomic status (SES) groups. Such a classification can be achieved in a straightforward manner through a multivariate k-means clustering analysis. K-means clustering is used to create groups such that observations within groups have low variability in their attributes, while there is relatively high between group variation in attributes (e.g., Weaver, 2015). Second, tests for equality of mean (or median, if appropriate) traffic congestion and public transit access between the derived SES groups allow for an assessment of the degree to which mobility-depressing traffic congestion and/or mobility-enhancing public transit access are (un)equally distributed across SES groups. The next subsection describes the demographic and socioeconomic status variables that were selected for the first (k-means cluster analysis) procedure.

Demographic and Socioeconomic Variables from the U.S. Census ACS

In addition to the public transit usage data that were obtained from the most recent U.S. American Community Survey (ACS) for research question #1 (see above), several demographic and socioeconomic variables were extracted from the ACS to facilitate the clustering of CBGs into SES groups. Drawing on literature that describe similar analytical exercises (e.g., Sampson et al., 2002; Manturuk et al., 2009; Weaver et al., 2016), the variables listed and described in the shaded portion of Table 1 were selected for this purpose.

Summary

In summary, this thesis is interested in the: (1) geographic distribution of traffic congestion in greater Austin, (2) relationships between traffic congestion, public transit access, and public transit usage in the city of Austin, and (3) the extent to which these variables are evenly distributed among socioeconomic groups in the city of Austin. The next chapter presents findings on these matters derived by carrying out the analytical operations described throughout this chapter.

IV. RESULTS AND PRELIMINARY DISCUSSION

Regression Results and the Geography of Traffic Congestion in Austin

The results from estimating the two regression models described in Figure 3 are presented in Table 5 and Table 6 below. The results for the first model are highly consistent with existing literature (e.g., Mohamad et al., 1998; Anderson et al., 2006). Namely, Average Daily Traffic (ADT) is directly and statistically significantly related to *number of lanes*, *surrounding population*, and *surrounding employment*. Moreover, as expected, *distance from the Austin CBD* is inversely related to ADT, suggesting that traffic eases the farther a road segment is from the urban core. ADT is higher in every year relative to the reference year of 2010, which implies that traffic has been on the rise as the Austin region continues to explode in population (e.g., Weissman, 2015). Likewise, relative to the reference *road functional classification* category of interstate, all other types of road segments experience lower traffic volumes. Two significant interaction effects between *road functional classification* and *distance from the Austin CBD* are also detected in the model. Overall, the R-squared value of the model is 0.85, which is consistent with or outperforms the findings from related literature (see the review by Lowry and Dixon, 2012).

Table 5. Regression Output for ADT Model.

Variable	**Coefficient**	**Standard Error**	
Number of Lanes	0.339	0.013	***
Distance from CBD	-0.017	0.005	**
Surrounding Population†	0.042	0.011	***

Table 5-Continued			
Surrounding Employment†	0.071	0.007	***
ADT Year = 2011	1.396	0.045	***
ADT Year = 2012/13	0.759	0.199	***
ADT Year = 2014	1.910	0.045	***
Road Functional Class = Local Road	-2.749	0.142	***
Road Functional Class = Major Collector	-1.398	0.137	***
Road Functional Class = Minor Arterial	-0.718	0.144	***
Road Functional Class = Principal Arterial	-0.675	0.137	***
Road Functional Class = Urban Freeway	-0.234	0.153	
Local Road*Distance from CBD	-0.002	0.005	
Major Collector *Distance from CBD	-0.007	0.005	
Minor Arterial *Distance from CBD	-0.012	0.006	*
Principal Arterial *Distance from CBD	0.002	0.006	
Urban Freeway *Distance from CBD	-0.031	0.008	***
Constant	6.907	0.175	***
n=8,168			
R-squared: 0.885			
Adjusted R-squared: 0.846			
Dependent variable: ADT†			

†Indicates variable was log-transformed; ***p<0.001; **p<0.010; *p<0.050; .p<0.100

As discussed above (see especially the workflow pictured in Figure 3), the parameter estimates reported in Table 5 were then used to predict ADT values for *all* road segments in the greater Austin region from Figure 2. As part of the prediction procedure, the *ADT Year* variable was set to 2014 for all road segments in the TxDOT dataset, in order to estimate contemporary levels of traffic volume throughout the greater Austin road network. The means and 95% confidence intervals of these predicted values, separated by *road functional class*, are presented in Table 3. They are also shown graphically in the left panel of Figure 4. Expectedly, interstates and other major roadways are characterized by the highest mean predicted ADT values, while local and minor roads are predicted to experience lower daily traffic volumes.

Table 6. Descriptive Statistics for Predicted ADT Variable.				
Road Functional Class	**Mean**	**Standard Deviation**	**Lower Bound of 95% Confidence Interval**	**Upper Bound of 95% Confidence Interval**
Interstate	81,068.95	55,515.92	72,001.35	90,136.55
Local Road	1,299.53	337.16	1,294.83	1,304.23
Major Collector	6,297.70	2,993.69	6,172.60	6,422.80
Minor Arterial	17,420.84	11,129.48	16,598.11	18,243.56
Principal Arterial	30,590.51	15,250.94	29,499.01	31,682.00
Urban Freeway	45,155.41	34,767.36	41,583.69	48,727.12

n=23,939 (all road segments in TxDOT dataset)

The results for the *Designed ADT* regression model, which are presented in Table 8, also match expectations. *Number of lanes* and *surrounding employment* are directly and statistically significantly correlated with *Designed ADT*, while *distance from the Austin CBD* is inversely related to *Designed ADT*. Relative to the reference road classification of interstate, all other road types are predicted to have been designed for lower traffic capacities. Similar to the ADT model, three significant interaction effects exist between *road functional class* and *distance from the CBD*.

Table 7. Regression Output for Designed ADT Model.			
Variable	**Coefficient**	**Standard Error**	
Number of Lanes	0.376	0.015	***
Distance from CBD	-0.015	0.006	*
Surrounding Population†	0.013	0.013	
Surrounding Employment†	0.059	0.008	***
Local Road	-4.592	0.156	*
Major Collector	-1.852	0.154	***
Minor Arterial	-1.347	0.161	***
Principal Arterial	-1.096	0.155	***
Urban Freeway	-0.091	0.176	
Local Road*Distance from CBD	-0.005	0.006	
Major Collector *Distance from CBD	-0.012	0.006	.

Table 7-Continued			
Minor Arterial *Distance from CBD	0.003	0.007	
Principal Arterial *Distance from CBD	0.021	0.007	**
Urban Freeway *Distance from CBD	-0.028	0.009	**
Constant	9.326	0.193	***
n=8,232			
R-squared: 0.810			
Adjusted R-squared: 0.810			
Dependent variable: Designed ADT†			

†Indicates variable was log-transformed in the regression model; ***p<0.001; **p<0.010; *p<0.050; .p<0.100

Following the workflow pictured in Figure 3 from Chapter II, the parameter estimates reported in Table 7 were used to predict *Designed ADT* for all 23,939 road segments in the TxDOT dataset. The means and 95% confidence intervals of these predicted values, separated by *road functional class*, are presented in Table 8. As above, they are also shown graphically in the right panel of Figure 4. Expectedly, interstates and other major roadways are characterized by the highest mean predicted Designed ADT values, while local and minor roads are predicted to have been designed for lighter traffic volumes.

Table 8. Descriptive Statistics for Predicted Designed ADT Variable.

Road Functional Class	Mean	Standard Deviation	Lower Bound of 95% Confidence Interval	Upper Bound of 95% Confidence Interval
Interstate	127,787.33	91,876.37	112,780.85	142,793.80
Local Road	268.37	69.73	267.40	269.34
Major Collector	5,116.90	2,645.83	5,006.34	5,227.46
Minor Arterial	16,117.56	8,935.67	15,457.01	16,778.11
Principal Arterial	36,515.14	17,631.05	35,253.30	37,776.98
Urban Freeway	84,029.27	68,278.42	77,014.90	91,043.64

n=23,939 (all road segments in TxDOT dataset)

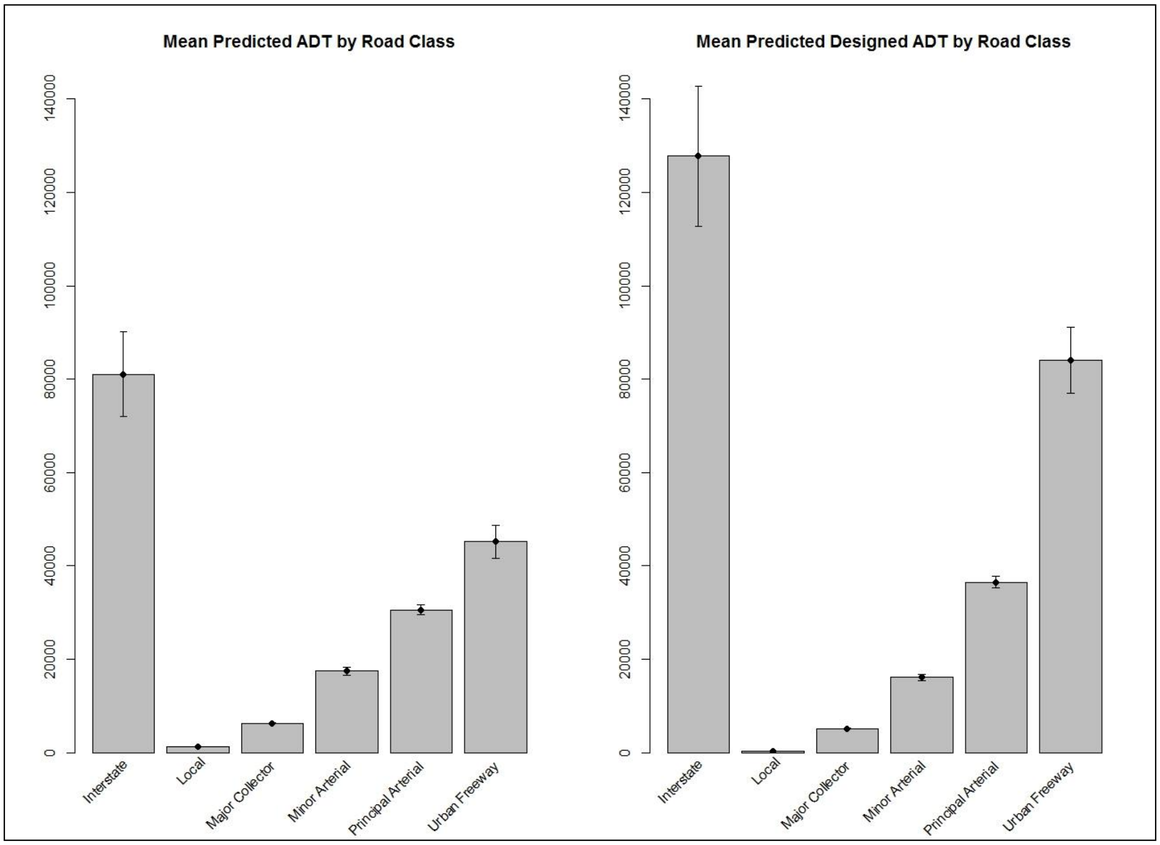

Figure 4. Mean predicted ADT and mean predicted Designed ADT by road class.

With extrapolated values for *ADT* and *Designed ADT* now available for all 23,939 road segments from the TxDOT dataset, the next step in the workflow (Fig. 3) for this project is to use kernel density estimation (KDE) to compute the density of predicted traffic volume (*ADT*) and predicted traffic capacity (*Designed ADT*) across the full spatial extent of the road network in greater Austin. KDE is adopted insofar as "density maps are particularly useful for looking at patterns rather than at the locations of individual) features, and for mapping areas of different sizes" (Mitchell, 1999: 70). Mitchell (1999: 70) continues to say that:

> "while [one] can see concentrations [e.g., of traffic volume] by simply mapping the locations of features, in areas with many features it may be difficult to see which areas have higher concentration than others. A density map lets [one] measure the number of features using a uniform areal unit…[to] clearly see the distribution [across an entire study area]".

In other words, density maps view traffic on road segments as existing within larger neighborhoods. Thus, the interest is not simply in the predicted *ADT* and *Designed ADT* values for each individual road segment. Rather, the interest is in how *ADT* and *Designed ADT* vary spatially within their broader geographic surroundings. That being said, KDE was applied to the 23,939 road segments using, first, *predicted ADT* as the population field of interest; and, second, *predicted Designed ADT* as the field of interest. In both cases, the uniform cell resolution was set to 195 meters, which was suggested for the dataset by a default cell size calculator algorithm in ArcGIS. The resulting two surfaces were then classified by ArcGIS's built-in stretching function using one standard deviation specification.[2] The *predicted ADT* surface is shown on the left-hand-side of Figure 5, while the *predicted Designed ADT* surface is on the right of the same figure. The figure further illustrates that the ArcGIS map algebra tool was used to divide the former by the latter, in order to create a unitless ratio of *predicted traffic volume* to *predicted traffic capacity*. The surface that corresponds to that unitless ratio is mapped in Figure 6.

[2] https://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/stretch-function.htm

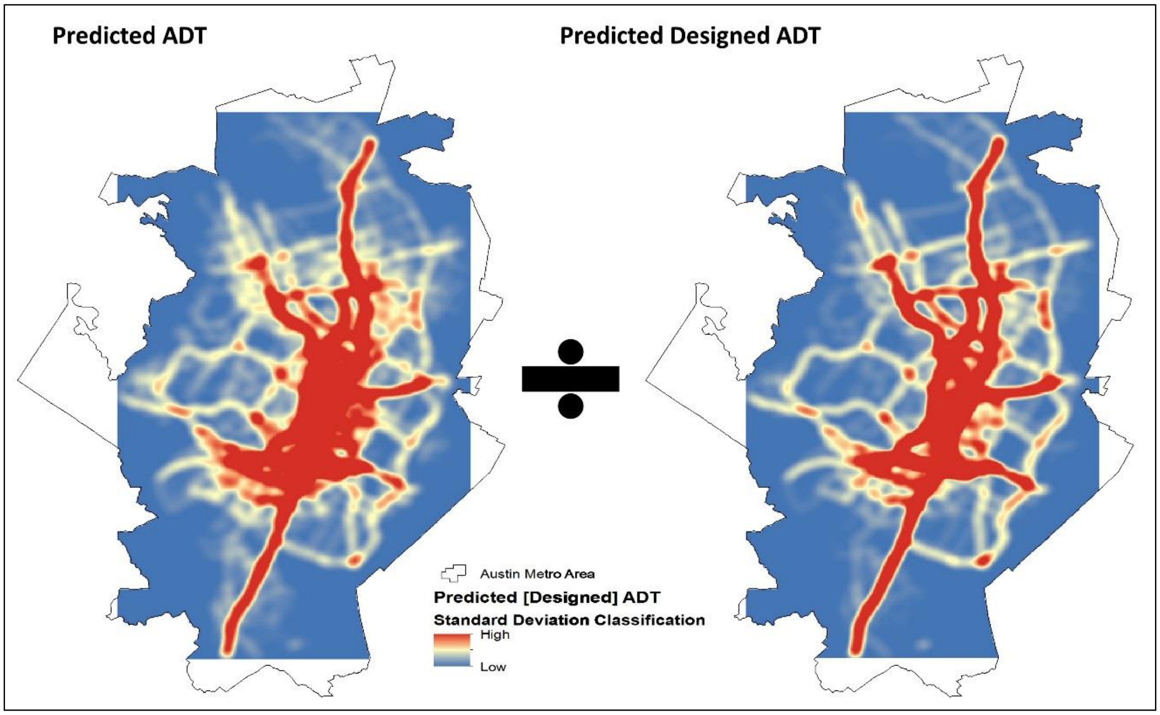

Figure 5. Map Algebra formula used to predict ratio of Predicted ADT to Designed ADT.

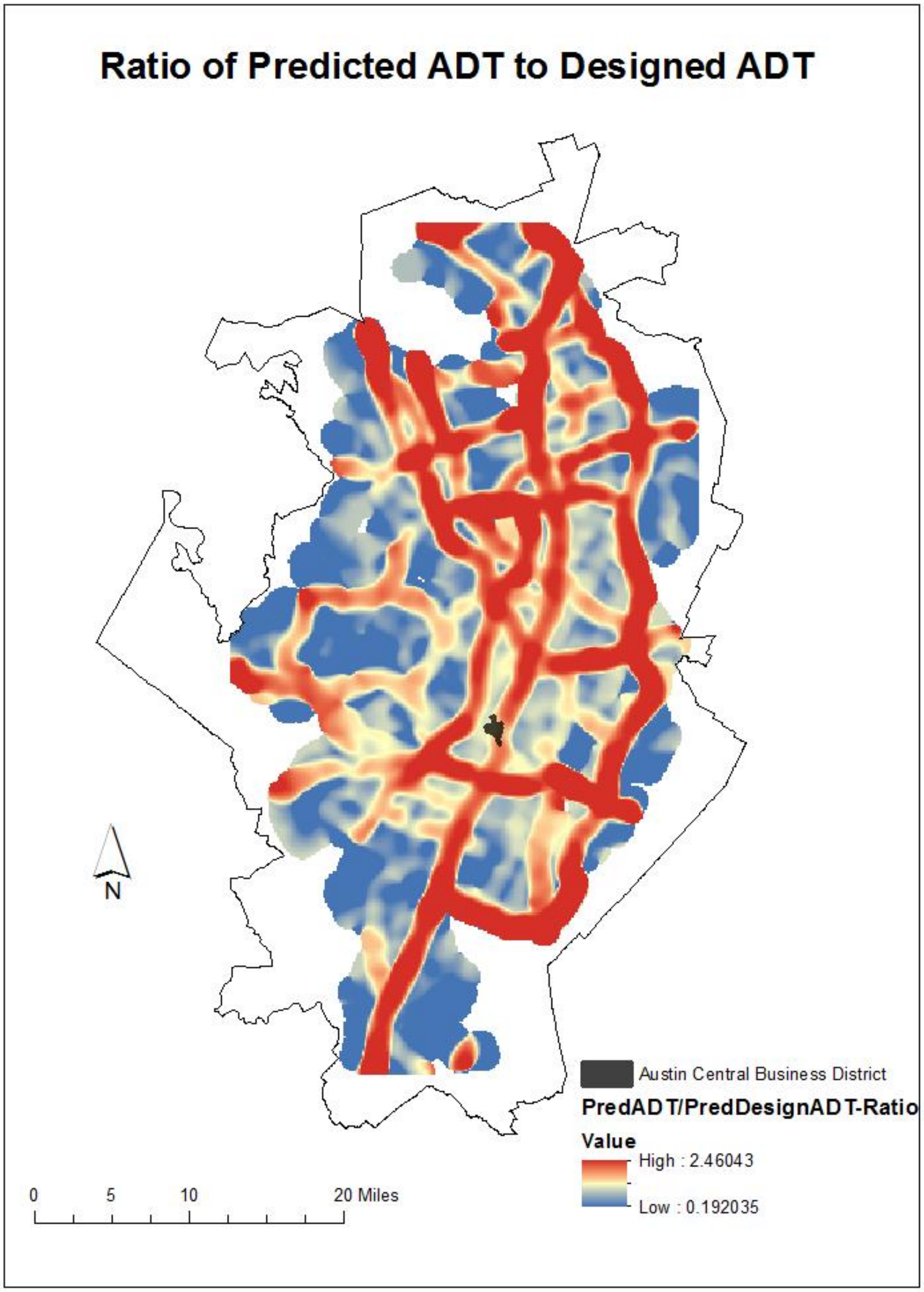

Figure 6. A map of the unitless ratio of Predicted ADT to Designed ADT.

As a next step, the continuously varying values of the traffic-to-capacity ratio pictured above in Figure 6 were reclassified into five discrete values, which were chosen by the researcher. Of the five categories, two describe locations in the road network where *predicted ADT* is less than *predicted Designed ADT*. In other words, such areas are considered to be “below capacity.” The remaining three classifications describe increasingly problematic cases in which areas along the road network are “above capacity,” or such that *predicted ADT* exceeds *predicted Designed ADT*. The five classes are pictured below in Figure 7.

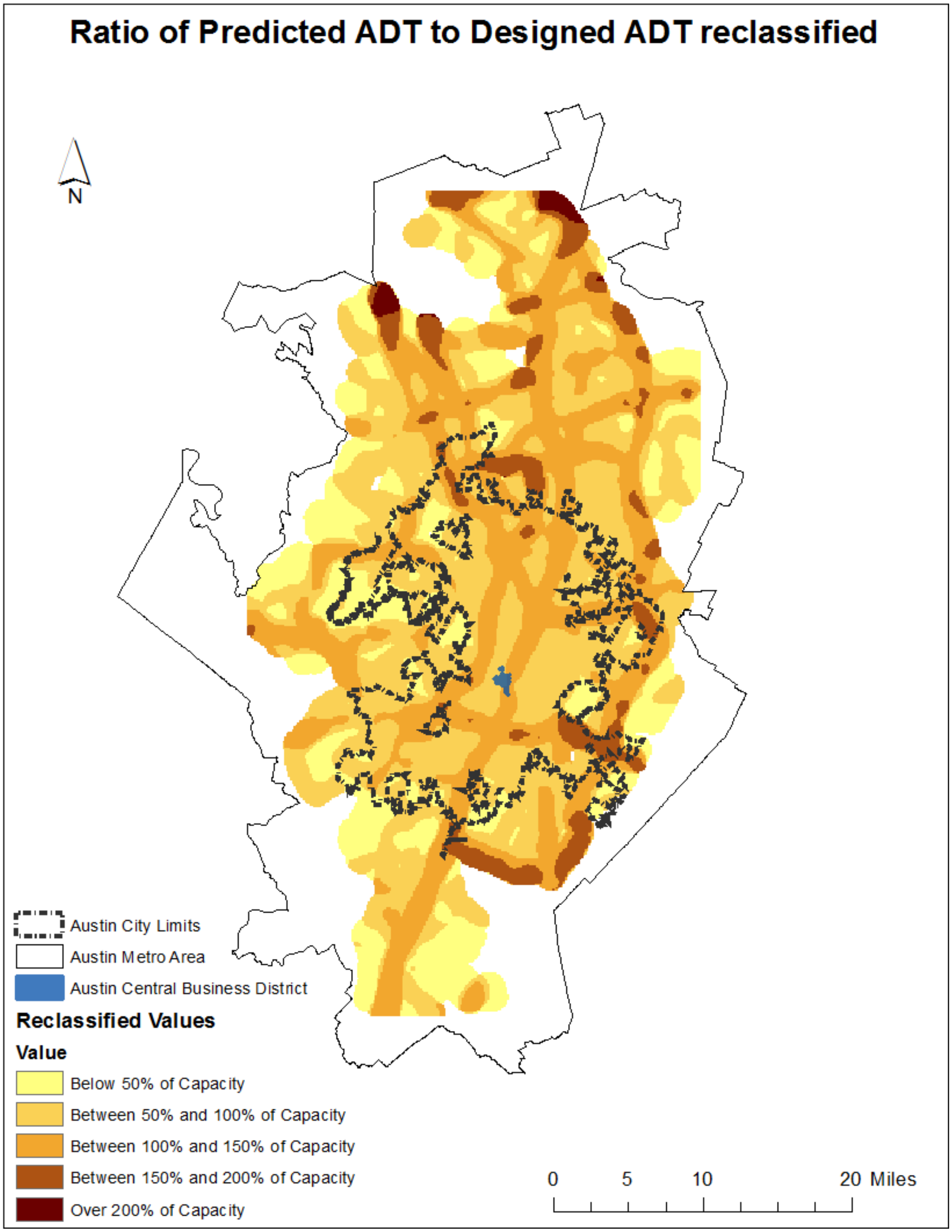

Figure 7. Ratio of Predicted ADT to Designed ADT, reclassified.

The reclassified figure shown above allows for an easier interpretation of the resulting patterns. In short, some of the largest traffic-to-capacity ratios—i.e., the areas of the network that are experiencing the operational definition of *congestion* defined earlier in this thesis—tend to exist at entry points to the city of Austin. Major roadways that move cars from surrounding suburban communities into and out of the Austin city limits are predicted to experience the greatest daily capacity issues. Commuter areas in the northern portion of the greater Austin area also experience severe capacity issues at the borders of the region. These capacity issues first decrease along major roadways from the northern boundaries of the region toward Austin; but they increase again at the city limits.

Whereas the reclassified raster surface pictured in Figure 7 provides a concise, general summary of daily traffic capacity issues in greater Austin, the analysis can be made more specific. In particular, each individual road segment is capable of being classified on the basis of its location in one of the five traffic-to-capacity categories visualized above. To achieve this objective, the reclassified raster surface must be converted into a vector (polygon) GIS data layer. This step is necessary to facilitate a spatial join, wherein road segments are given the attributes—here, the traffic capacity category—of the new vector data layer. Such a join cannot be facilitated between the vector (line-based) road segment GIS data layer and a raster data layer. The results of this join operation are pictured in Figure 8, where all road segments are symbolized on the basis of the traffic-to-capacity category in which they fall. This figure presents an even clearer picture of the notion that major roadways that connect suburban communities to the Austin city limits experience capacity issues, particularly where they act as entry points into the city.

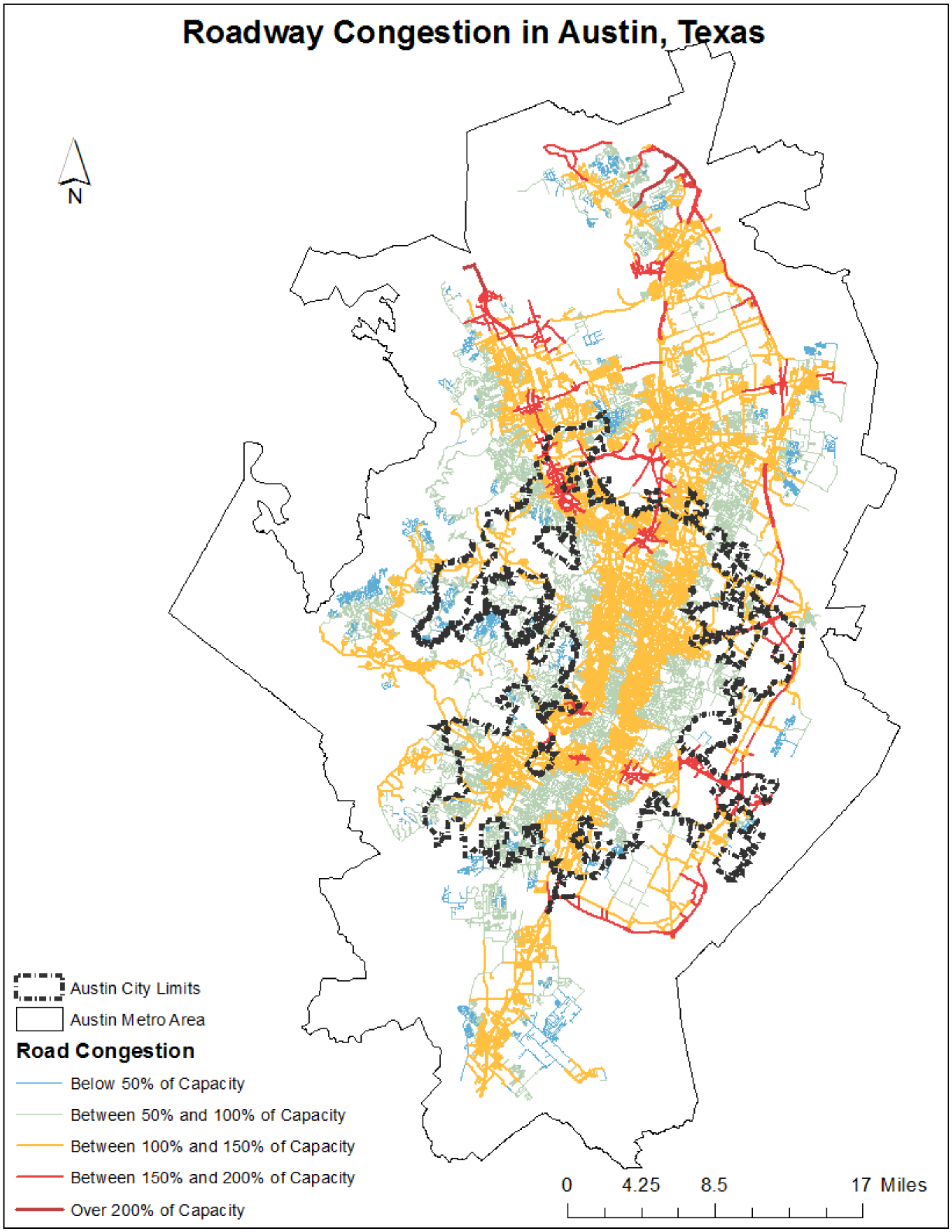

Figure 8. A map of road congestion in Austin for every street in the dataset.

Correlation between Congestion, Transit Access, and Transit Usage

The final figure presented in the preceding subsection (Figure 8) offers several interesting insights with respect to the first part of research question #1—namely, the geographic distribution of average daily traffic congestion in the greater Austin area. To explore the subparts of this question—i.e., those that examine relationships between congestion, public transit access, and public transit usage—it is necessary to keep in mind that *public transit access* and *public transit usage* were measured at the census block group (CBG) level of analysis. Thus, prior to addressing the subquestions of research question #1, it is first required that *traffic congestion* be measured at this same level. Within Esri's ArcGIS software, the Zonal Statistics tool allows users to compute various statistics for polygon "zones," such as CBGs, from raster surfaces. For this project, the Zonal Statistics tool was used to sum the (1) *predicted ADT* values and (2) *predicted Designed ADT values*, both shown in Figure 5, for each CBG. Hence, for each CBG, we obtained the aggregate *traffic volume* (sum of *predicted ADT*) and aggregate *traffic capacity* (sum of *predicted Designed ADT*). Next, the former value can be divided by the latter to create a ratio of *traffic-to-capacity* for each CBG in the greater Austin study area. Figure 9 visualizes these ratios using a simplified, three-category classification scheme. The mean of the mapped ratio variable for the 814 CBGs in greater Austin is 1.11, with a standard deviation of 0.42.

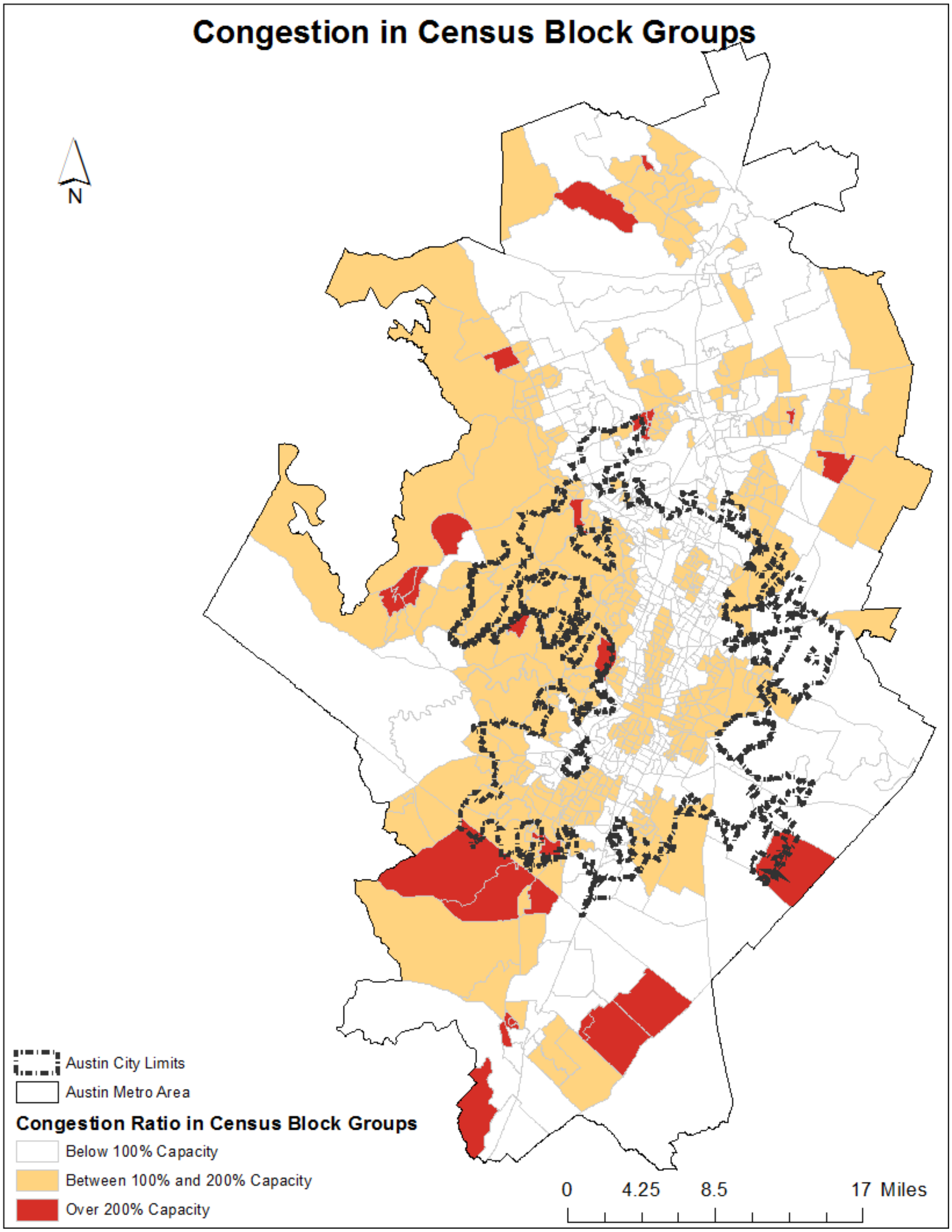

Figure 9. Congestion in Census Block Groups.

Recall now that the public transit data described above were only available for 385 CBGs that fall within Austin's city limits. Accordingly, to explore relationships between *traffic congestion*, *public transit access*, and *public transit usage*, correlation analysis was performed exclusively on this sample of 385 CBGs. As described above, the first step in answering the two subparts of research question #1 involved computing bivariate Pearson correlations between the variables of interest. These correlations are reported in Table 9.

Table 9. Bivariate Correlations.

Research Question	Variable #1	Variable #2	Pearson Correlation Coefficient	p-value
#1a	Traffic Congestion	Transit Access	+0.174	0.001**
#1b	Traffic Congestion	Transit Usage	-0.054	0.293

**p<0.01

Concerning the relationship between *traffic congestion* and *transit access* (research question #1a), the bivariate Pearson correlation coefficient is +0.174 and statistically significant at a 99% level of confidence. With respect to the relationship between *traffic congestion* and *transit usage* (research question #1b), the Pearson correlation coefficient is -0.054 and not statistically significant. These results have two immediate takeaways. First, the significant *positive* relationship between congestion and transit access on its face appears counterintuitive. Indeed, one might reason that better access to public transit ought to reduce traffic congestion. Second, public transit *usage* does appear to have a negative relationship with traffic congestion, as one might expect. However, the relationship is relatively weak in magnitude and does not achieve statistical significance. In essence, then, the bivariate correlation results appear to undermine—or, minimally, do not support—the notion that public transit can be an effective means for alleviating traffic congestion issues in the city of Austin.

Recall from Chapter II, though, that public transit usage was hypothesized to contribute to a reduction in traffic congestion *for a given level of transit access*. In other words, the non-significant relationship between congestion and transit usage from Table 9 does not account for the underlying variation in our adopted index of *transit access*. Thus, a more appropriate method for analyzing this relationship is arguably a partial correlation analysis, in which the Pearson correlation is computed for *traffic congestion* and *transit usage* while controlling for *transit access*. Using this method, the correlation coefficient between congestion and transit usage increases in magnitude to -0.108, and achieves statistical significance at a 95% level of confidence (p=0.035). Although this magnitude is still relatively weak, the finding suggests that when access is accounted for, the greater the fraction of the population that uses public transit, the less problematic are the issues of daily traffic congestion in the city of Austin.

k-Means Cluster Analysis Results for Grouping Austin CBGs by Socioeconomic Status

In order to assess equity in the distributions of both mobility-constraining traffic congestion problems and mobility-enhancing public transit access on the basis of socioeconomic status (SES), k-means cluster analysis was used to classify census block groups (CBGs) on various demographic and socioeconomic dimensions. Drawing on existing literature (e.g., Sampson et al., 2002; Manturuk et al., 2009; Weaver et al., 2016), the seven variables were used for this purpose: (1) minority population; (2) adults without a high school diploma or equivalency degree; (3) unemployment rate; (4) households on public assistance; (5) renter-occupied housing units; (6) cost-burdened households; and (7) size of the poor and struggling population (see Table 1 from Chapter II for additional information). These seven variables featured in an initial run of a k-means cluster analysis, where the number of data clusters (*k*) was allowed to vary from two to 20. For each value of *k*, a pseudo F statistic was computed to quantify the ratio of intra-group homogeneity to inter-group heterogeneity. The larger the value of this ratio, the more appropriate the given value of *k* for the analysis (Calinski and Harabasz, 1974). That being said, a *k* value of 2—which corresponds to two socioeconomic status (SES) groups—had the highest pseudo F statistic and was adopted for the remainder of the analysis. Stated differently, the 385 city of Austin CBGs were classified into two groups on the basis of the seven variables listed above. The resulting grouping structure is mapped in Figure 10 below, and Table 10 summarizes descriptive statistics for the two groups. Insofar as one group, on average, appears to be more disadvantaged in the relevant socioeconomic variables relative to the other, one group is named the “low SES” group and the other the “high SES” group. The low SES group is predominantly concentrated in east Austin, which is widely known to be the disadvantaged area of the city (e.g., Herrick, 2008).

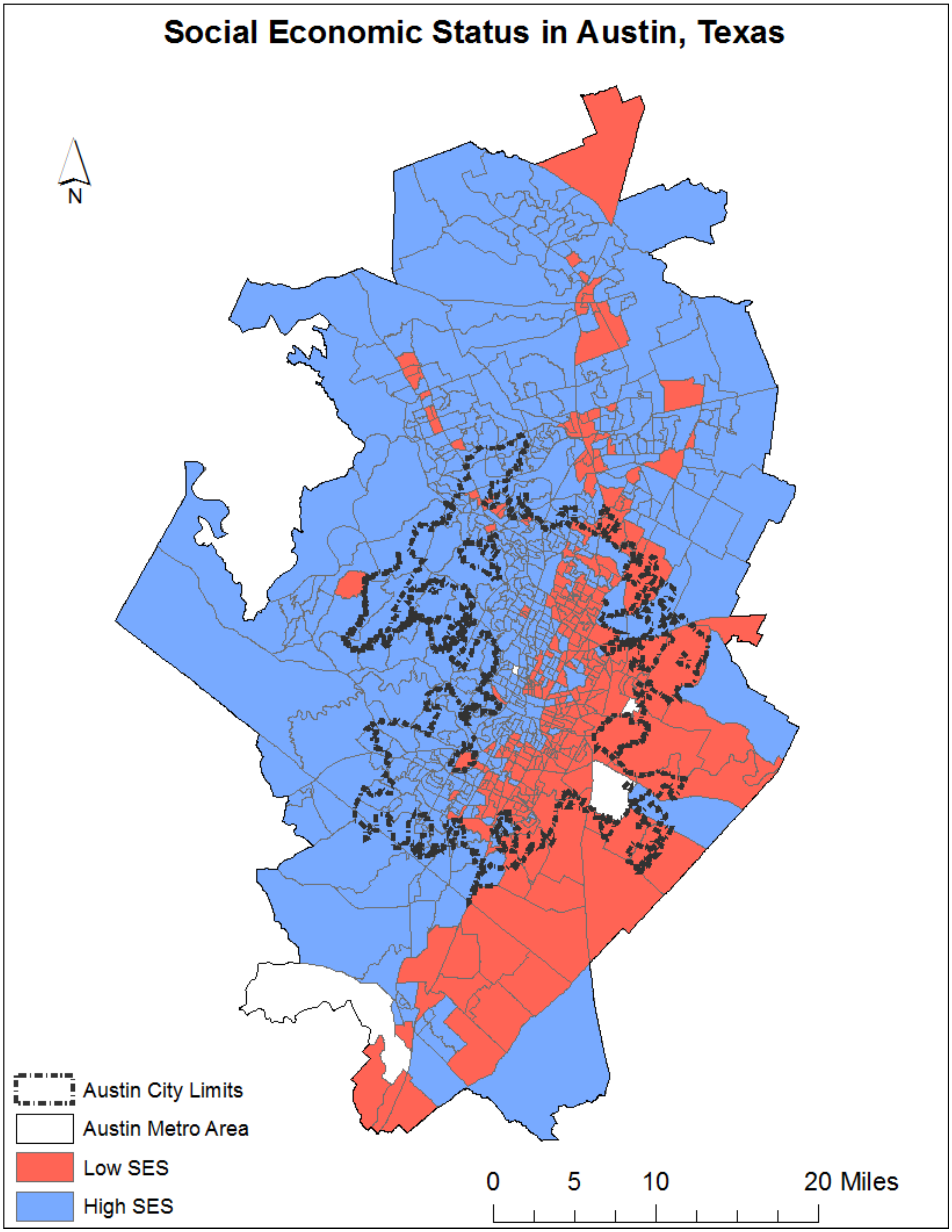

Figure 10. The city of Austin divided into Low and High Socioeconomic Status.

Table 10. Descriptive Statistics for Socio-economic Groups.				
	Low SES Group		**High SES Group**	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
% Minority population	0.738	0.177	0.335	0.166
% Adults without a high school diploma	0.285	0.156	0.055	0.060
Unemployment rate	0.098	0.062	0.054	0.041
% Households on public assistance	0.024	0.031	0.009	0.017
% Renters	0.667	0.246	0.517	0.264
% House burdened	0.587	0.165	0.411	0.115
% Poor or struggling	0.631	0.154	0.259	0.132

n=385 CBGs in the Austin city limits

T-Tests and Mann-Whitney Results for Equal Measures of Central Tendency in Traffic Congestion and Traffic Access Between SES Groups

Given the delineation of Austin CBGs into high and low SES groups (Fig.10), it is possible to compare measures of central tendency in key variables of interest for these two groups in order to evaluate the degree to which those variables are (in)equitably distributed across the city of Austin. Per research question #2, we begin by looking at results that compare the means and [effectively] the medians of the key *traffic congestion* and *transit access* variables for the two SES groups. Once again, the first variable (*congestion*) is a hindrance to mobility. It is measured here as a ratio of total average daily traffic (ADT) in a CBG to the total capacity (*Designed ADT*) in that block group. A ratio equal to 1.0 indicates that a CBG is at capacity—the daily volume of traffic experienced is the same as the volume for which the CBG-based road network was designed. Ratios greater than 1.0 indicate situations in which the road network within a CBG is over-capacity, i.e., *congested*. The second variable, *transit access*, was measured as a multiplicative function of distance to the nearest transit stop in a CBG and the frequency of transit trips within that CBG. This composite index ranges in value from 0 (no access) to 1 (highest access in the Austin study area). The SES group-specific means of these two variables were compared using a t-test with Welch's correction for unequal variances. Their medians were [effectively] compared with the nonparametric Mann-Whitney U test. Even though the parametric t-test is relatively robust to departures from distributional assumptions (e.g., Rogerson, 2015), both tests were performed here for reasons of comprehensiveness. The results from both tests, for both variables, are reported in Table 11.

Table 11. Tests for Equality in Means and Medians of Variables for Research Question #2.

Variable	Measure of Central Tendency	Low SES	High SES	Test Statistic	p-value
Traffic Congestion	Mean	1.08	1.00	t = -3.32 (df=317.6)†	0.001**
	Median	1.09	0.95	z = -3.20††	0.001**
Transit Access	Mean	0.22	0.15	t = -4.85 (df=258.0) †	<0.001***
	Median	0.18	0.12	z = -4.89††	<0.001***

***p<0.001 **p<0.01; †approximate degrees of freedom for Welch's correction (unequal variances); ††z-score is based on a nonparametric Mann-Whitney U statistic

The results from Table 11 speak to somewhat of a paradox. Namely, while mean (median) *transit access* is statistically significantly higher in low SES neighborhoods relative to high SES neighborhoods, so is *traffic congestion*. In other words, on one hand there are greater opportunities to use public transit in low SES neighborhoods. But, on the other, these opportunities do not coincide with lower traffic congestion in these areas. This seeming paradox may be due to the fact that several major roadways and interstates cut through low SES neighborhoods (Herrick, 2008) and are used by commuters. The end result is that traffic congestion is unevenly distributed across Austin, and the problems may be most severe in the most disadvantaged neighborhoods. More optimistically, though, to the extent that individuals within low SES areas tend to be the least mobile, in Austin access to public transit is greatest in these spaces. Moreover, supplementary t- and Mann-Whitney tests on the *transit usage* variable from the Census American Community Survey show that workers in low SES CBGs use public transit at significantly higher rates than workers in high SES CBGs. Hence, the increased transit access available in these areas appears to be strongly linked to increase usage.

Table 12. Supplemental Test for Equality in Means and Medians of Transit Usage.

Variable	Measure of Central Tendency	Low SES	High SES	Test Statistic	p-value
Transit Usage	Mean	0.07	0.04	t = -5.31 (df=259.2†)	<0.001***
	Median	0.06	0.02	z = -5.67††	<0.001***

***p<0.001 **p<0.01; †approximate degrees of freedom for Welch's correction (unequal variances); ††z-score is based on a nonparametric Mann-Whitney U statistic

Summary of Results Chapter

To recap, the results from this chapter uncover important patterns for planners. First, the regression model used to estimate ADT for all of Austin's roadways produced results similar to those found by Mohamad et al. (1998) and Anderson et al. (2006). Namely, ADT counts are a function of number of lanes, population, and employment. As number of lanes, population, and local employment increase, so does ADT. Also, the further away from the CBD, the lower the traffic volume. The second regression model, of Designed ADT on relevant explanatory variables, then allowed for a comparison of predicted ADT to predicted road capacity. The results suggest that many roadways in the Austin area were designed to handle lower traffic capacity than what they are presently experiencing—perhaps traffic planners did not anticipate such a large population surge.

The second part of research question #1 produced some surprising results. Essentially, the analysis found that access to public transit might actually contribute to congestion [before controlling for transit use]. That being said, when controlling for public transit usage, access is in fact negatively and significantly correlated with congestion; however, the magnitude of this relationship is relatively small. This finding suggests that when people have better access to public transit and choose to use it, they may be less likely to experience as much congestion.

Next, to address research question #2, census block group data were used to divide Austin into two groups: disadvantaged and not disadvantaged (Low SES and High SES). The results here have very important implications. First off, the lower SES neighborhoods have greater access to public transit. However, the research also shows that traffic congestion is also higher in these areas. This is troublesome because it shows that individuals who rely most on public transit are experiencing the most delays. However, one can be cautiously optimistic because it also appears that individuals in low income neighborhoods use public transit at a much higher rate than those in higher income neighborhoods. Therefore, it can be concluded that these areas have high usage because of high access. Low income neighborhoods contribute least to congestion but are most negatively affected by it. This is certainly an area that further research will need to explore.

V. DISCUSSION, CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH

Revisiting Research Question #1: Geographies of Congestion and Public Transit

This study was divided into two separate but important questions. Research question #1 asked about the geographic distribution of traffic congestion in the greater Austin area. That question was broken down into two components: a. what is the relationship between traffic congestion and access to public transit? b. what is the relationship between traffic congestion and public transit usage. The workflow for studying this question is illustrated in Figure 3. Recall that for this study, congestion will be measured as a ratio of ADT (Average Daily Traffic) to Designed ADT. If a roadway has a ratio above 1, that roadway will be considered congested, whereas if the ratio is below 1, it will be considered not congested.

Also recall that access in this study is measured by distance from a Census Block Group's (CBG) population-weighted centroid to the nearest public transit stop and aggregate frequency of public transit service (number of transit trips) per square mile with a CBG. Remember that the distance variable has an inverse relationship to access, meaning as distance increases, transit stops become less available. It is important to remember that for this study, access will be a function of both distance and functionality (Bullard 2008). The previously mentioned variables accomplish this task. This is important for a few reasons. First, suppose an area has many transit stops but is not serviced frequently. In this case, travel times will be longer because even though there are many stops, there are not enough busses/transit lines/etc. to timely transport people.

Second, it is possible that there is a transit stop that is serviced frequently, but it may be the only stop in a certain distance. In this case, the frequency of service is essentially negated by the fact that the stop is located far away from a significant portion of the population. Therefore, it is important that access factor both proximity and frequency because both are equally important.

The results have important implications when compared to previous literature and the overall context. The most congested roadways in Austin are the major thoroughfares. Interstate highways, primary arterials, and major collectors all have statistically significant relationships with congestion according to the regression model. Specifically, State Highway 130, portions of Interstate 35 in northern Austin, and portions of US Highway 183 in western and northwestern Austin show up as the most congested roadways in the city. Overall, the map in Figure 8 reveals that the majority of roadways in this study are at least at capacity, and a significant portion of them are above or significantly above capacity. Also note that roadways near interstates or urban freeways generally have more congestion than roads further away. This is a trend that was picked up in previous literature (Eom et al, 2006). This is significant because it limits options for travelers who encounter congestion and clogs up major side roads that are often frequented by busses.

Furthermore, these results can be tied back into sprawl. Recall the three statements from Chapter I: 1. the roadways of sprawling metropolises are frequently characterized by substantial traffic congestion (e.g., Ewing, 2008); (2) accessibility to and usage of public transit has the capacity to alleviate some of the costs of sprawl, including traffic congestion (e.g., Bernick and Cervero, 1997); and (3) access to and usage of public transit tend to be unevenly distributed in many cities (e.g., Bullard et al., 2000; Agyeman, 2005). Indeed, Austin certainly fits the first and third statements. The map in Figure 8 clearly supports the first statement while tables 11 and 12 support the third statement. The second statement from the previous paragraph leaves many things to be desired from Austin. First, there is only one metro/light rail line in the entire city. While this light rail covers a respectable distance from Leander (northwest Austin) to downtown, it leaves many areas of the city unable to use light rail for any practicality. However, results from table 9 suggest that transit usage and transit access are important, even in cities like Austin which have underdeveloped networks. The most important finding here is that when access is accounted for, the greater the population that uses public transit, the less problematic the issues of congestion are. These results make an argument for more public transit (especially in sprawling cities like Austin) and support the second statement made by Bernick and Cervero (1997).

Revisiting Research Question #2: Congestion, Transit, and Socioeconomic Status

The second part of the study addressed the problems of traffic and access to public transit in relation to distribution between socio-economic groups in Austin. In this study, socioeconomic status was defined on the basis of seven variables: (1) minority population; (2) adults without a high school diploma or equivalency degree; (3) unemployment rate; (4) households on public assistance; (5) renter-occupied housing units; (6) cost-burdened households; and (7) size of the poor and struggling population. These variables were chosen based on previous studies (Sampson et al., 2002; Manturuk et al., 2009; Weaver et al., 2016). The study used a k-means clustering and found that the optimal number of socioeconomic groups in the dataset, based on the seven variables, was two. For the purpose of this study, these two groups were called “disadvantaged” and “not-disadvantaged” based on their in-group values of the selected variables.

There are several important implications from this study, but one may stand out as most important. This study found that transit access is statistically significantly higher in low SES neighborhoods (disadvantaged populations) than in higher SES neighborhoods (less disadvantaged populations). This finding certainly seems positive, given that lower SES populations tend to rely more on public transit and, in Austin, such populations appear to have better access to this mode of mobility. However, the study also found that traffic congestion in these areas is higher than in areas with relatively advantaged populations. Essentially, lower SES neighborhoods have greater access to transit, but perhaps because public transit in Austin consists of almost entirely buses, it seems that the worst traffic problems are also in the most disadvantaged neighborhoods. Additionally, this outcome might be partially caused by the fact that major roadways and interstates in Austin go through low SES neighborhoods (Herrick, 2008). This finding should not be overlooked, especially considering that those who commute into the city may be contributing most to sprawl.

The above results have several meaningful consequences. The literature stresses that lower SES groups may be more dependent on public transit for every day uses and errands. Many members of these groups do not own automobiles. Therefore, public transit is one of the only options for daily travel. As such, if the highest traffic congestion is in these neighborhoods, it makes it even more burdensome to travel by bus. Along those lines, recall two of the four observations made by Handy (2005): (1) investing in light rail systems will increase densities (2) adopting New Urbanism design strategies will reduce automobile use. Given the findings from this study, it is likely that a city such as Austin might benefit from such policies. If the city of Austin invests in new light rail, it might experience some relief in traffic (some of which is two times that of which it was designed to carry). That being said, elements of New Urbanism can already be seen in Austin. However, these places are usually not located near low SES neighborhoods and are usually somewhat unaffordable for the lowest income residents. The New Urbanist Mueller Development, a neighborhood created from an old airport in east Austin, offers some affordable housing options, but it is a considerable distance (timewise) from the central city. The above statement shows similar findings to that of Bullard (2008). "Whether highway or airport sprawl is "good" or "bad" will almost always depend on where you live, and whether or not you own a car" (Bullard, 2008: 241). In the case of Atlanta, nearly 4 out of 10 Black households do not have access to a car. Furthermore, only about a third of the jobs in that area were within a one-hour public transit ride of low income neighborhoods (Bullard, 219). An area of future study for Austin would be to see where exactly in the city are jobs located, and perhaps more importantly, is there access to those jobs. Also, a study similar to Bullard (2008) could be done to see the percentage of lower income residents that have access to automobiles.

In sum, the analysis of research question #2 calls attention to important social justice issues in the transportation geography of Austin. Lower SES neighborhoods have might greater access to public transit—and use it more frequently than higher SES neighborhoods; but they are also most [negatively] affected by congestion. In the words of one transportation planner: “The traffic problem in Austin can’t be solved right now, we can only make it less bad” (Rain Nox, panelist on “managing urbanization” at Texas State University, 30 November, 2015). Statements like this are alarming for a city that continues to grow at a fast rate. This growth is pushing lower income families further out, where housing is more affordable. City leaders and planners struggle with this issue because younger generations can afford to live in downtown in new developments. This is a problematic phenomenon because of lower SES individuals relying more on public transit in areas where the networks may be sparse.

Limitations and Future Research

There are several limitations to this study. First, the average daily traffic (ADT) data used to operationalize congestion measures aggregate traffic throughout an entire day. It does not factor in things such as peak hours of congestion, time of day, or day of the week. It is simply a count of all of the cars that passed through a certain roadway segment in a given day. Therefore, it is entirely possible that there are times of the day where some of the most congested roads *on a daily basis* are not congested at all. Conversely, there may be times of day where roads that are not considered congested by our operational definition are quite congested.

Additionally, Eom et al (2006) suggest that traffic on local roadways may not be correlated to traffic on freeways, even if the two types of roadways are close together. This observation is important to this study because a significant amount of congestion in Austin occurs on secondary and tertiary roads near freeways—yet the TxDOT dataset does not provide ADT data for most of these roadways. As such, conventional spatial extrapolation techniques (e.g., Lowry and Dixon, 2012) were used to predict ADT across the greater Austin road network from data that were mostly measured for freeways.

Next, there is always uncertainty in statistical modelling. Zhao and Chung (2014) add that regression models in their current forms may not be adequate enough to meet the standards of engineers or designers, but that they are improving. Their study found that functional class of roadway outperformed other variables, but that they did not explain the causes that determine ADT because they were also determined by other factors including traffic volume. Functional classes were not consistently related to ADT, but models that excluded ADT completely had the worst performance (Zhao and Chung 2014). Anderson et al (2006) also add that even though their regression models account for 82% of variability in ADT, future work could focus on reducing uncertainty by including additional variables in the analysis.

Furthermore, because Austin public transit consists primarily of bus routes, it is highly likely that these buses run on the most trafficked—and thus most congested—roads in the city. Looking again at the map in Figure 8, a large part of the city is either at or above capacity, indicating congestion. Buses use many of these roads, so it is possible that they are contributing to traffic.[3] Downtown Austin has bus only lanes, but these lanes are mostly restricted to the central city and do not extend very far past the central business district (CBD). Therefore, some citizens may be skeptical about using public transit because it might take considerably longer to get to their destinations. Recall from Chapter II that even in Helsinki, a city with an extensive public transit network, mean

[3] Crucially, this fact may be a driving force behind the positive and significant bivariate correlation detected between *traffic congestion* and *transit access* in Table 11.

travel times were found to be longer than private automobile (Salonen and Toivonen, 2013). Future studies could look at travel times within the city of Austin, comparing private vehicle trips to public transit trips, similar to Salonen and Toivonen (2013).

In addition, data from Capital Metro, the Austin public transit authority, is fairly difficult to obtain. Some of it was incorporated into the U.S. Environmental Protection Agency (EPA) Smart Location Database (SLD). However, outside of that dataset, we were unable to obtain reliable data for alternative time periods or locations beyond Austin's city limits. Anderson et al (2006) experienced similar problems in their study. They note that a major limitation in these models is the availability of data that can fit the rigorous calibration required to use them effectively. Salonen and Toivonen (2013) were able to acquire comprehensive data for their study. However, this is not the norm. Wang and Kockelman (2009) also mention variables like housing price, trip generation rates, pavement conditions, and crash rates could further enhance studies. However, they note that finding such data is difficult.

Also, the neighborhood typologies in this study are very basic. This study looks at only seven dimensions of socioeconomic status (SES), and even then does not consider every possible variable. It is broken down into very simplistic categories (disadvantaged and not disadvantaged groups) that were suggested by data, not by bottom-up community-based initiatives. Future work can look at different social and economic variables and determine their relationship to congestion and public transit. It will also be important to collect primary data through surveys and other community-engaged methods, given that secondary data do not always reflect local conditions and local knowledge.

Also, this study was limited by one of the most important geographic themes: scale. This analysis was performed on the Census Block Group (CBG) scale, therefore the results and conclusions can only be made at this scale. Further research could be performed to see how scale impacts this study and studies similar to it. Here, the Modifiable Areal Unit Problem (MAUP) affects this study.

Finally, one aspect that has not been mentioned thus far is environmental impact. The focus of this study was on the spatial distribution of congestion and neighborhood access to public transit. The SLD database, and a related H+T index database,[4] contain several environmental variables that could measure the relationship between congestion and environmental impact. Of particular interest are the potential effects that various features of traffic congestion—e.g., idling, emissions, etc.—have on variables like air quality. In turn, harmful effects on air quality can pose serious health risks to surrounding populations. Given that low SES neighborhoods in Austin seem to be disproportionately affected by traffic congestion (see Table 11 from Chapter IV), such effects are likely to give rise to several environmental injustices. Hence, future research on environmental impacts, especially from the environmental and social justice perspectives, is an important next step for this area of inquiry. The results of this study could be compared to measures of pollution in Austin. Because the study found relatively strong connections in regards to socio-economic status and congestion, a comparison to environmental impact could make an even stronger case for policy implications.

[4] http://htaindex.cnt.org/map/

LITERATURE CITED

Anderson, M., Sharfi, K., and Gholston, S. (2006). "Direct Demand Forecasting Model for Small urban Communities Using Multiple Linear Regression." *Transportation Research Record: Journal of the Transportation Research Board*, 1981(-1), 114-117.

Beauregard, Robert A. (2006). *When America Became Suburban.* Minneapolis, MN: University of Minnesota Press.

Beauregard, Robert A. (2011). "Time, Action, Space". *Journal of Urban Geography.* 32(4), 470-475.

Bernick, Michael, and Robert Cervero. (1998). "*Transit Villages in the 21st Century.*" East Lansing: H-Urban publishing.

Bernier, Nathan. (2014). *This is Project Connect's $1.4 Billion Proposal for Austin \Urban Rail.* Austin: News 90.5 KUT. http://kut.org/post/project-connects-14-billion-proposal-austin-urban-rail (last accessed 7 April 2016).

Briggs, Xavier de Souza. (2009). *Democracy as Problem Solving.* Cambridge, MA: The MIT Press.

Bullard, Robert D. (2007). *Growing Smarter: Achieving Livable Communities, Environmental Justice, and Regional Equity.* Cambridge, MA: The MIT Press.

Bullard, Robert D, et al. (2000). *Sprawl City: Race, Politics, and Planning in Atlanta.* Washington D.C.: Island Press.

Eom, J., Park, M., Heo, T., and Huntsinger, L. (2006). "Improving the Prediction of Annual Average Daily Traffic for Nonfreeway Facilities by Applying a Spatial Statistical Method." *Transportation Research Record: Journal of the Transportation Research Board*, 1968(-1), 20-29.

Ewing, Reid H. (2008). "Characteristics, Causes, and Effects of Sprawl: A Literature Review." *Urban Ecology. An International Perspective on the Interaction between Humans and Nature.* New York: Springer.

Handy, Susan. (2005). "Smart Growth and the transportation-land use connection: what does the research tell us?" *International Science Review,* 28(2), 146-167.

Herrick, Clare. (2008). To the west and east of Interstate-35: Obesity, philanthropic entrepreneurialism, and the delineation of risk in Austin, Texas. *Environment and Planning A* 40(11): 2715-2733.

Hess, D.B. (2005). “Access to employment for adults in poverty in the Bufallo-Niagara region.” *Urban Studies,* 42(7), 1177–1200.

Kawabata, M., Shen, Q. (2006). “Job accessibility as an indicator of auto-oriented urban structure: a comparison of Boston and Los Angeles with Tokyo.” *Environment and Planning*. B, 33(1), 115–130.

Kawabata, M., Shen, Q. (2007). “Commuting inequality between cars and public transit: The case of the San Francisco Bay Area, 1990–2000.” *Urban Studies,* 44(9), 1759–1780.

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

Leyden, Kevin M. and Abraham Goldberg. (2015). “Untangling what makes cities liveable: happiness in five cities.” *Proceedings of the Instiution of Civil Engineers – Urban Design and Planning.* 165(3), 127-136.

Lowry, Michael, and Michael Dixon. (2012) "GIS Tools to Estimate Average Annual Daily Traffic." *National Institute for Advanced Transportation Technology*.

Manturuk, et al. (2009). “Homeownership and Local Voting in Disadvantaged Urban Neighborhoods.” *Cityscape,* 11(3), 213-230.

Mitchell, Andy. (1999). *The Esri Guide to GIS Analysis Volume 1: Geographic Patterns & Relationships.* Redlands, CA. ESRI Press.

Mohamad, D., Sinha, K., Kuczek, T., and Scholer, C. (1998). “Annual Average Daily Traffic Prediction Model for County Roads.” *Transportation Research Record: Journal* of the *Transportation Research Board*, 1617(-1), 69-77.

O’Sullivan, D., Morrison, A., Shearer, J. (2000). “Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach.” *International Journal of Geographical Information Science,* 14(-1), 85–104.

Ramsey, K and Bell, A. (2014). *Smart Location Database Version 2.0 User Guide.* Washington, D.C.: United States Environmental Protection Agency.

Register, Richard. (2006). *EcoCities: Rebuilding Cities in Balance with Nature.* Gabriola Island, British Columbia: New Society Publishers.

Rodrigue, Jean-Paul. (2006). “Transportation and the geographical and functional integration of global production networks.” *Growth and Change: a Journal of Urban and Regional Policy,* 37(4), 510-525.

Sampson et al. (2002). “Assessing ‘Neighborhoods Effects’: Social Processes and New Directions in Research.” *Annual Review of Sociology*, 28, 443-478.

Sanchez, T. and Marc Brenman. (2007). *The Right to Transportation: Moving to Equity.* Chicago: APA Planners Press.

Salonen, M. and Toivonen, T. (2013). “Modelling travel time in urban networks: Comparable measures for private car and public transport.” *Journal of Transportation Geography*, 31, 143-153.

Silverman, Bernard W. (1986). *Density Estimation for Statistics and Data Analysis.* Boca Raton, FL: CRC Press.

“Smart Location Mapping.” (2016). EPA. https://www.epa.gov/smartgrowth/smart-location-mapping (last accessed 31 May, 2016).

Tang, Y. F., Lam, W. H. K., and Ng, P. L. P. (2003). “Comparison of Four Modeling Techniques for Short-Term AADT Forecasting in Hong Kong.” *Journal of Transportation Engineering*, 129(3), 271-277.

Teaford, Jon C. (2007). *The American Suburb: The Basics.* New York: Routledge Press.

Torrens, Paul M. (2008). “A Toolkit for Measuring Sprawl.” *Applied Spatial Analysis and Policy,* 1(1), 5-36.

Wang, X., and Kockelman, K. (2009). “Forecasting Network Data.” *Transportation Research Record: Journal of the Transportation Research Board*, 2105(-1), 100-108.

Weaver, Russell, et al. (2016). *Shrinking Cities: Understanding Urban Decline in the United States.* New York: Routledge Press.

Weissman, Jordan. (2015). “Austin, Texas, is Blowing away every other big city in population growth.” *Slate Magazine.* New York.

Zhong, M., and Hanson, B. (2009). “GIS-based travel demand modeling for estimating traffic on low-class roads.” *Transportation Planning and Technology*, 32(-5), 423-439.