The Thesis Committee for Han Park
Certifies that this is the approved version of the following thesis:

Ridership Analysis at the Stop Level: Case Study of Austin, TX

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Ming Zhang

Terry Kahn

Ridership Analysis at the Stop Level: Case Study of Austin, TX

by

Han Park, B.S.; M.S.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Community and Regional Planning

The University of Texas at Austin

December 2011

Abstract

Ridership Analysis at the Stop Level: Case Study of Austin, TX

Han Park, MSCRP

The University of Texas at Austin, 2011

Supervisor: Ming Zhang

Transit ridership analysis has been advancing towards the use of disaggregate spatial and boarding data. This study attempts to improve the understanding of factors influencing transit ridership by estimating/comparing ridership models at the route, the segmented route, and the stop level in the Austin area.

Spatial and statistic analysis methods are used in this study. The dependent variable is ridership at the transit route, the segmented route, and the stop level, whereas independent variables consist of traveler characteristics, land use, transit service characteristics, and other contextual factors. Spatial analysis is conducted using Geographic Information System (GIS) to compile data within a quarter-mile buffer from each transit stop, each segregated route, and each route. Linear and semi-log models of ridership are estimated using Statistical Analysis System (SAS). Initial analysis confirms the qualitative understanding that traveler demographics such as population and employment densities, ethnic background, and income significantly affect transit ridership. Land use composition, measured by the shares of single-family homes, multi-

family homes, commercial, civic uses, as well as the total area of paved parking, all influence transit use. Service qualities such as headway and transfer opportunities also matter. Sensitivity tests of these factors affecting ridership are carried out to compare model performance among the route, segmented route, and the stop level analyses.

It is expected that the study findings will help to better inform transit agencies and local communities in optimizing existing transit operations, planning for new services, and developing transit-friendly environments.

Primary data were obtained from the Capital Metropolitan Transit Authority and the Census Bureau, and secondary data was processed by GIS analysis.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Public transportation is an essential travel means for city residents, especially for low- and mid-income households. There is always a need for mobility, both within cities and from cities to outlying areas. As cities grow larger, more transit services are needed to serve commuting or other travel purposes.

Most previous research analyzed ridership using entire routes as a basic unit of analysis. Some advanced studies used route segments as a basic observation unit to determine spatial variation. Although more advanced studies have increased the fitness of analysis, there remains a limitation that the length of each segment is different from zone to zone and that the area of each fare zone is different. Also, not all transit routes serve all of the fare zones. In addition, it is more difficult to represent transit service variables at the segmented level and to analyze the inter-route relationships with segmented routes. Due to the limited advantages of recent segmented studies, a stop-level analysis seems more appropriate. Stop-level ridership analysis is able to show clear reasons why certain bus stops have greater ridership and explain the extent to which each individual independent variable affects the ridership demand.

The main body of this research includes two parts. First, the literature review clarifies the characteristics of bus ridership analyses. Previous studies are organized according to the analysis level of the dependent variable: aggregate-level analysis, route/segregated route-level analysis, and stop-level analysis. Second, empirical analysis shows the influence of each variable related to ridership. Within the empirical analysis, route-level, segmented route-level, and stop-level regression models will be developed from independent variables including traveler characteristics, land use, transit service features, and other contextual factors. Each analysis is conducted with two types of regression model (linear regression and semi-log regression models) to better understand the influence of each independent variable.

Chapter 2: Literature Review

1. Aggregate-Level Analysis

1) General Review

Most of the variation in transit ridership between urbanized areas can be explained by the size (both population and area) of the metropolitan area, the vitality of the regional economy (measured in terms of median housing costs), and the share of the population with low levels of private vehicles access (measured in terms of the percentage of zero-vehicle households). Transit patronage is also explained, to a lesser but still significant extent, by transit service levels and fares. The relative influence of transit service levels on ridership is greater than that of transit fares. Finally, by separating the service supply variable into an instrumental control variable and a residential policy variable, it can be estimated that large changes in transit service levels are likely to have far less influence on transit ridership than many of the previous aggregate models of transit patronage would suggest (Taylor and Miller, 2003).

Estimating transit demand functions is complex, because the perceived utility and disutility of transit trips vary significantly between individuals and trips (even for the same person). First, the utility of a transit trip is, largely, a function of the utility of the activity from which the demand for a transit trip is derived. While the utility, and hence demand, for a particular good, service, or activity can be ascertained, transit is likely just one of many possible ways to access the desired good, service, or activity. Second, the perceived disutility of transit trip costs varies dramatically. Numerous studies have found that travelers perceive out-of-vehicle time (walking to and from transit stops, transferring, and waiting at transit stops) as more onerous (and therefore more costly) than in-vehicle time (Horowitz et al., 1986; Small, 1992; Small et al., 1999; Wardman, 2001). Therefore, someone who lives and works near transit stops on a particular line will likely perceive lower costs for a peak-hour and peak-direction transit trip than will a person traveling between the same two stops, but who lives and works farther from the stops; and/or who is traveling at night or weekends, when service is less frequent. Third, while some people do not have practical substitutes for transit trips, most do. Relatively fast, flexible private vehicles dominate metropolitan travel and even walking trips now far exceed the number of trips made on public transit in the US. Thus, most travelers find the relative utility of traveling by other modes (particularly private vehicles) to be greater than that of public transit for most trips. (Taylor et al., 2008)

2) Time-Series Analysis

Time-related variables with local characteristics help to better understand the behaviors of transit riders. Since a single cross-section of travel behavior only represents ridership at a certain time, it would be impossible to test for the significance of adjustment lags in response to changes in some of the variables such as fares, waiting time, etc. and very difficult to simultaneously evaluate the impact of many weather-related comfort variables such as temperature, rainfall, snowfall, and cumulative snowfall. Some variables, such as the price of non-transportation goods, are easier to incorporate within a time-series study (Gaudry, 1974).

3) Transit-Oriented Development

Arrington and Cervero (2009) reported that the most effective strategy to increase ridership in Transit Oriented Development (TOD) areas is to maximize development densities near transit stops. In terms of TOD cases, the employment densities at trip destinations are more influential than the population densities at trip origins. However, relative travel time, which is estimated by transit time in relation to auto time, is still more influential than any land use factors such as density, diversity of uses, or urban design.

They also reported that mixed land uses in TODs make it possible for transit services to be used for various trip purposes. Although this is not a main purpose for prospective TOD residents, whose primary reason for living in TOD areas is better access to their work from their home, mixed uses and urban design treatment are important for amenity and design value in attracting residents and visitors. Urban design and local land use mix can attract prospective TOD residents and transit riders. Also, physically well-designed TOD areas allow more desirable trips to and from the neighborhoods.

4) External Factors

TranSystems et al. (2009) explained the relationship between ridership and external factors in a straightforward way. For instance, population and employment increases within a region can raise transit demand by expanding the potential ridership base. Alternatively, other factors such as fuel prices, parking availability and prices, and regional development patterns affect transit ridership by influencing the relative attractiveness of transit versus automobile use. While transportation agencies may not be able to control external factors, they can anticipate their potential impacts on transit demand and therefore initiate appropriate responses.

5) Variables

At the metropolitan level, there exists a consensus that urban population size and density, households without cars, local economy vitality, and transit supply all affect transit usage (Kohn, 2000; Chatterjee et al., 2002; Yoh et al., 2003; Taylor et al., 2003). Parker et al. (2002) asserted that TOD can increase ridership by 20% to 40% at an individual station, and up to 5% overall at the regional level. A 1994 survey in Portland, Oregon showed that areas with TOD characteristics and good transit services had a much higher transit modal share (Lawton, 1997). Land use types and land use mix influence transit use, although less so than density (Parsons Brinckerhoff, 1996a).

2. ROUTE-LEVEL/SEGMENTED ROUTE-LEVEL ANALYSIS

1) General Review

Most transportation researchers agree that there are many factors that affect both the demand for and supply of urban transit. These include family size (families with children may choose to use personal vehicles rather than urban transit because the monthly cost of transit passes may be, or is perceived to be, more expensive), economic change (e.g., employment opportunities, taxes, fuel costs, parking fees, automobile insurance costs, vehicle operating costs, subsidies, investment, etc.), demographic impacts (e.g., population growth, immigration rates, fertility rates, parking rates, and distance to work), and other factors such as convenience of access to services and expensive downtown parking (Kohn, 2000).

2) Walking Distance and Diminished Transit Use

Walking distance to the station/stop has a negative influence on transit use. According to 200 Southeast Florida TOB survey data, most trips origins were within 1,800 ft (0.35 mile) of transit stops and few trips originated more than 2,700 ft (half a mile) from transit stops. This finding was consistent with other studies reporting that transit use begins to decrease after 300 ft (0.06 mile) and vanishes beyond 1,900 ft (0.36 mile) from a transit stop (Lam and Morrall, 1982; Levinson and Brown-West, 1984; Zhao et al., 2003).

3) Transit Supply Aspect

Peng et al. (1997) described a model demonstrating that a service improvement has a twofold impact on ridership: it may increase ridership on the route with service changes, but it may also reduce the ridership on competing routes, so that the net change in ridership may be small.

The coefficient of population was negative and statistically significant in the supply model, while it was positive and statistically significant in the demand model. This could be explained by the fact that Tri-Met, the public transportation provider for Portland, Oregon, is under pressure to provide services to a broad area, extending to low density suburbs. Even within low density suburb, the policy requires buses to complete the entire length of a route. Therefore, a negative relationship was identified between population and service supply in the Portland area. Service increase and population growth in core areas will have a more favorable impact on transit use than corresponding increases in suburban areas (Peng et al., 1997). In this sense, a variable may have a different role in two different transit models: transit supply and demand.

4) Geographically Weighted Model

Zhao et al. (2005) presented the development of geographically weighted regression models to predict the use of public transit for home–workplace commuting purposes. A large array of potential transit use predictors were considered, including demographic, socioeconomic, land use, transit supply quality, and pedestrian environment variables. The results showed that the percentage of Black ethnic population in a Transportation Analysis Zone (TAZ) had a positive influence on transit ridership in general, and the effect was stronger in the downtown area. In the case of Black population and employment density, although these variables were expected to be positively correlated with transit use, they were found to be negatively correlated in some areas (Zhao et al. 2005).

5) Variables

Peng and Deuker (1995) categorized their data as follows: ridership, transit service variables, and demographic and socioeconomic data. Ridership data usually refer to boarding trips at a spatial unit such as a bus stop, a route segment, or a whole route. Transit service data include service quantity variables such as bus frequency, hours of service, and route length of service and service quality variables such as on-time performance. Transit service variables also include the location, usage, and capacity of Park-and-Ride sites provided by transit agencies. Demographic and socioeconomic characteristics at the place of residence are used to estimate potential transit users at trip origins, and include population and age structure, household income, and levels of auto ownership. The demographic and socioeconomic variables at the destination are used to represent characteristics at places where people wish to access for specific purposes, including employment and high school enrollment.

A Transit Cooperative Research Program (TCRP) study suggested that ridership was most significantly impacted by Central Business District (CBD) employment size and density, residential density, and availability of bus feeder services (Parsons Brinckerhoff, 1996b). The study supported the findings of Pushkarev and Zupan (1982), that CBD size and distance to CBD are influential factors.

3. STOP-LEVEL ANALYSIS

1) General Review

Stop-level analysis estimation can reduce aggregation errors in some variables. However, at the same time, the allocation of demographic variables like population and income to individual stops may encounter difficulties. For example, because most transit stops in urban areas are less than a quarter mile apart, the stop catchments largely overlap.

2) Variables

Kikuchi and Miljkovic (2001) developed transit ridership models at the bus stop level with household auto ownership, number of households, average household income, bus stop condition, bus stop accessibility, commercial activities, and transit service quality. Another bus stop level ridership model, T-BEST, was developed by the Florida Department of Transportation. T-BEST predicts ridership by route, direction, and time-of-day based on the number of bus runs, bus stop catchment characteristics (population, percentage of households without cars, percentage of African-American population, average household income, percentage of population aged 65 or older, and employment data), accessibility characteristics, and the effects of alternative routes and network design configurations.

Cervero et al. (2010) researched direct ridership on bus rapid transit in Los Angeles County. They suggested three key sets of variables. First, service attributes such as frequency of buses (headways and number of buses per hour), operating speeds, feeder bus connections (number of lines or buses), dedicated lane (dummy), and vehicle and brand marketing; second, location and neighborhood attributes such as population and employment densities, mixed land use measures (0–1 scale), median household incomes, vehicle ownership levels, distance to nearest stop, accessibility levels (e.g., number of jobs which can be reached within 30 minutes over transit network in peak periods), terminal station (dummy), street density (directional street miles divided by land area), and connectivity indices (e.g., links or nodes of street network); and third, bus stop or site attributes such as bus shelters (dummy), next-bus passenger information (dummy), passenger seating areas (dummy), far-side bus stops (dummy), Park-and-Ride lots (dummy, or number of spaces), and bus bulbs/bus borders.

3) Findings

The analysis by Cervero et al. (2010) of 69 Bus Rapid Transit (BRT) stops in Los Angeles County revealed three important factors related to high BRT ridership. First, BRT ridership increased as the frequency of BRT and feeder bus services both increased. Second, high levels of intermodal connections supported high BRT usage. In particular, rail transit connections and Park-and-Ride provisions were highly associated with significant increases in daily ridership. Third, population density and employment density were significant contributors to the increase in ridership.

Huang and Zhang (2011) explained ridership using the number of bus routes at stops and the number of stops in stop areas. The number of bus routes had a positive influence while the number of bus stops in a bus stop area had a negative influence on boarding. In this study, accessibility was calculated using rasterized data allocation from large geographies. In addition, trip demand in the overlapping service area between adjacent stops was estimated with a step-wise spatial spreading approach.

Lin and Shin (2008) explained two effects of land use diversity on transit ridership: diverse land uses attract more trips by increasing the vitality of the area and improving economic development and residents' quality of life; but land use diversity simultaneously reduces ridership, as shorter distances between locations providing different activities reduce the need to leave the metro station area. Therefore, mixed land use theoretically has both positive and negative effects on ridership. They also mentioned the need to consider the catchment radius of a station in a metro service area and the effects of relative location and transit service conditions of stations.

Chapter 3: Empirical Analysis

1. Hypothesis

1) Traveler Variables

Traveler variables represent each traveler's influence on bus ridership at origins and destinations: Population, household, household size, race, and income level function at origins while employment is the dominant influence at destinations.

Since population is representative of how many people are living at journey origins, higher population densities increase the number of prospective public transit riders. The number of households is also similar to population: If more households are located in a certain area, more household members might use public transit, including bus. In the case of employment, if more jobs are available in a certain area, it is more plausible that a number of employees tend to commute to their work using public transit methods, including bus. Employment functions also support a range of related functions within the surrounding areas. It is also expected that larger household size will result in more bus riders in the end. In terms of ethnicity, empirical results showed that Caucasian groups are less likely to use buses than Black and Hispanic residents. In the case of income level, if people are poorer, they are more likely to use public transit, including bus.

2) Land Use Variables

Land use composition around bus stops and bus routes is a critical influence on ridership levels. Amongst land use categories, some usages allowed more people to actively behave in a city while others merely provide people with active behaviors.

People living in a single-family house are less likely to use buses than those living in a multi-family dwelling. More commercial/office and civic areas along transit corridors are expected to generate more transit trips, including bus usage. On the other hand, industrial uses do not increase public transit ridership, since work-related journeys tend to involve commercial vehicles or individual vehicles for cargo transport rather than public transportation. In the case of parking, if total parking space or the number of Park-and-Ride facility increases around bus stops and bus routes, it is expected that the number of bus riders would increase.

3) Transit Service Variables

The service level of transit systems is also influential on the demand for transit. Previous studies demonstrated that transit supply and demand affect each other. Bus service quality variables include the number of bus services per day, hours of operation, revenue hours, and headway.

Initially, the number of bus services per day might have a positive influence on ridership. With a larger number of daily services, people would be more likely to use transit rather than driving; the same is predicted for the effect of longer hours of transit operation. The case of revenue hours includes two variables, the number of bus services and hours of operation. Therefore, this revenue-hour variable has a positive influence on ridership, but it cannot be used with two endogenous variables simultaneously in the same model. Conversely, longer headway between services discourages ridership.

There are also special bus routes such as University of Texas (UT) shuttle bus lines, Express buses, and Route Number 1 (Route #1) buses, including Route Number 1L (Route #1L) and Route Number 1M (Route #1M). These are major bus lines running throughout the Austin area and the dependence on these bus routes is significant. Many UT students use shuttle buses every day and Route #1 serves passenger transport running through the center-line of the city. Express buses also serve a rapid transit role for individuals who are dependent on transit or who would otherwise face a long-distance car commute.

4) Contextual Variables

In addition to demographics, land use characteristics, and transit service qualities, levels of ridership are also influenced by contextual variables. If the number of transit stops within a quarter mile buffer increases, this has an influence in two different ways. The possibility of ridership is predicted to increase due to the good accessibility to transfer to other buses, but sharing ridership with other stops within a quarter mile buffer also has the potential to reduce ridership because a greater number of local stops may reduce the number of passengers boarding per stop through distribution effects. Also, there are two unique neighborhood areas in Austin: University of Texas and Downtown. These two neighborhoods are located in the center of Austin and most Austin bus lines cross these two areas. In particular, there are many students living within the UT Neighborhood and many shuttle bus services are operated for students commuting to the neighborhood. Many bus routes also run through the Downtown Neighborhood to serve work and social activities.

Categorization	Description	Name
Traveler	Characteristics of Bus Users	• Population • Number of Household • Employment • Household Size • Ethnicity (White/Black/Asian/Other) • Income Level
Land Use	Land Use related Characteristics	• Single-family Use Acreage • Multi-family Use Acreage • Commercial Use Acreage • Office Use Acreage • Civic Use Acreage • Industrial Use Acreage • Open Space Use Acreage • Other Use Acreage • Parking Service Acreage • Park-and-Ride
Transit Service	Characteristics of Bus Service and Operation	• Total Number of Services • Operation Hours per Service • Revenue Hours • Headway • Bus Route Categories (UT Shuttle Bus, Express Bus, Route #1)
Context	Characteristics of Surrounding Area of Bus Routes/Stops	• Total Number of Bus Stops • Local Center (University of Texas) • Downtown

Table 1: Categorization and Description of Independent Variables in Regression Models

2. ROUTE-LEVEL REGRESSION ANALYSIS

1) Route-Level Variable Description

The data set used in this analysis was mainly obtained from the Capital Metropolitan Transportation Authority (Cap Metro) and 2010/2000 Census Data. Variable names and explanations for the route-level analysis are shown in Table 2. The dependent variable, weekday boarding of each route, was obtained from the Cap Metro Service Plan 2020 Final Report. The logarithmic value was then taken for the semi-log model analysis.

Data on passenger ethnicity was available from 2010 Census Data. The White_Prct variable is the percentage of residents from Caucasian ethnic groups within a quarter mile buffer. This percentage was obtained from the prorated ethnic distribution from each census tract to each quarter mile buffer. In the case of the income variable, 2000 census data was used. INC_LS_25000_PC is the percentage of households with an annual income less than 25,000 dollars as a proportion of all households (2000 census data). The prorated number of income distribution was obtained through spatial analysis using a GIS shape file derived from 2000 Census Tract data within a quarter mile buffer of every bus route. The MF_acreage variable was obtained from GIS analysis of land uses within a quarter mile buffer of every bus route. These two files were overlaid and then all land use acreages within the quarter mile buffer were calculated. The Park_Ride_D variable was obtained from CapMetro. If there is at least one Park-and-Ride site within a quarter mile buffer area, then the route observation has the value of 1; otherwise, it is zero. The Revenue_Hours variable is the total hours for which each route is operated per day. The Headway variable is the time that separates two buses operating the same route. Headway data was obtained from the Cap Metro Service Plan 2020. In the case of the RT_1_D variable, if a route is either Route #1L or Route #1M, then it has the value of 1; otherwise, it is zero.

Variable	Description	Source
Weekday_ Boardings	Dependent variable of linear route model	Cap Metro Service Plan 2020 - Final Report (2008)
Ln_WKBD	Dependent variable of semi-log route model	Logarithm Conversion
White_Prct	Proportion of Caucasian residents within ¼ mile buffer	Census Tract Data: Race (2010)
INC_LS_25000_PC	Percentage households with annual income < $25k / All households within ¼ mile buffer	Census Tract Data: Income (2000)
MF_acreage	Multi-family acreage within ¼ mile buffer	City of Austin: Land Use (2008)
Park_Ride_D	Park & Ride = 1, No Park & Ride = 0 within ¼ mile buffer	Cap Metro Park & Ride Location (2011)
Revenue_Hours	Revenue hours of each bus route	Cap Metro Service Plan 2020 - Final Report (2008)
Headway	The time separating two buses operating the same route (Minutes)	Cap Metro Service Plan 2020 - Final Report (2008)
RT_1_D	Route #1 = 1, Others = 0	Cap Metro: Bus Schedule (2011)

Table 2: Route-Level Variable Description

2) Route-Level Descriptive Statistics

Table 3 shows basic statistics for variables used in the linear and semi-log models. Initially, the dependent variable, Weekday Boarding shows an average of 1,644 people using each bus route during a weekday. However, the standard deviation is so high that the maximum ridership was approximately 8,000 while the minimum ridership was only 14. This shows there is great variation in ridership levels between bus routes. The average value for White_Prct was almost 70%, indicating that, overall, Caucasian ethnic groups comprised more than half of the residents within the quarter mile buffer zones around each route. On average, approximately 37% of households had an annual income of less than 25,000 dollars. The average value of multi-family acreage within the quarter-mile buffer of each route was approximately 300 acres. Twenty-three percent of routes had a Cap Metro-managed Park-and-Ride site within the quarter mile buffer of the route corridor. The average number of revenue hours was 43 hours, and this variable has a relatively high standard deviation. The average headway was almost 74 minutes, but this variable also has a high standard deviation. The Route #1 dummy variable shows that Route #1 category services take up only 3 percent of total routes in the Austin area.

Variable	N	Mean	Std Dev	Minimum	Maximum
Weekday_Boardings	80	1,644.00	2,015.00	14.00	8,027.00
Ln_WKBD	105	4.95	3.11	0.00	8.99
White_Prct	105	68.23	8.99	38.25	85.94
INC_LS_25000_PC	105	37.51	10.91	9.31	75.42
MF_acreage	105	304.14	201.99	4.41	931.72
Park_Ride_D	105	0.23	0.42	0.00	1.00
Revenue_Hours	80	43.65	38.23	1.00	147.10
Headway	80	74.05	168.72	5.00	720.00
RT_1_D	80	0.03	0.16	0.00	1.00

Table 3: Route-Level Descriptive Statistics

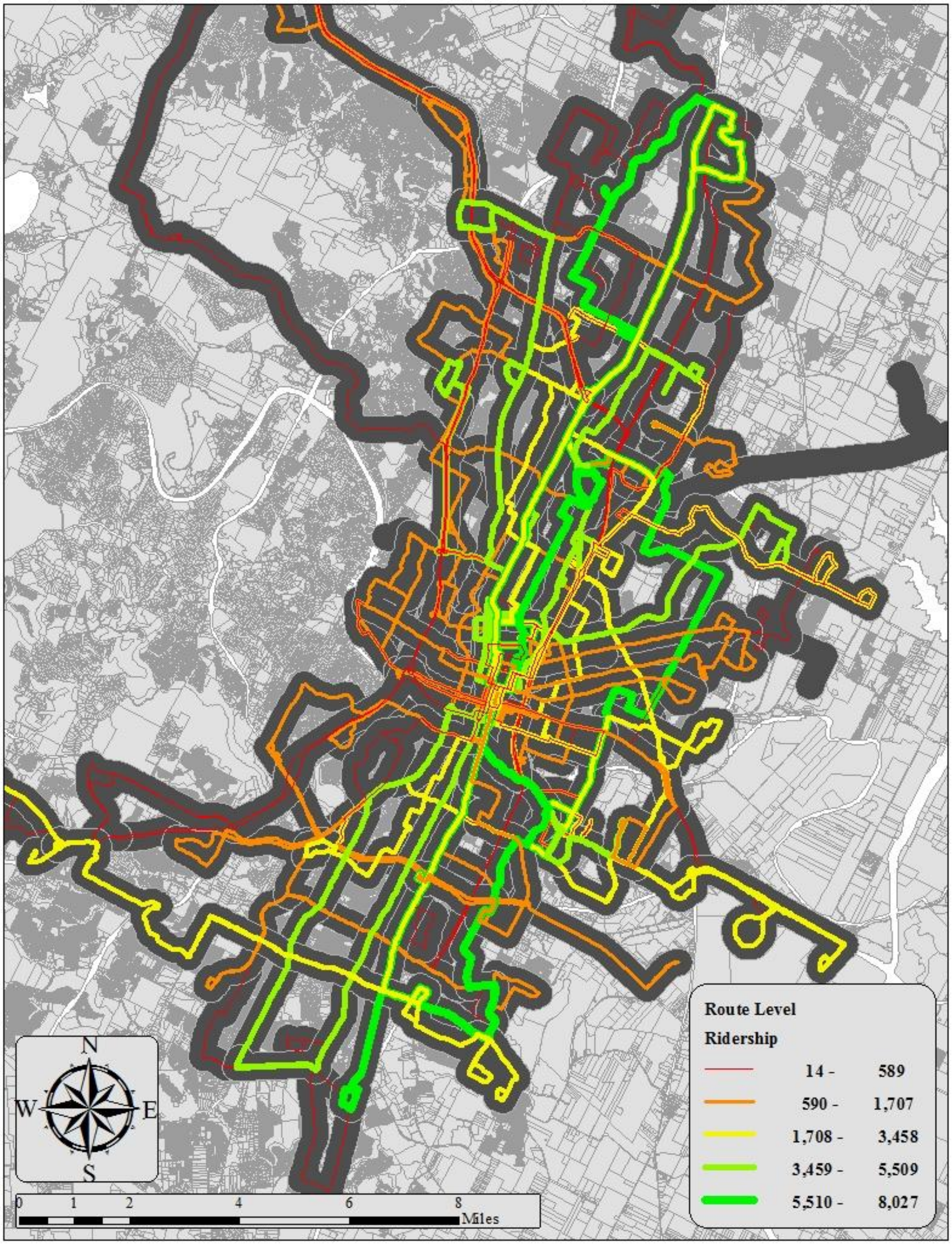

Figure 1: Route Level Analysis Dependent Variable - Ridership

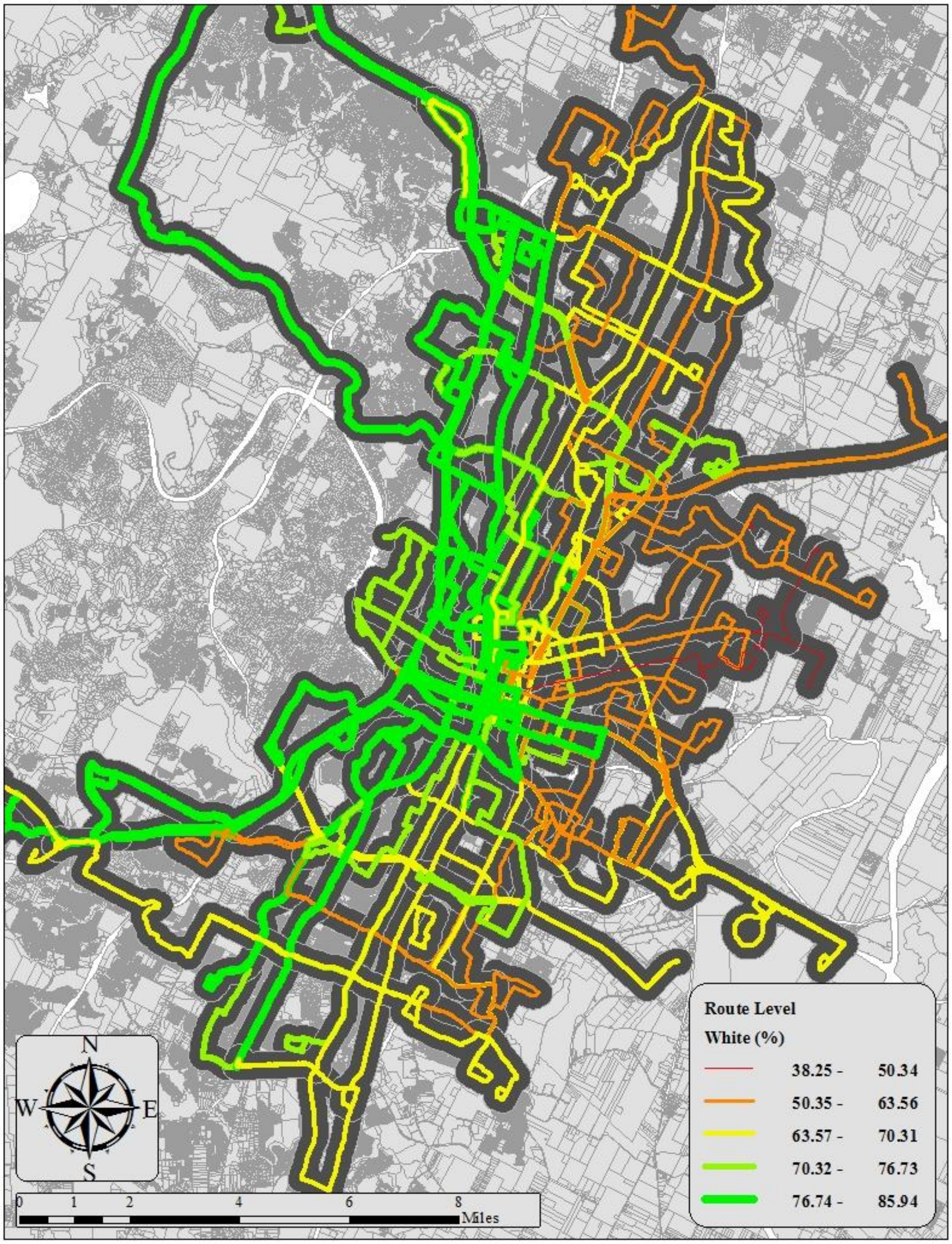

Figure 2: Route Level Analysis Independent Variable 1 - White Percentage

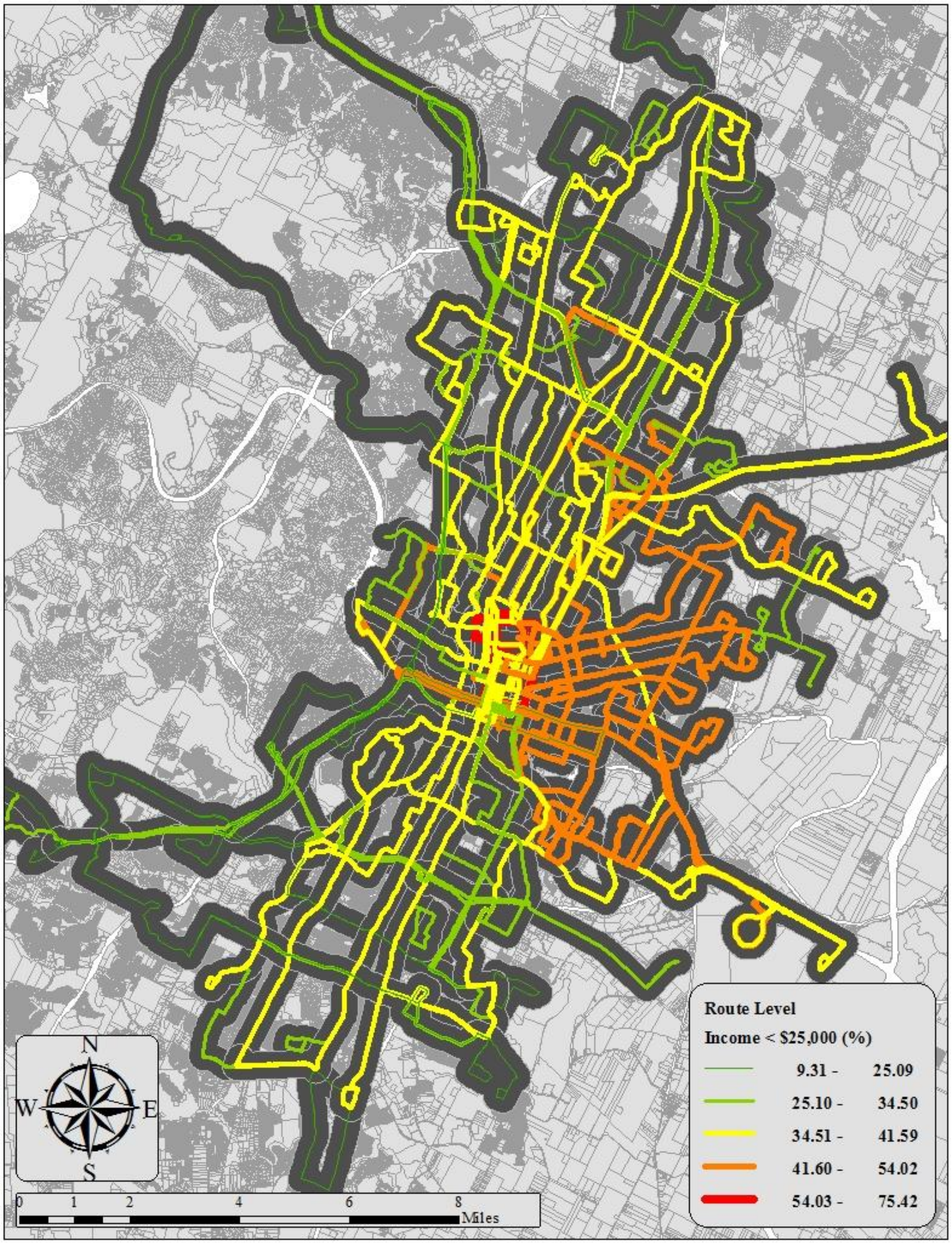

Figure 3: Route Level Analysis Independent Variable 2 - Income Percentage

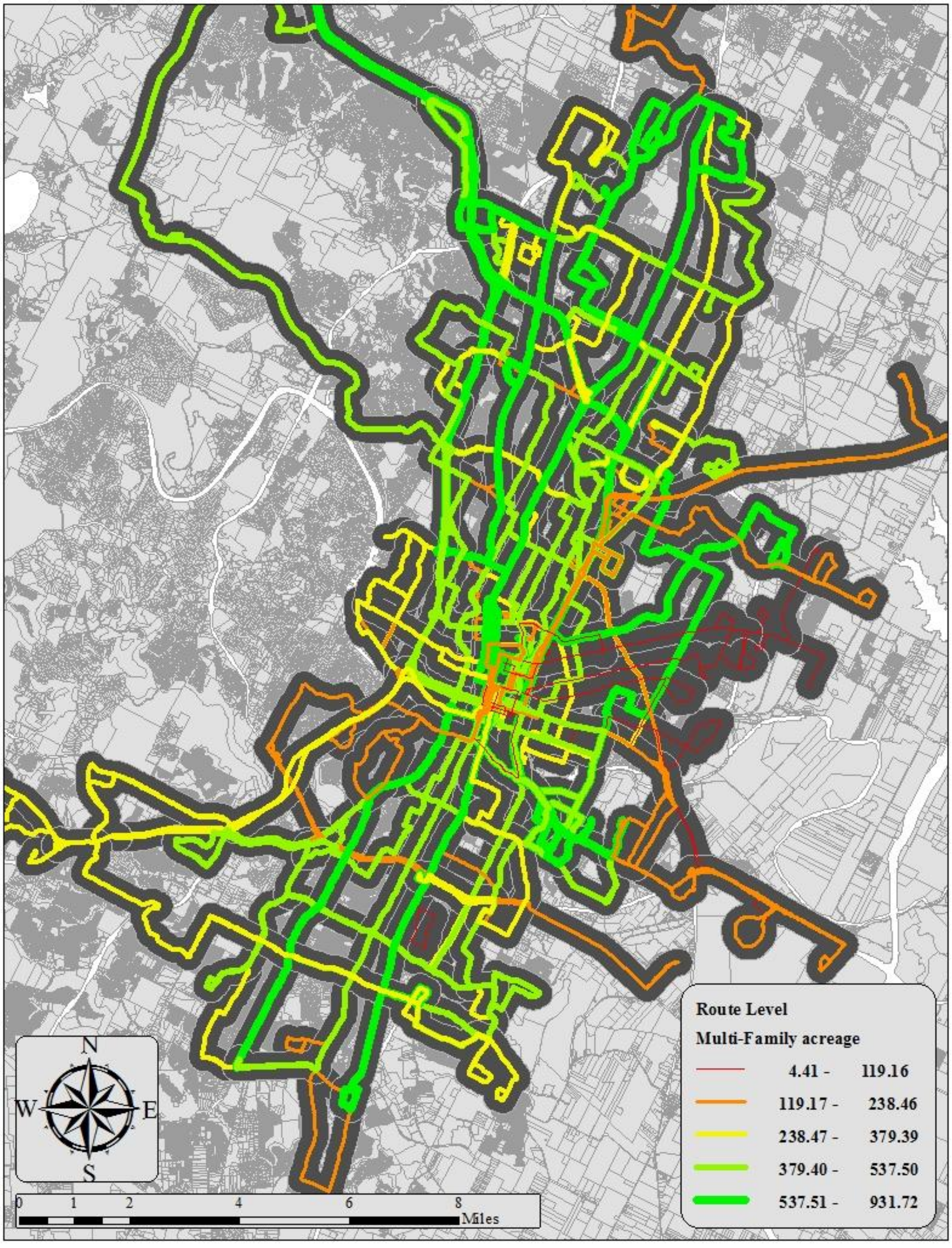

Figure 4: Route Level Analysis Independent Variable 3 - Multi-Family Acreage

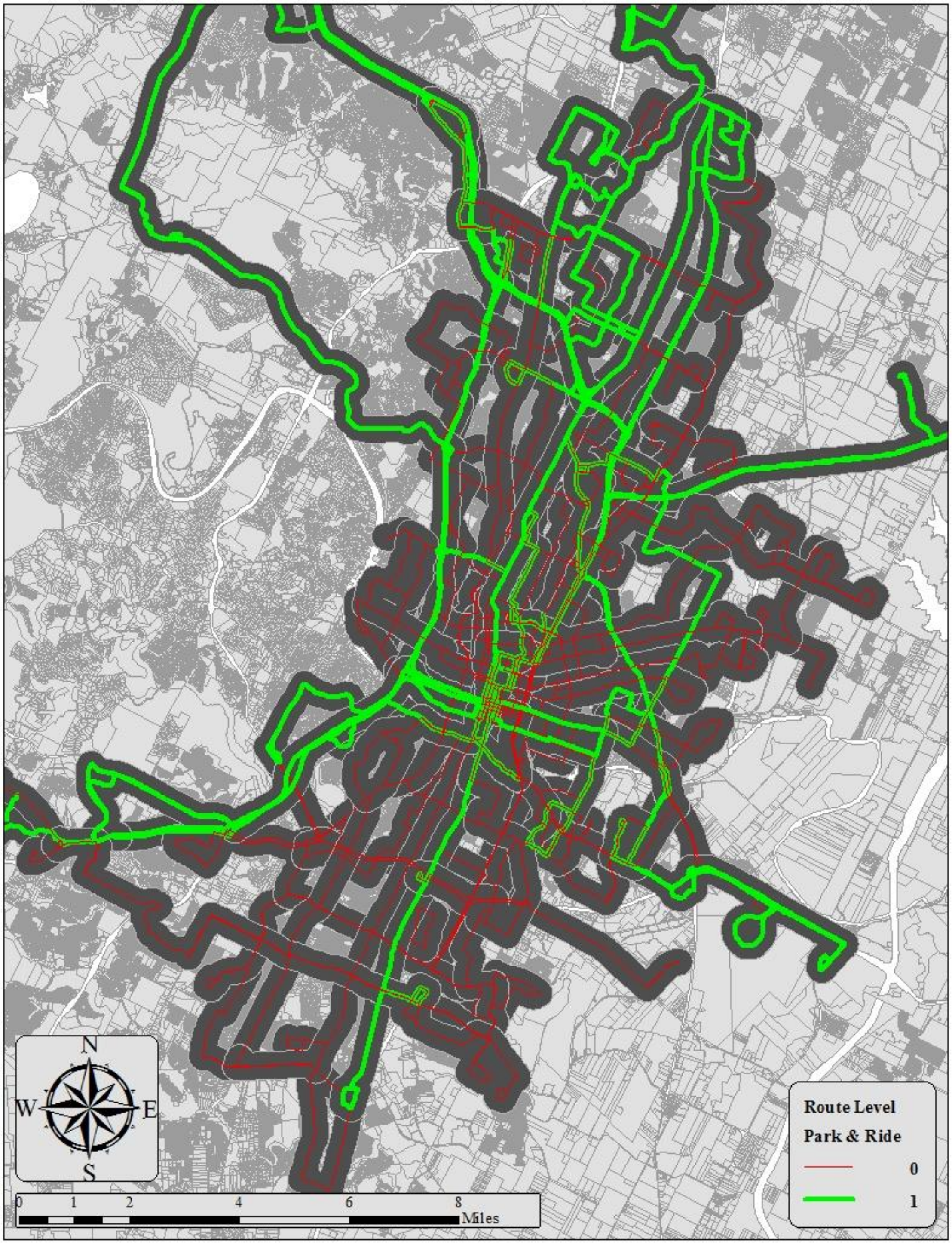

Figure 5: Route Level Analysis Independent Variable 4 - Park & Ride

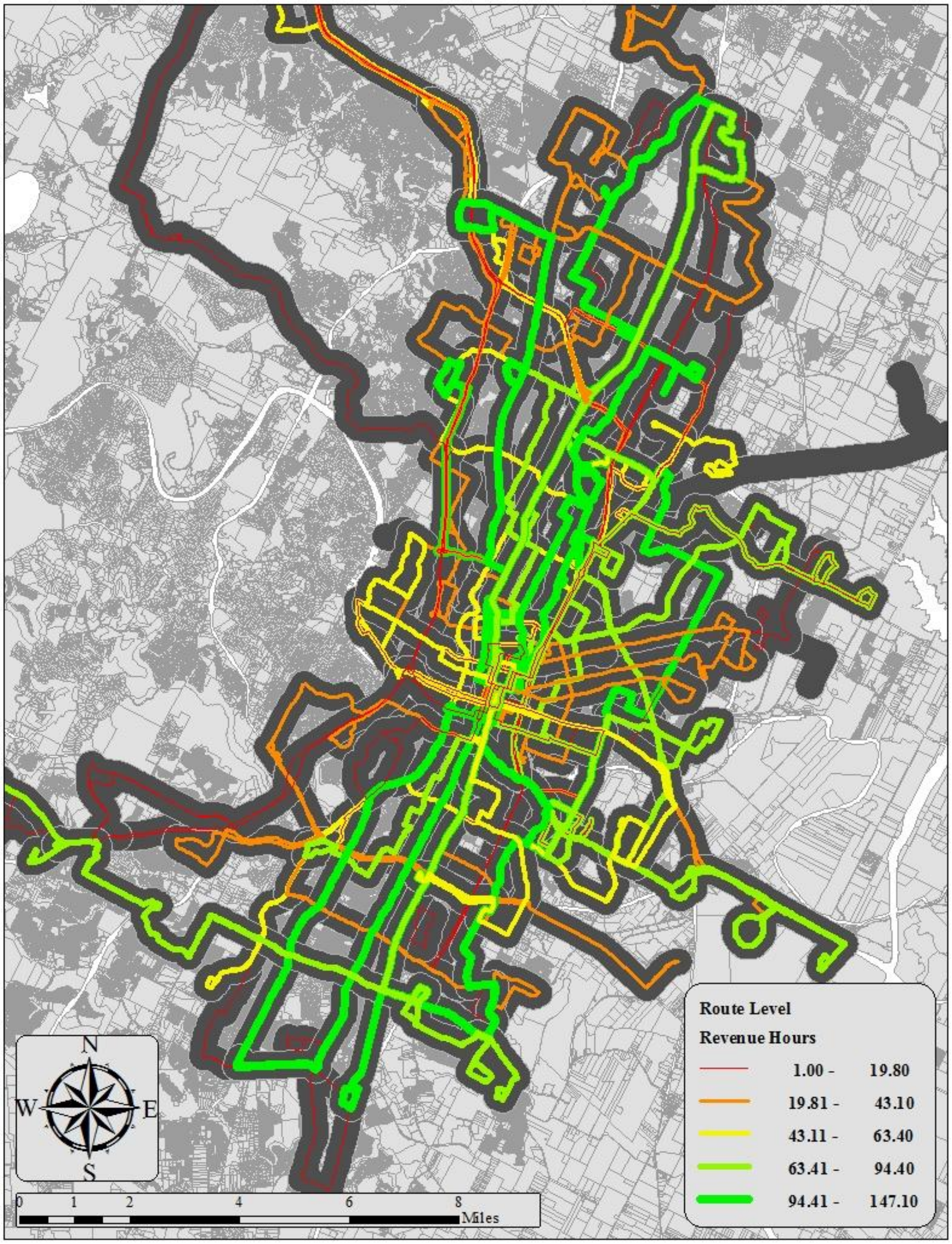

Figure 6: Route Level Analysis Independent Variable 5 - Revenue Hours

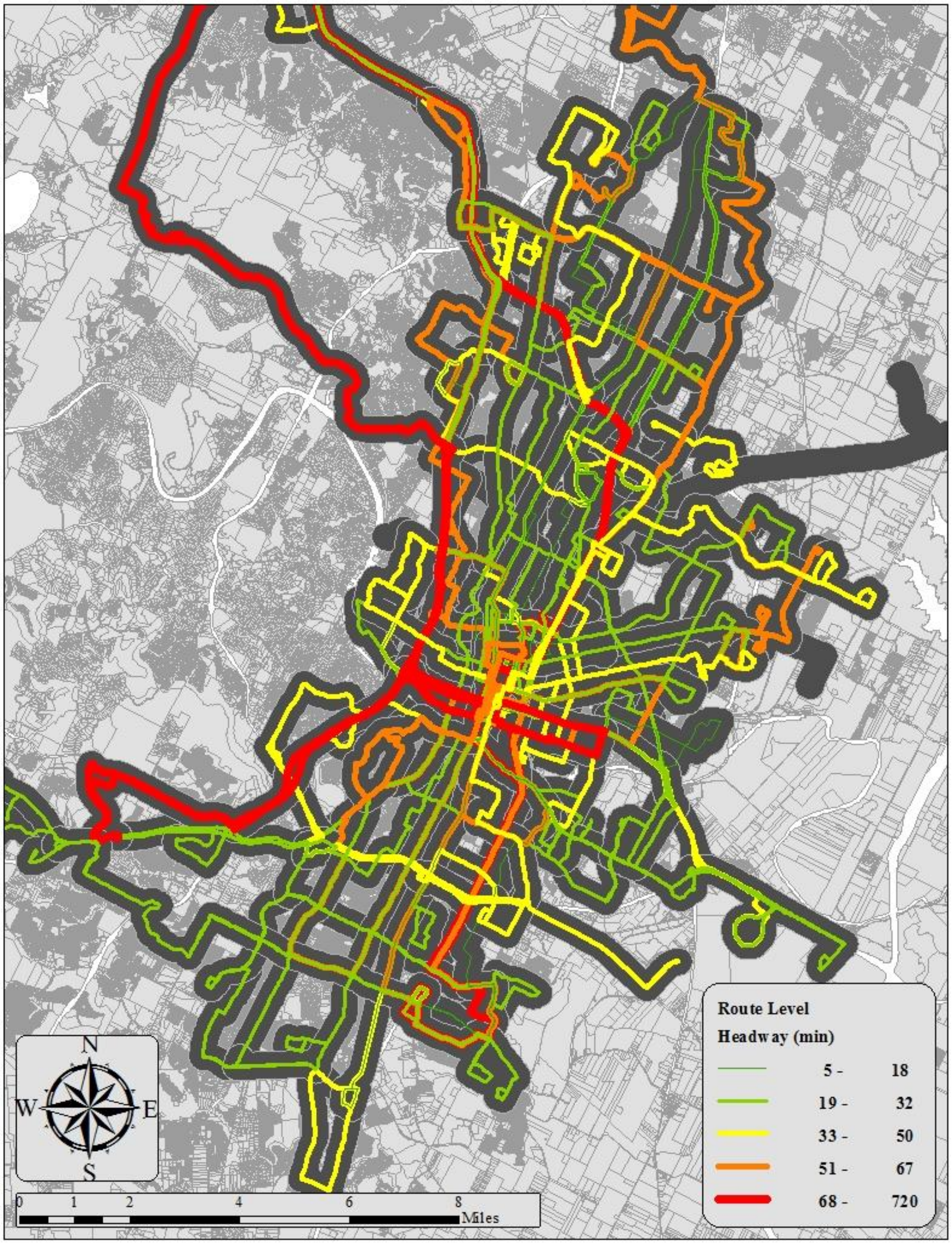

Figure 7: Route Level Analysis Independent Variable 6 - Headway

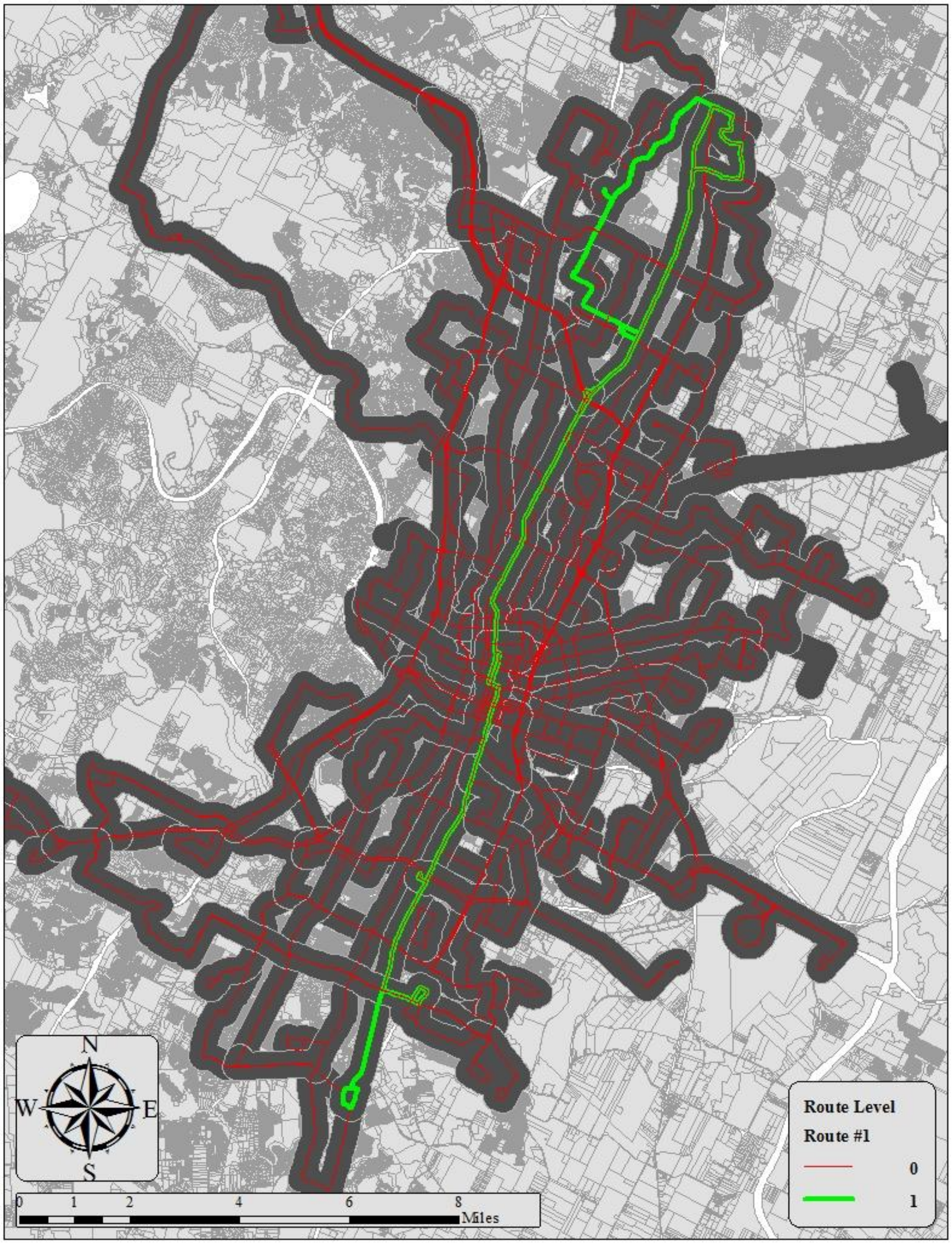

Figure 8: Route Level Analysis Independent Variable 7 - Route #1

3) Route-Level Correlation Analysis

Correlation analysis was conducted prior to the regression model analysis. Generally, a coefficient of more than 0.5 within a 10% statistical significance level between two independent variables indicates that they are correlated. Two correlated variables cannot be used simultaneously in regression models. There was no strong correlation between the seven independent variables (White_Percentage, Percentage of Household Income with less than 25000, Park-and-ride Dummy, Revenue Hours, Headway, and Route #1 Dummy), since every coefficient was less than 0.5, and none of these correlations was statistically significant.

Category	White_ Prct	INC_LS _25000 _PC	MF_ acreage	Park_ Ride_D	Revenue _Hours	Headwa y	RT_1_D
White_ Prct	1.000	-0.256	0.018	0.125	-0.014	0.173	-0.019
		0.008	0.854	0.204	0.903	0.124	0.871
INC_LS_ 25000_PC	-0.256	1.000	-0.160	-0.304	0.197	-0.273	-0.018
	0.008		0.102	0.002	0.080	0.014	0.871
MF_ acreage	0.018	-0.160	1.000	0.372	0.301	0.183	0.367
	0.854	0.102		<.0001	0.007	0.105	0.001
Park_ Ride_D	0.125	-0.304	0.372	1.000	-0.004	0.303	0.260
	0.204	0.002	<.0001		0.974	0.006	0.020
Revenue _Hours	-0.014	0.197	0.301	-0.004	1.000	-0.317	0.431
	0.903	0.080	0.007	0.974		0.004	<.0001
Headway	0.173	-0.273	0.183	0.303	-0.317	1.000	-0.060
	0.124	0.014	0.105	0.006	0.004		0.596
RT_1_D	-0.019	-0.018	0.367	0.260	0.431	-0.060	1.000
	0.871	0.871	0.001	0.020	<.0001	0.596	
- Upper Cell: Coefficient between independent variables - Lower Cell: Significance Level							

Table 4: Route-Level Correlation Analysis

4) Route-Level Regression Result

Overall, the fit of the linear model was slightly better than that of the semi-log model. In the case of the White_Percentage variable, it was not statistically significant although it had a negative coefficient value. The Low Income variable has a positive influence on ridership at the route level. It was found that, for each percentage increase in households with income less than 25,000 dollars, 56 more people use buses. This was significant at 1% level. The Multi-family variable was not statistically significant and the sign of coefficient was negative. The Park-and-Ride dummy variable was also non-significant, although it had a positive coefficient. In the case of service variable group, one hour of revenue hour increase was associated with 40 additional passengers on each bus route (significant at the 1% level). Although headway was not statistically significant, the Route #1 dummy variable had a coefficient with the value of 1,574, (significant at the 5% level). This means that the Route #1 buses attract an additional 1,574 passengers per bus daily. Amongst independent variables in the linear model, Low Income, Revenue Hours, and Route #1 Bus variables were statistically significant, with Revenue Hours being the most important, followed by Income and Route #1 according to results obtained using standardized estimation.

The semi-log model shows that low income still has a positive influence on increase in ridership. A one percentage increase in the proportion of low income households, which is the unit of the INC_LS_25000_PC variable in the analysis, results in a 2 percent increase in ridership. Also, Revenue Hours has a significant positive influence, while increased Headway has a negative influence. On the other hand, Route #1 dummy variable results in the opposite sign of coefficient to the linear model result.

Model		Linear Model		
		Coefficient		Standardized Coefficient
	Intercept	-1,994.9043	*	0.0000
Traveler Variables	White_Prct	-6.0066		-0.0255
	INC_LS_25000_PC	56.0518	***	0.2840
Land Use Variable	MF_acreage	-0.1110		-0.0109
	Park_Ride_D	364.1061		0.0812
Service Variable	Revenue_Hours	40.7379	***	0.7730
	Headway	0.6891		0.0577
	RT_1_D	1,574.8030	**	0.1228
R-Square		0.8219		
Observation		80		
***: 1% significance level, **: 5% significance level, *: 10% significance level				

Table 5: Linear Model (Route-Level Ridership Analysis)

Model		Semi-Log Model		
		Coefficient		Standardized Coefficient
	Intercept	4.3057	***	0.0000
Traveler Variables	White_Prct	0.0014		0.0076
	INC_LS_25000_PC	0.0221	**	0.1398
Land Use Variable	MF_acreage	-0.0004		-0.0440
	Park_Ride_D	0.2378		0.0663
Service Variable	Revenue_Hours	0.0342	***	0.8110
	Headway	-0.0020	***	-0.2110
	RT_1_D	-1.2123	*	-0.1182
R-Square		0.7918		
Observation		80		
***: 1% significance level, **: 5% significance level, *: 10% significance level				

Table 6: Semi-Log Model (Route-Level Ridership Analysis)

3. SEGMENTED ROUTE-LEVEL REGRESSION ANALYSIS

1) Segmented Route-Level Variable Description

Variable names and explanations for the segmented route-level analysis are shown in Table 7. The dependent variable, Ridership_Stop_Route, was obtained from the Cap Metro Service Plan 2020 Final Report. First, the number of routes running through each stop was counted; Second, ridership at each stop was divided by the number of routes intersecting each specific stop; Third, ridership on each route was proportionally distributed to the stops based on the proportion of each stop ridership from the second step. These steps were performed to proportionally distribute ridership at each route to intersecting stops, based on the number of riders at the stops. In the final step, newly distributed ridership numbers at each stop were summed, based on the boundary of each neighborhood. This dependent variable for the linear regression model is termed Ridership_StRt. The logarithm (termed Ln_Ridership_StRt) was also calculated for use in the semi-log model analysis. Where the value of Ln_Ridership_StRt was less than 0, it took the value of 0 for the regression analysis.

Data on ethnicity of passengers was available from 2010 Census Data. The variable White_Prct represents the percentage of residents from Caucasian ethnic groups within a quarter mile buffer of each route within each neighborhood. This percentage was derived from the prorated ethnic distribution from each census tract into each quarter mile buffer. The income variable (INC_LS_25000_PC) was the same as in the route-level model. The MF_acreage variable was obtained by GIS analysis of land use within quarter mile buffers of each bus route inside of neighborhood. These two files were overlaid and then all land use acreages within the quarter mile buffer were calculated. The Park_Ride_D variable was the same as that used in the route-level analysis. The

Revenue_Hours and Headway variables were available by bus routes, so were the same as those used in the route-level analysis. In the case of the UT_D variable, if a route is located on one of the UT shuttle bus lines then it has the value of 1; otherwise, it has a value of zero.

Variable	Description	Source
Ridership_StRt	Dependent variable of linear segmented route model (Proportional distribution of ridership of each route into stops and then summation of the number of riders of the stops within each neighborhood boundary)	Cap Metro Service Plan 2020 - Final Report (2008), GIS analysis
Ln_Ridership_StRt	Dependent variable of semi-log segmented route model	Logarithm Conversion
White_Prct	Proportion of Caucasian residents within ¼ mile buffer inside of each neighborhood boundary	Census Tract Data: Race (2010)
INC_LS_25000_PC	Percentage households with annual income < $25k / All households within ¼ mile buffer	Census Tract Data: Income (2000)
MF_acreage	Multi-family acreage within ¼ mile buffer inside of each neighborhood boundary	City of Austin: Land Use (2008)
Park_Ride_D	Park & Ride = 1, No Park & Ride = 0 within ¼ mile buffer	Cap Metro Park & Ride Location (2011)
Revenue_Hours	Revenue hours of each bus route	Cap Metro Service Plan 2020 - Final Report (2008)
Headway	The time separating two buses operating the same route (Minutes)	Cap Metro Service Plan 2020 - Final Report (2008)
UT_D	UT shuttle buses = 1, Others = 0	Cap Metro: Bus Schedule (2011)

Table 7: Segmented Route-Level Variable Description

2) Segmented Route-Level Descriptive Statistics

Table 8 shows the basic statistics for variables used in the linear and semi-log models. The dependent variable, Ridership_StRt shows an average of 190 people using bus stops in each segmented route each weekday. This is comparably small if considering that the average ridership at route-level was 1,644. As in the route-level analysis, the standard deviation remains so high that the maximum ridership was approximately 7,715 while the minimum ridership was zero. This result indicates that there are also major variations between segmented bus routes. Caucasian residents comprised almost 70%, on average, although the range was almost 60%. On average, 38% of households had an annual income of less than 25,000 dollars. The average value of multi-family acreage within the quarter-mile buffers of each route within neighborhood boundaries was approximately 35 acres. Thirty percent of route segments had a Cap Metro-managed Park-and-Ride site, which was slightly higher than in the route-level analysis. The average number of revenue hours was 54 hours and the maximum value of revenue hours was 147 hours while the minimum value was 1 hour. The average headway was almost 64 minutes, but it also has a relatively high standard deviation of 145 minutes. The longest time between two buses operating on the same route was 720 minutes, while the shortest headway was 5 minutes. The UT shuttle bus dummy variable shows that UT shuttle category services take up 12 percent of all routes in the Austin area.

Variable	N	Mean	Std Dev	Minimum	Maximum
Ridership_StRt	692	190.05	439.06	0.00	7,715.00
Ln_Ridership_StRt	692	3.74	1.98	0.00	8.95
White_Prct	692	69.23	13.85	29.10	92.90
INC_LS_25000_PC	692	37.58	7.65	14.81	75.42
MF_acreage	692	35.34	56.13	0.00	519.96
Park_Ride_D	692	0.30	0.46	0.00	1.00
Revenue_Hours	692	54.27	45.54	1.00	147.10
Headway	692	63.84	145.45	5.00	720.00
UT_D	692	0.12	0.32	0.00	1.00

Table 8: Segmented Route-Level Descriptive Statistics

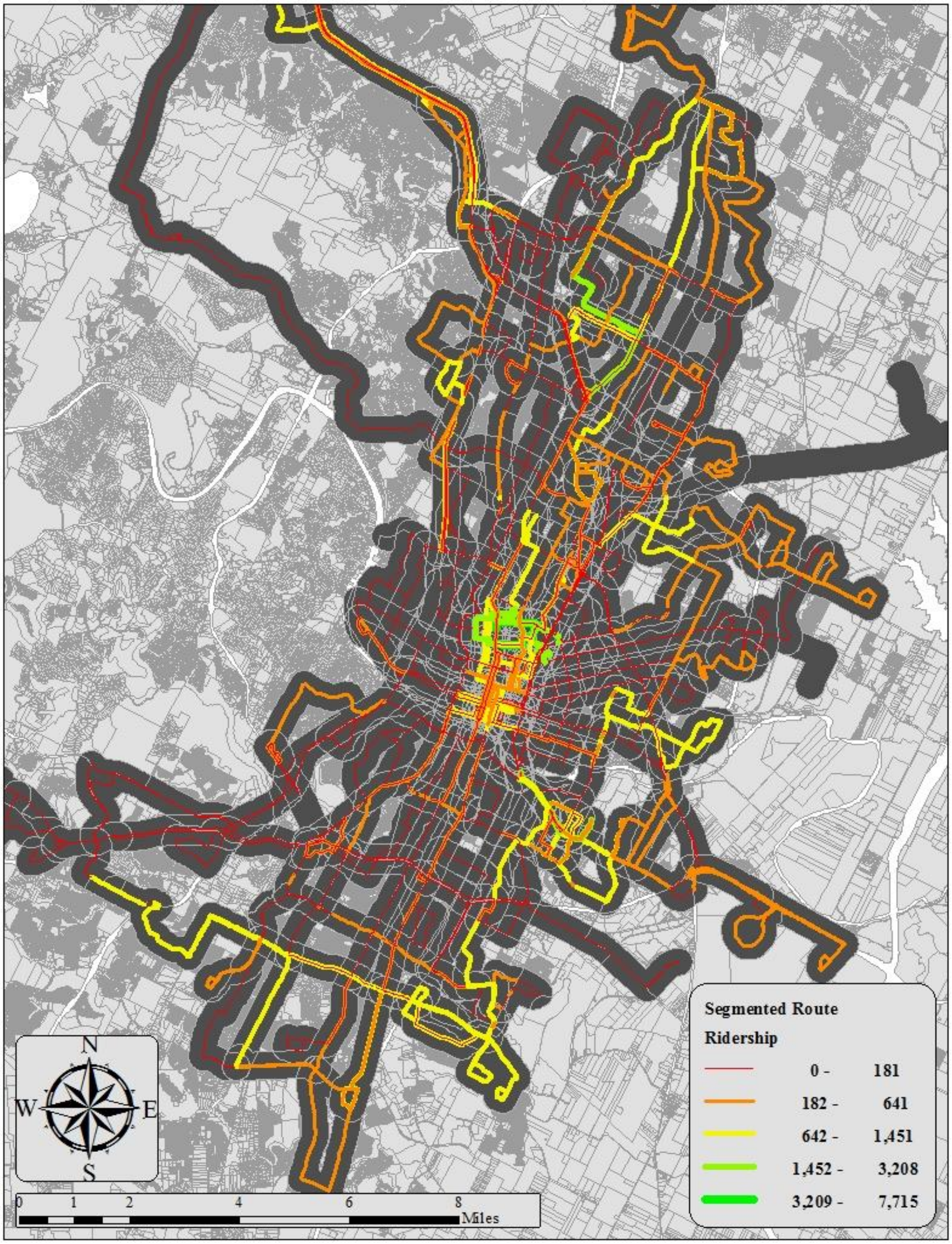

Figure 9: Segmented Route Level Analysis Dependent Variable - Ridership

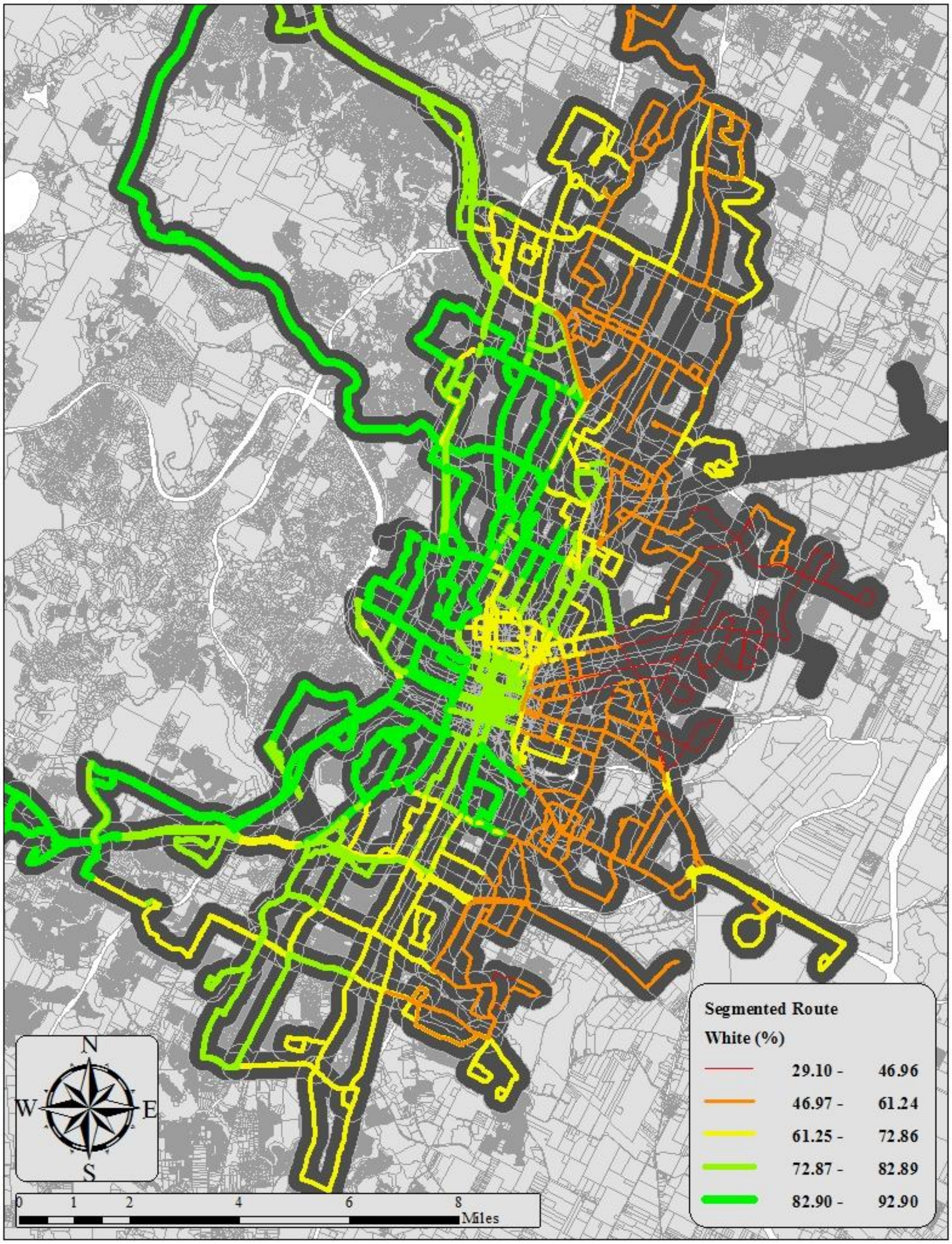

Figure 10: Segmented Route Level Analysis Independent Variable 1 - White Percentage

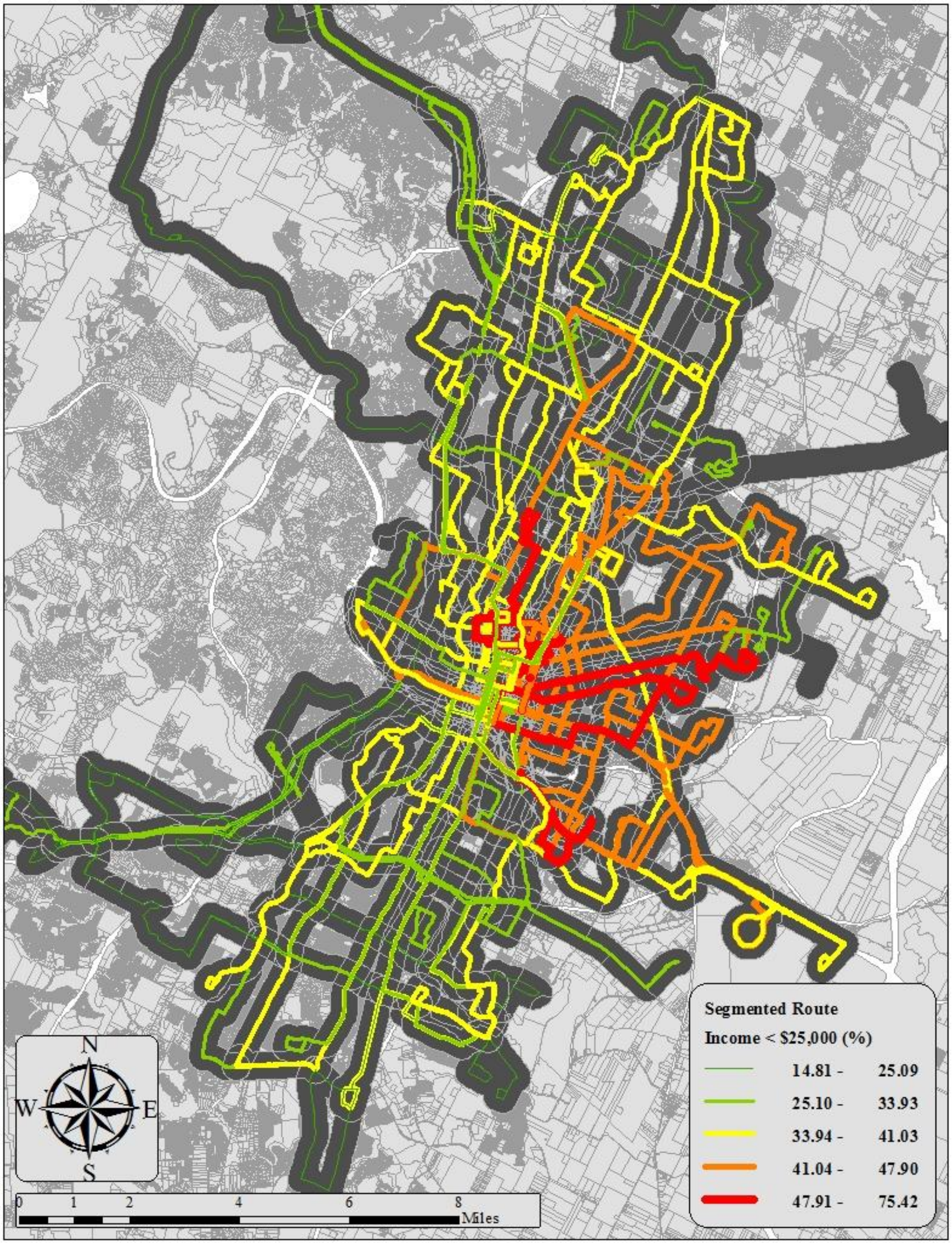

Figure 11: Segmented Route Level Analysis Independent Variable 2 - Income Percentage

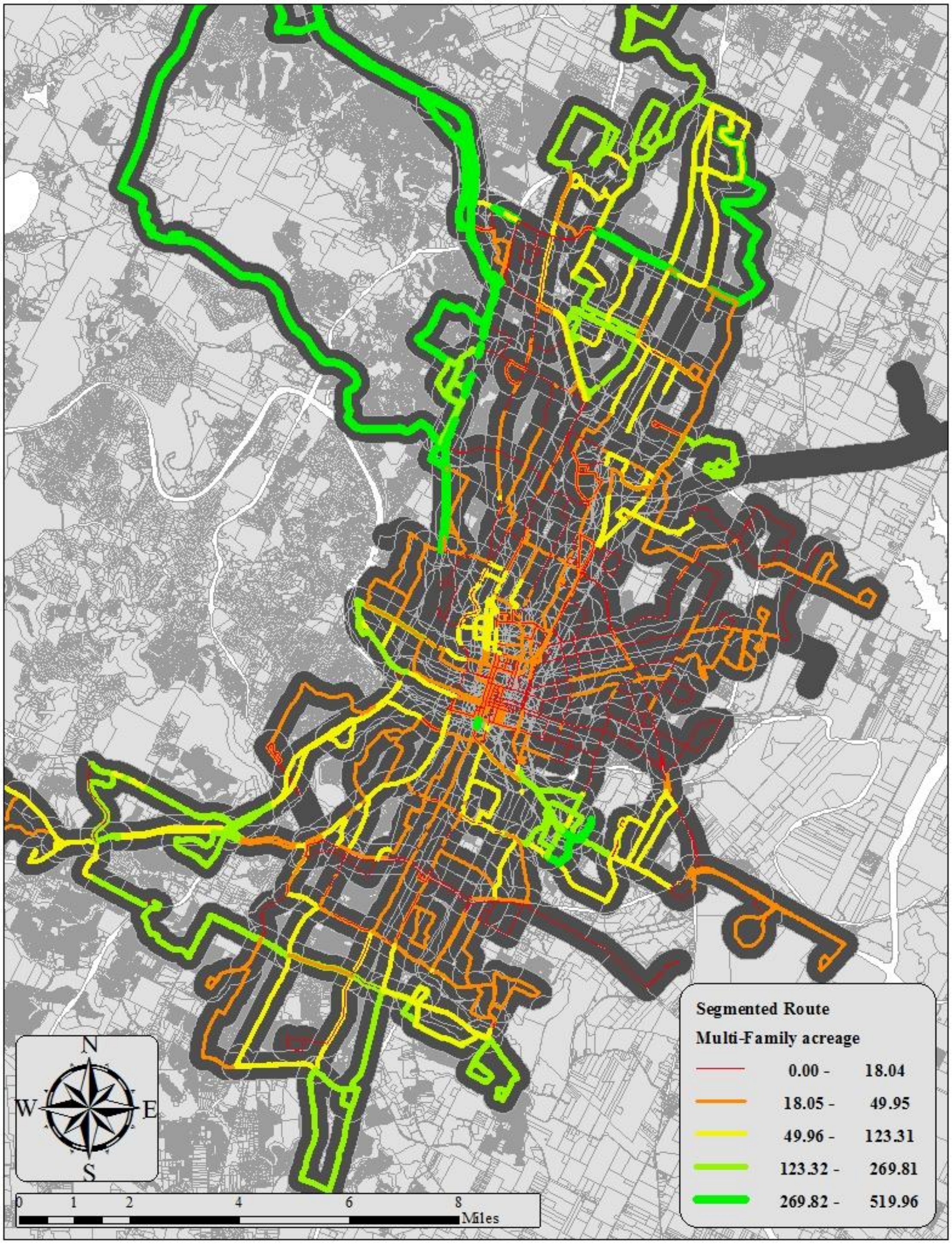

Figure 12: Segmented Route Level Analysis Independent Variable 3 - Multi-Family Acreage

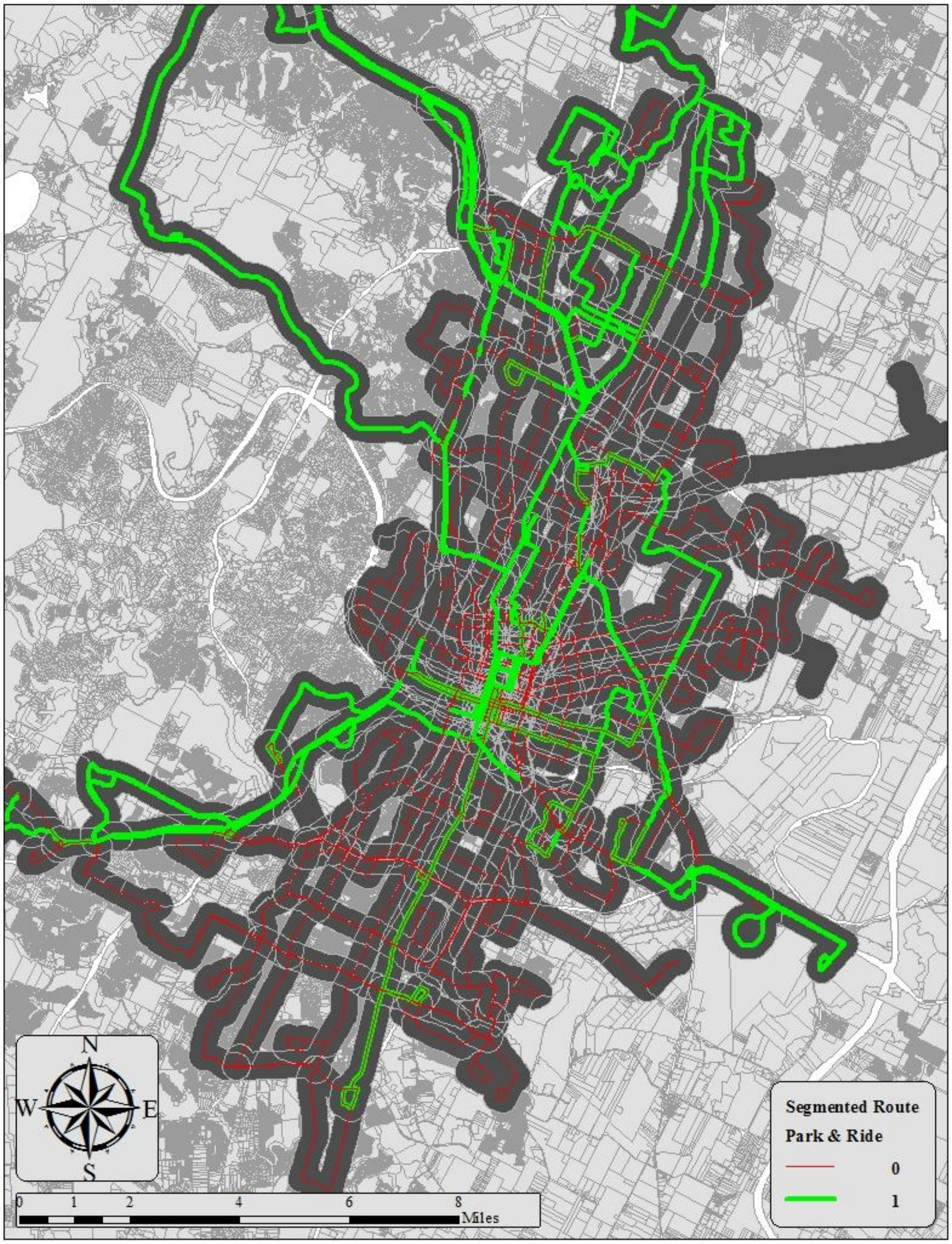

Figure 13: Segmented Route Level Analysis Independent Variable 4 - Park & Ride

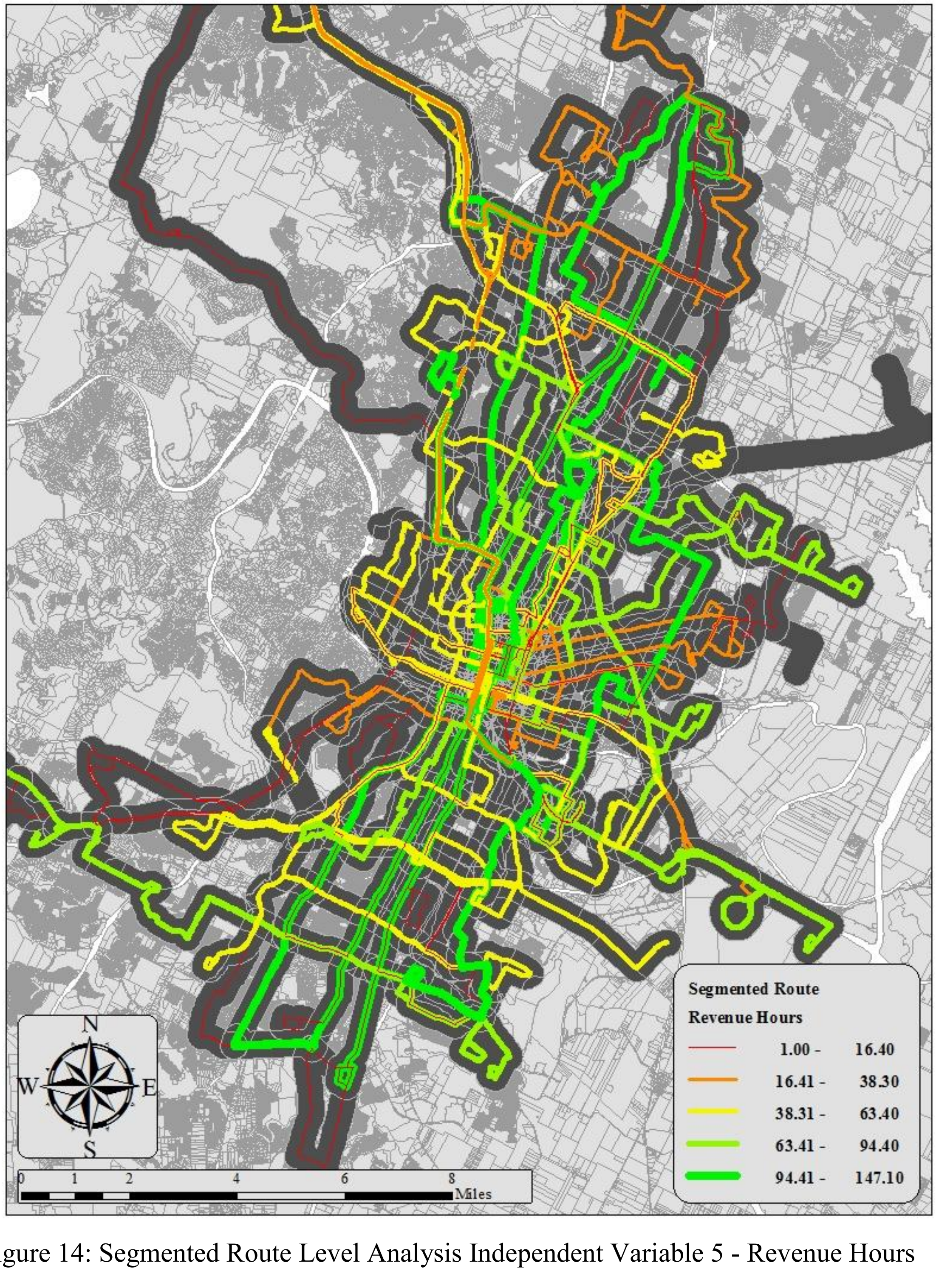

Figure 14: Segmented Route Level Analysis Independent Variable 5 - Revenue Hours

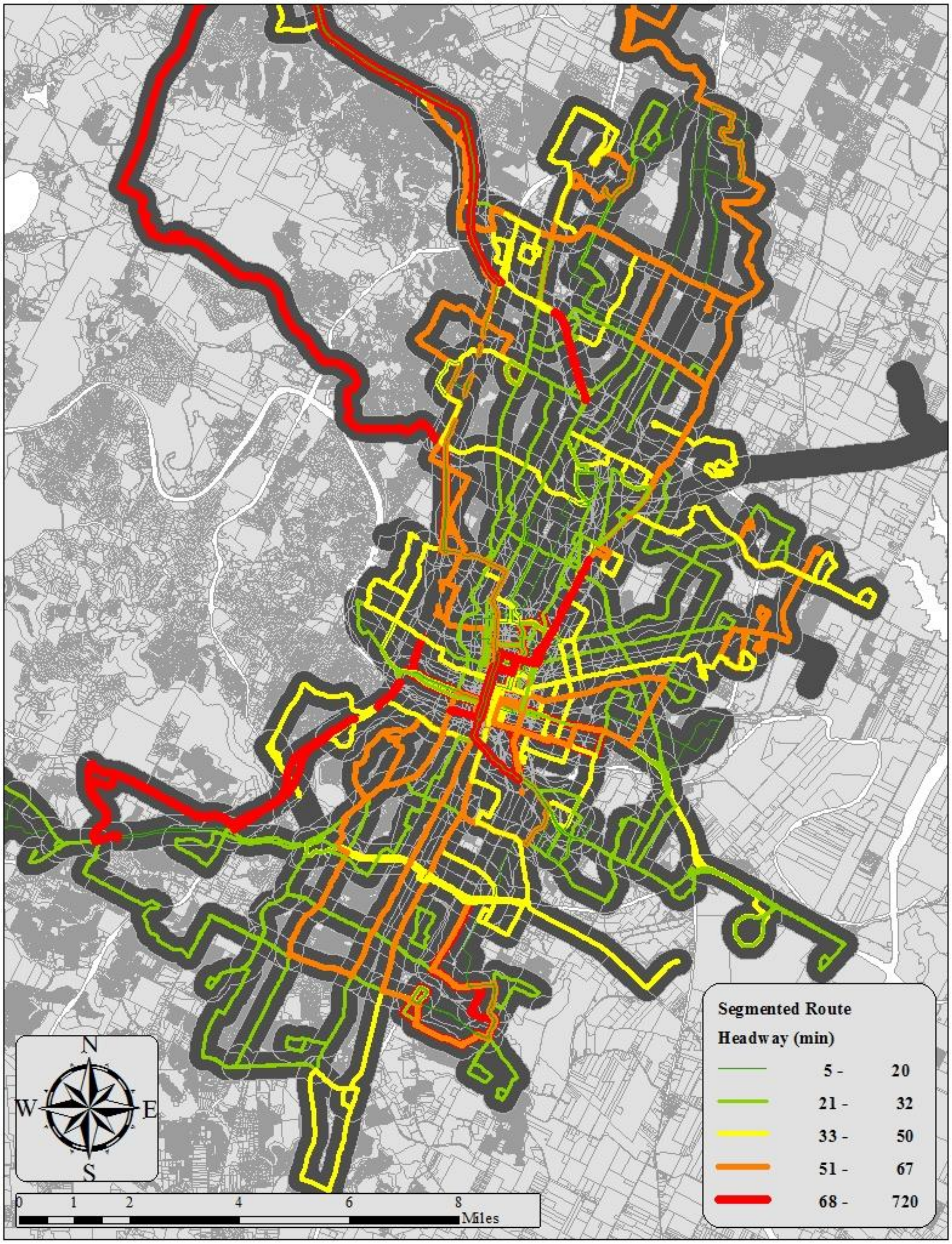

Figure 15: Segmented Route Level Analysis Independent Variable 6 - Headway

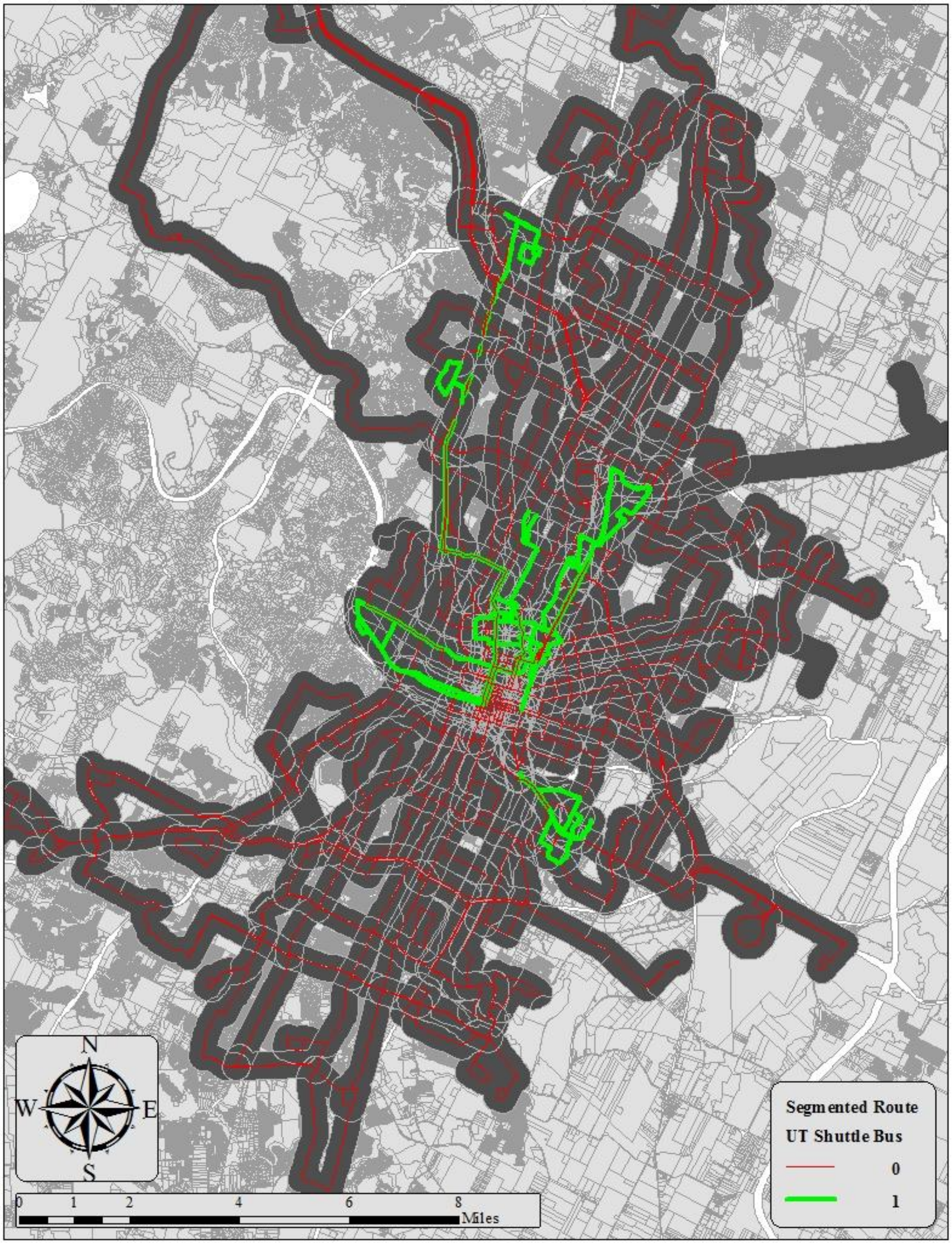

Figure 16: Segmented Route Level Analysis Independent Variable 7 - UT Shuttle Bus

3) Segmented Route-Level Correlation Analysis

Correlation analysis was conducted prior to regression modeling. In order to eliminate independent variables with coefficients of more than 0.5 within 10 % statistical significance level between two independent variables, the correlation analysis step was conducted. Although some variables have a tendency to make the regression model better, correlated variables might distort the overall fit of the model; moreover, this is an indication that variables are not actually independent of each other. No strong correlation was found between the 7 independent variables used in the model (White_Percentage, Percentage of Household Income with less than 25000, Park-and-ride Dummy, Revenue Hours, Headway, and UT Shuttle Bus Dummy), as every coefficient was less than 0.5 and none of them was statistically significant.

Category	White_Prct	INC_LS_25000_PC	MF_acreage	Park_Ride_D	Revenue_Hours	Headway	UT_D
White_Prct	1.000	-0.225	-0.082	0.081	0.027	0.047	-0.016
		<.0001	0.032	0.033	0.472	0.217	0.675
INC_LS_25000_PC	-0.225	1.000	-0.088	-0.234	0.151	-0.276	0.393
	<.0001		0.020	<.0001	<.0001	<.0001	<.0001
MF_acreage	-0.082	-0.088	1.000	0.091	-0.021	0.067	0.124
	0.032	0.020		0.017	0.590	0.079	0.001
Park_Ride_D	0.081	-0.234	0.091	1.000	0.154	0.208	-0.199
	0.033	<.0001	0.017		<.0001	<.0001	<.0001
Revenue_Hours	0.027	0.151	-0.021	0.154	1.000	-0.302	-0.056
	0.472	<.0001	0.590	<.0001		<.0001	0.141
Headway	0.047	-0.276	0.067	0.208	-0.302	1.000	-0.117
	0.217	<.0001	0.079	<.0001	<.0001		0.002
UT_D	-0.016	0.393	0.124	-0.199	-0.056	-0.117	1.000
	0.675	<.0001	0.001	<.0001	0.141	0.002	

- Upper Cell: Coefficient between independent variables
- Lower Cell: Significance Level

Table 9: Segmented Route-Level Correlation Analysis

4) Segmented Route-Level Regression Result

Overall, the fit of the semi-log model was better than the linear model. In the linear model, the White_Percentage variable was statistically significant (coefficient -1.8814), which means that a one percentage increase in Caucasian ethnic population results in a 1.88 ridership decreases at a bus route within each neighborhood segmentation. Low income had a positive influence on segmented route-level ridership, as observed in the route-level analysis. A one-percentage increase in households with income less than 25,000 dollars was associated with 10 more people using buses. This was significant at the 1% level, as in the route-level analysis. The Multi-family variable was statistically significant at the 1% level (coefficient 1.3716), meaning that a one-acre increase in multi-family land use within a quarter mile buffer of every route within each neighborhood is responsible for an increase of 1.37 people using buses in the neighborhood. The Park-and-Ride dummy variable had a positive coefficient but had no significant effect. The Park-and-Ride result was similar to the route-level analysis. In the case of the service variables group, a one hour increase in revenue hours results in 2.59 additional passengers at each bus route in a neighborhood (significant at the 1% level). Although headway was not statistically significant, the UT shuttle bus dummy variable had a coefficient of 253, which was significant at the 1% level. This means a bus had an additional 253 riders in a neighborhood if a bus stop was located on one of the UT shuttle bus routes.

Amongst independent variables in the linear model, Caucasian residents, Low Income, Multi-family Acreage, Revenue Hours, and UT Shuttle Bus variables were statistically significant. It was found that Revenue Hours was the most important variable, followed by UT Shuttle Bus, Income, Multi-Family, and White_Percentage through the results of standardized estimation.

The results of the semi-log model show that an increase in the proportion of Caucasian residents was associated with a reduction in bus ridership, where a 1% increase in Caucasian residents was associated with a 0.42% decrease in ridership of bus routes within a neighborhood. It was also found that low income still has a positive influence on increase in ridership. A one-percentage increase in low-income households results in a 0.86 percent increase in ridership in a neighborhood. Multi-Family acreage still has the positive sign with the value of 0.0010, which means one acre increase in multi-family land use results in 0.10 percent increase in ridership at the segmented route-level separated by neighborhood boundaries. Revenue Hours has a statistically positive influence on ridership at the segmented route level while increased Headway has a negative influence on the ridership, where a one hour increase in revenue hours results in a 0.13 percent increase in ridership, while a one-minute increase in headway results in a 0.04 percent decrease in ridership. UT shuttle bus dummy variable results in the positive sign of coefficient with the value of .1928, which means that a UT shuttle bus increases ridership within a neighborhood by 19.28 percent.

Model		Linear Model		
		Coefficient		Standardized Coefficient
	Intercept	-301.5728	**	0.0000
Traveler Variables	White_Prct	-1.8814	*	-0.0593
	INC_LS_25000_PC	10.5150	***	0.1832
Land Use Variable	MF_acreage	1.3716	***	0.1754
	Park_Ride_D	14.0249		0.0146
Service Variable	Revenue_Hours	2.5910	***	0.2687
	Headway	0.0528		0.0175
	UT_D	253.5762	***	0.1868
R-Square		0.2125		
Observation		692		
***: 1% significance level, **: 5% significance level, *: 10% significance level				

Table 10: Linear Model (Segmented Route-Level Ridership Analysis)

Model		Semi-Log Model		
		Coefficient		Standardized Coefficient
	Intercept	0.4892	***	0.0000
Traveler Variables	White_Prct	-0.0042	***	-0.0797
	INC_LS_25000_PC	0.0086	*	0.0591
Land Use Variable	MF_acreage	0.0010	***	0.2500
	Park_Ride_D	0.1312		0.0085
Service Variable	Revenue_Hours	0.0013	***	0.5671
	Headway	-0.0004	***	-0.1166
	UT_D	0.1928	**	0.0793
R-Square		0.4660		
Observation		80		
***: 1% significance level, **: 5% significance level, *: 10% significance level				

Table 11: Semi-Log Model (Segmented Route-Level Ridership Analysis)

4. Stop-Level Regression Analysis

1) Stop-Level Variable Description

Variable names and explanations for the stop-level analysis are shown in Table 12. The dependent variable, or Total, was obtained from Cap Metro. The logarithmic value was taken for the semi-log model analysis and the variable was termed Ln_Total. Where the value of Ln_Total was less than 0, it took the value of 0 for the regression analysis.

Data on the ethnicity of residents was obtained from 2010 Census Data. The variable White_Prct represents the percentage of Caucasian residents within a quarter mile buffer of each bus stop. This percentage was derived from the prorated number of ethnic distribution from each census tract into each quarter mile buffer. In the case of income, data from the latest 2000 census was used. The variable INC_LS_25000_PC represents the percentage of households within a quarter mile buffer that have an annual income less than 25,000 dollars. The prorated number of income distribution was derived from GIS spatial analysis using 2000 Census Tract GIS shape file based on quarter mile buffers from each bus stop. The MF_acreage variable was obtained by GIS analysis of land use within quarter mile buffers of each bus stop. These two files were overlaid and the land use acreages within the quarter mile buffer were calculated. The Parking_acreage variable was obtained from CapMetro. The value was also calculated by GIS analysis after overlaying a parking shape file onto the quarter-mile buffer shape-file. The Frequency_per_day variable is the number of daily bus departures at each bus stop. This was calculated using the Cap Metro bus route schedule. UTline_D is a dummy variable showing whether a stop is located on any UT shuttle bus routes. UT_NBHD is a dummy variable representing whether a stop is located within the UT neighborhood.

Count indicates the number of bus stops within a quarter mile buffer from each bus stop, and is used to estimate the impact of clustered bus stops.

Variable	Description	Source
Total	Dependent variable of linear stop model	Cap Metro
Ln_Total	Dependent variable of semi-log stop model	Logarithm Conversion
White_Prct	Proportion of Caucasian residents within ¼ mile buffer	Census Tract Data: Race (2010)
INC_LS_25000_PC	Percentage households with annual income < $25k / All households within ¼ mile buffer	Census Tract Data: Income (2000)
MF_acreage	Multi-family acreage within ¼ mile buffer	City of Austin: Land Use (2008)
Parking_acreage	Parking acreage within ¼ mile buffer	Cap Metro Park & Ride Location
Frequency_per_day	Number of bus services per day	Cap Metro: Bus Schedule (2011)
UTline_D	UT shuttle bus route = 1, Others = 0	Cap Metro: Bus Schedule (2011)
UT_NBHD	UT neighborhood area = 1, Others = 0	City of Austin: Neighborhood (2010)
Count	Number of bus stops within ¼ mile buffer from each bus stop	GIS analysis

Table 12: Stop-Level Variable Description

2) Stop-Level Descriptive Statistics

Table 13 shows statistics for variables included in the linear and semi-log models. The average ridership calculated across all stops was 79.63 with high standard deviation. The busiest bus stop has approximately 5,000 passengers while one of the stops in Austin has only 0.04 riders. Caucasian groups made up approximately 67 percent of residents and 33 percent of households earned less than 25,000 dollars annually. The average value of multi-family acreage within quarter-mile buffers from each bus stop was approximately 12 acres. The average value of parking acreage estimated by the Cap Metro was 13.74 acres. The average bus frequency per day at each stop was 34.68 bus departures (maximum 141 per day, minimum 1). Almost 12 percent of bus stops are on all UT shuttle bus routes. However, in this study, due to technical difficulties, every bus stop on UT bus routes was included in this dummy variable regardless of whether or not the UT shuttle buses actually stopped there. Only 3 percent of bus stops are located within the UT neighborhood boundary and there is an average of 9.53 bus stops located within the quarter mile buffers of each bus stop.

Variable	N	Mean	Std Dev	Minimum	Maximum
TOTAL	3016	79.63	270.00	0.04	5,305.00
Ln_TOTAL	3016	2.97	1.61	0.00	8.58
White_Prct	2877	67.02	15.55	28.97	96.05
INC_LS_25000_PC	2877	33.31	15.08	2.29	82.97
MF_acrerage	2879	11.90	16.50	0.00	115.14
Parking_acreage	3016	13.74	12.32	0.00	58.85
Frequency_per_day	2632	34.68	17.57	1.00	141.00
UTline_D	3016	0.12	0.32	0.00	1.00
UT_NBHD	2879	0.03	0.16	0.00	1.00
Count	2879	9.53	7.26	1.00	56.00

Table 13: Stop-Level Descriptive Statistics

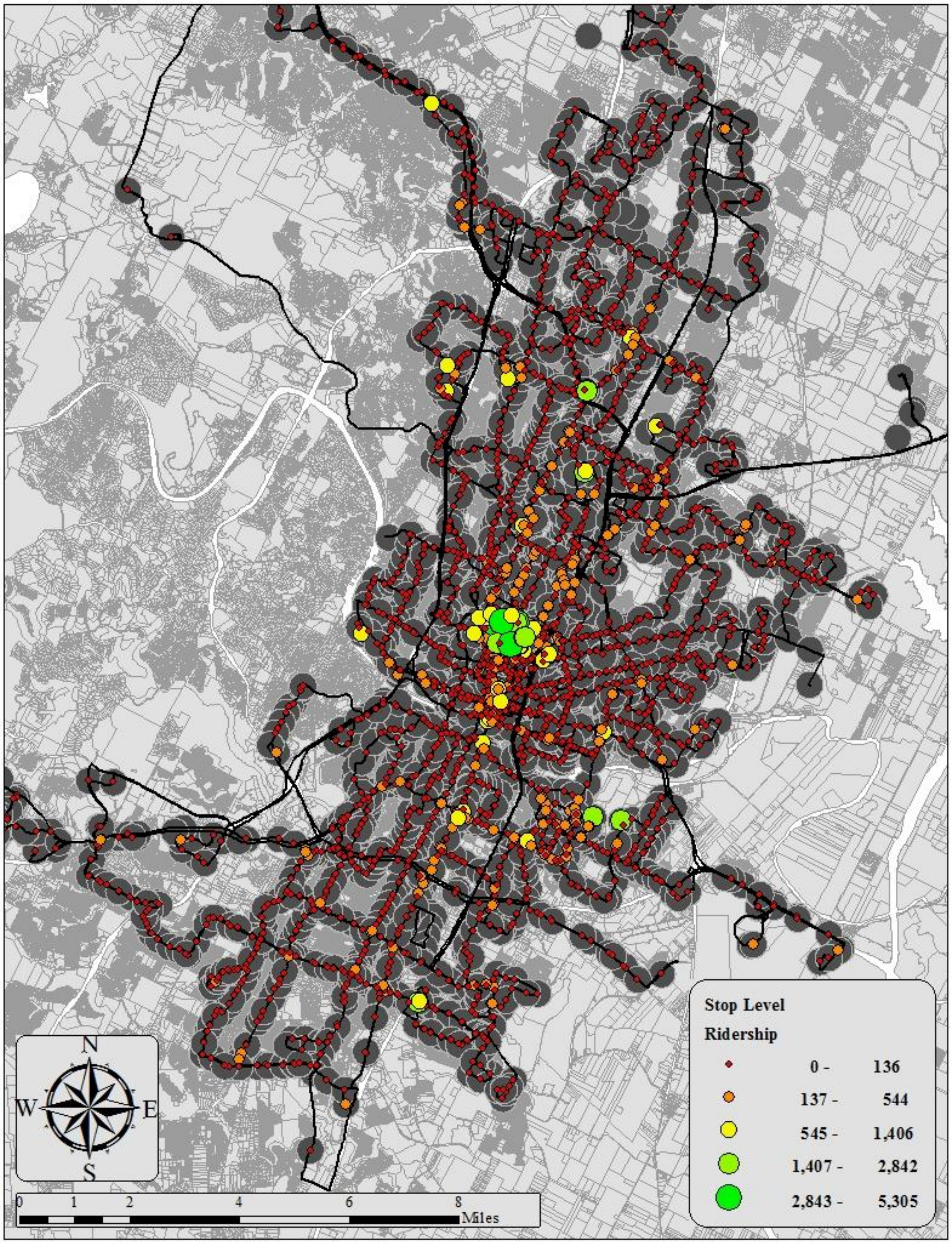

Figure 17: Stop Level Analysis Dependent Variable - Ridership

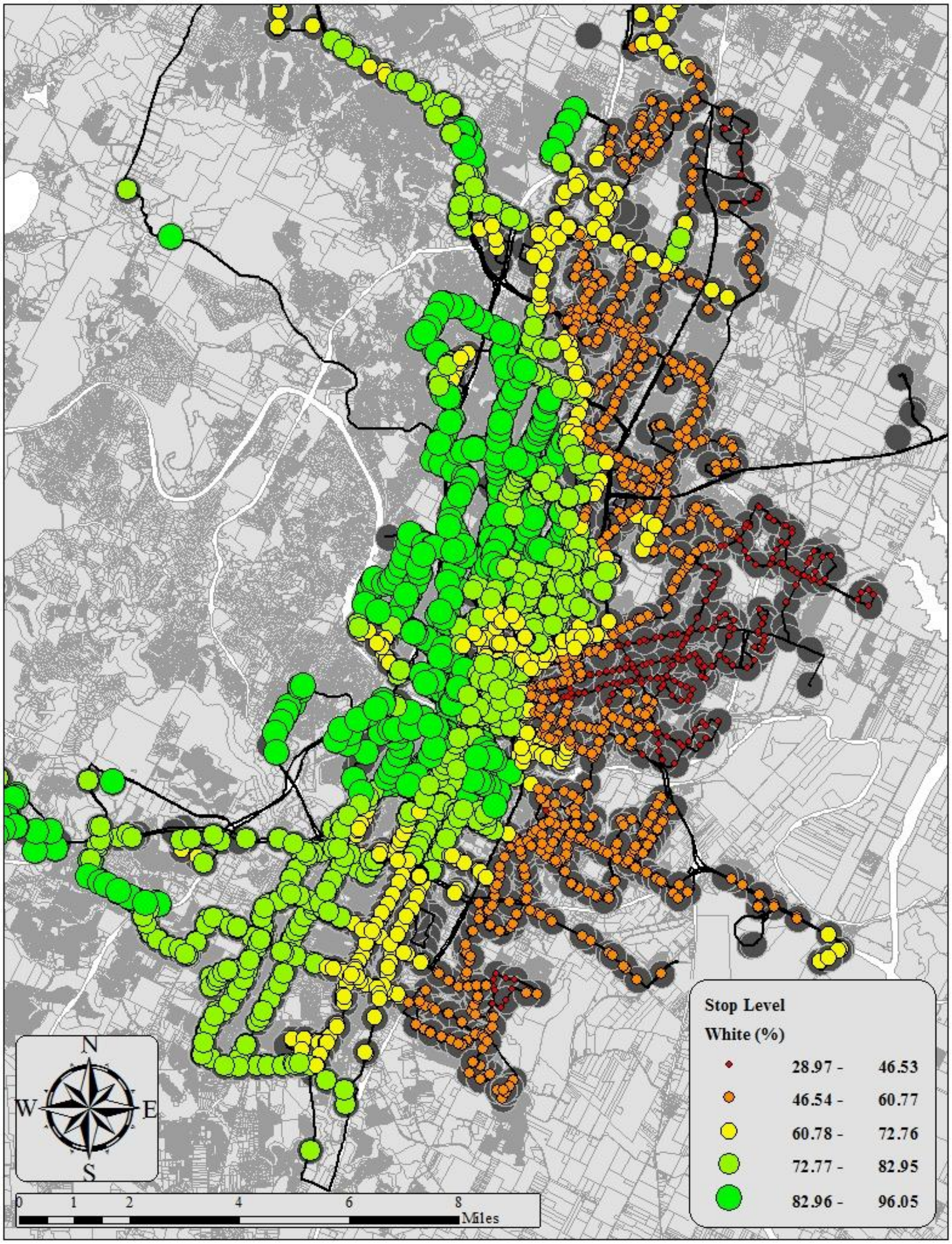

Figure 18: Stop Level Analysis Independent Variable 1 - White Percentage

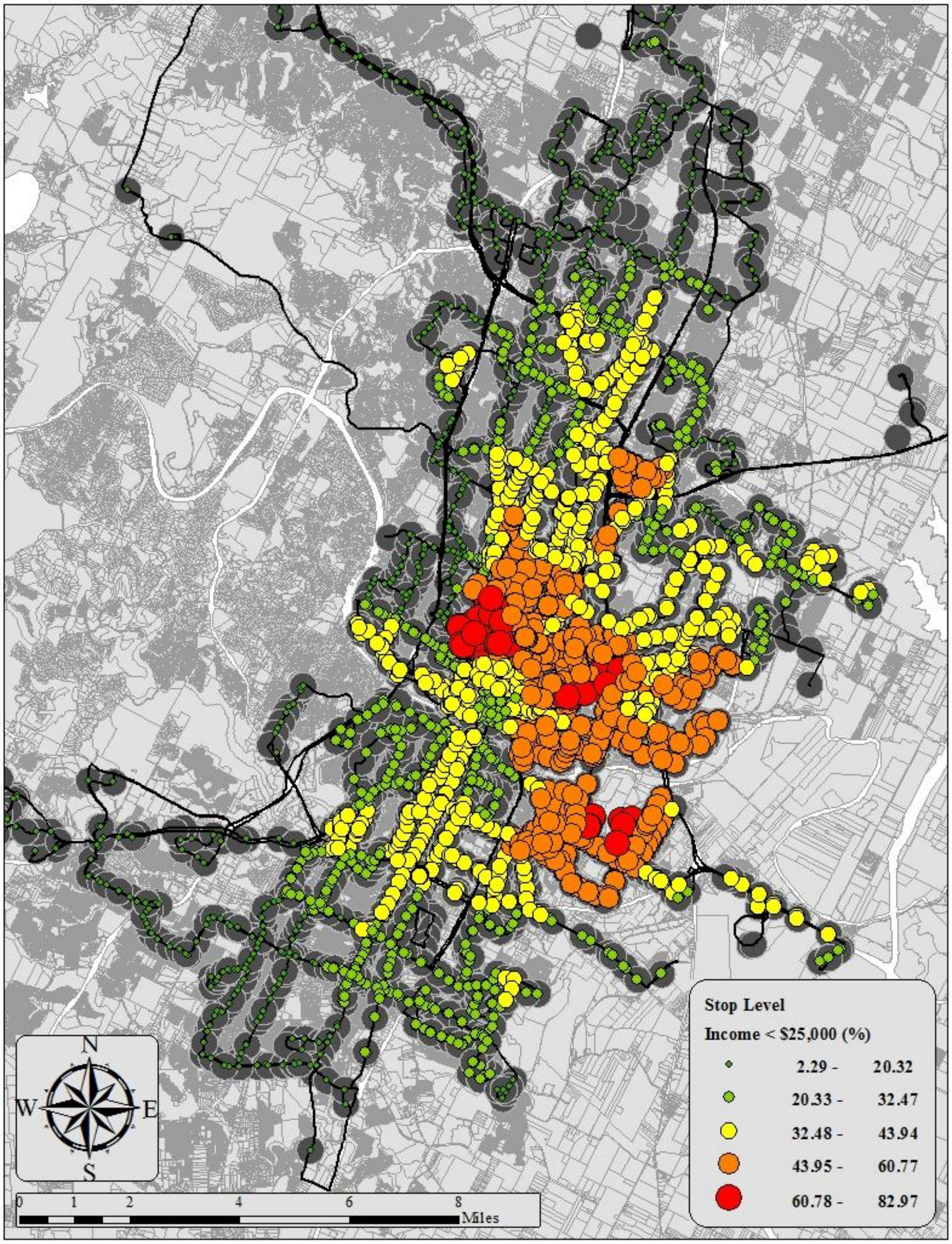

Figure 19: Stop Level Analysis Independent Variable 2 - Income Percentage

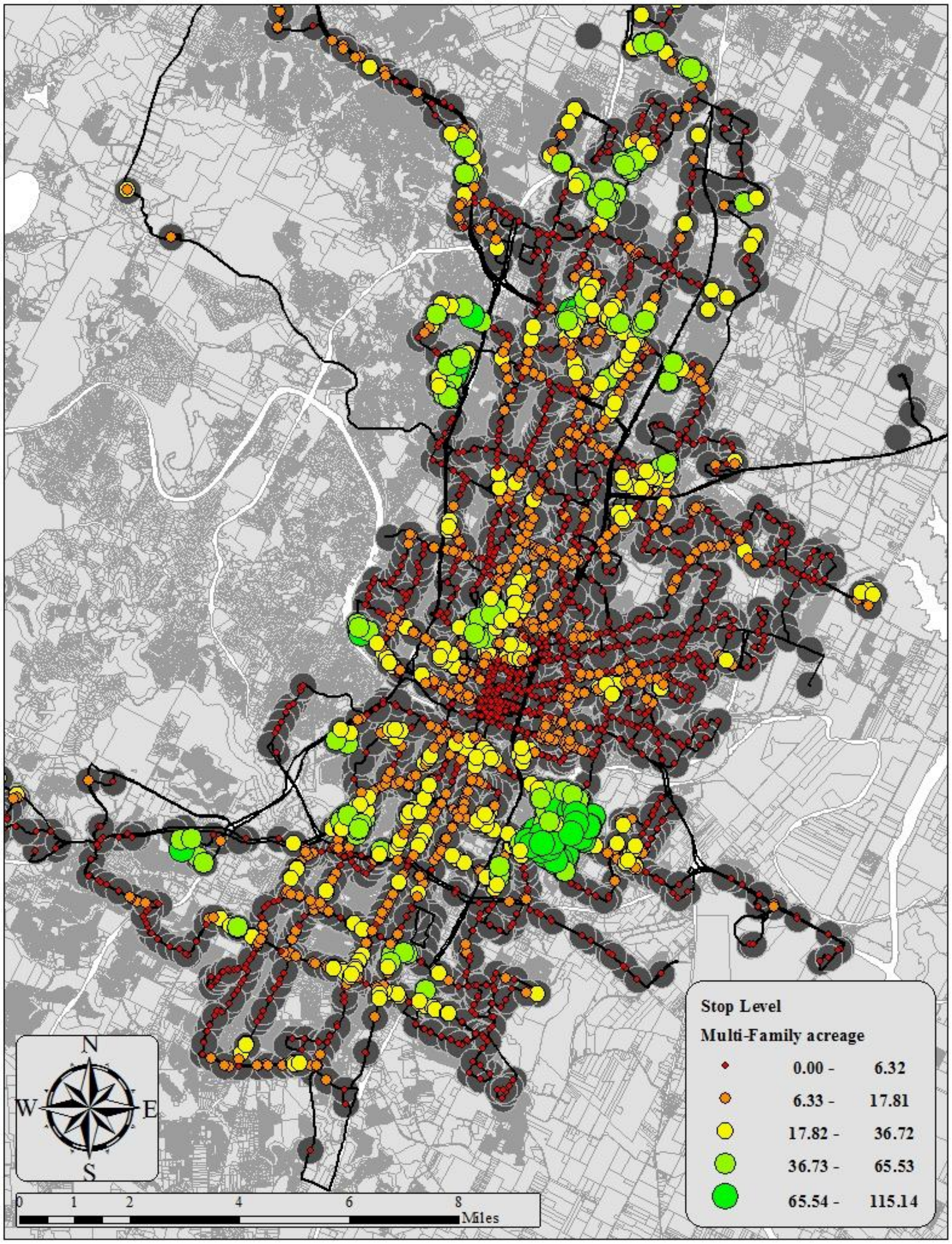

Figure 20: Stop Level Analysis Independent Variable 3 - Multi-Family Acreage

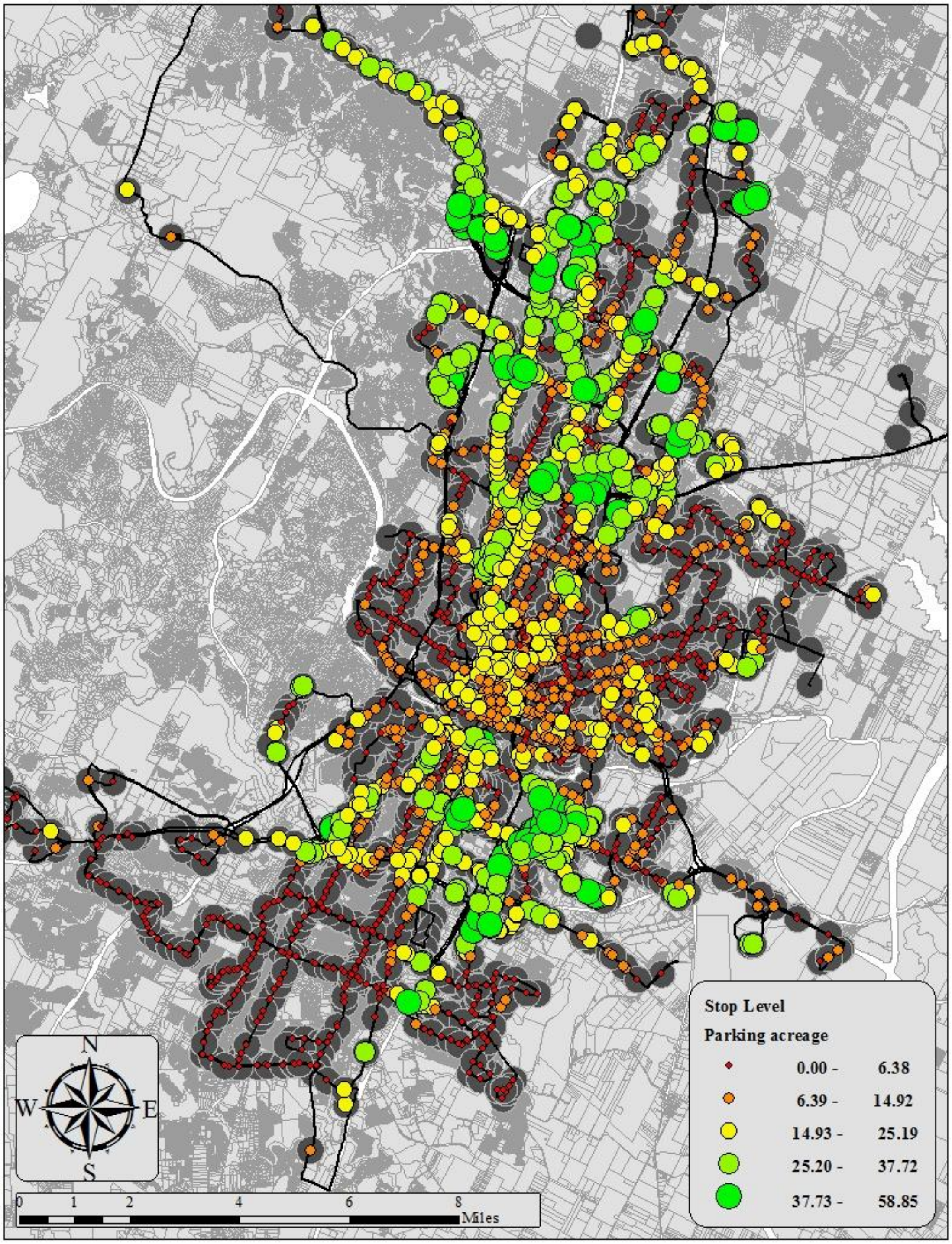

Figure 21: Stop Level Analysis Independent Variable 4 - Parking Acreage

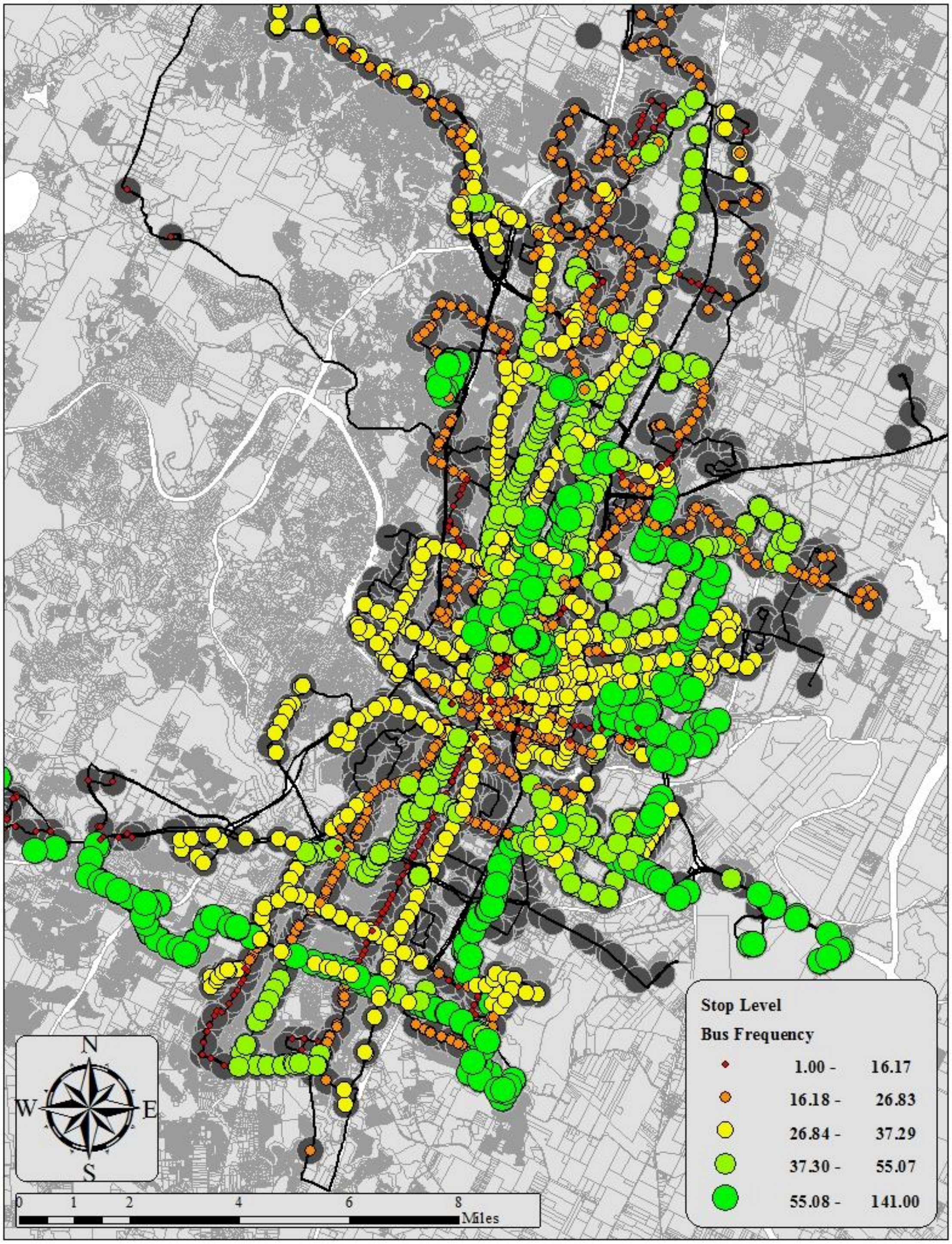

Figure 22: Stop Level Analysis Independent Variable 5 - Bus Frequency

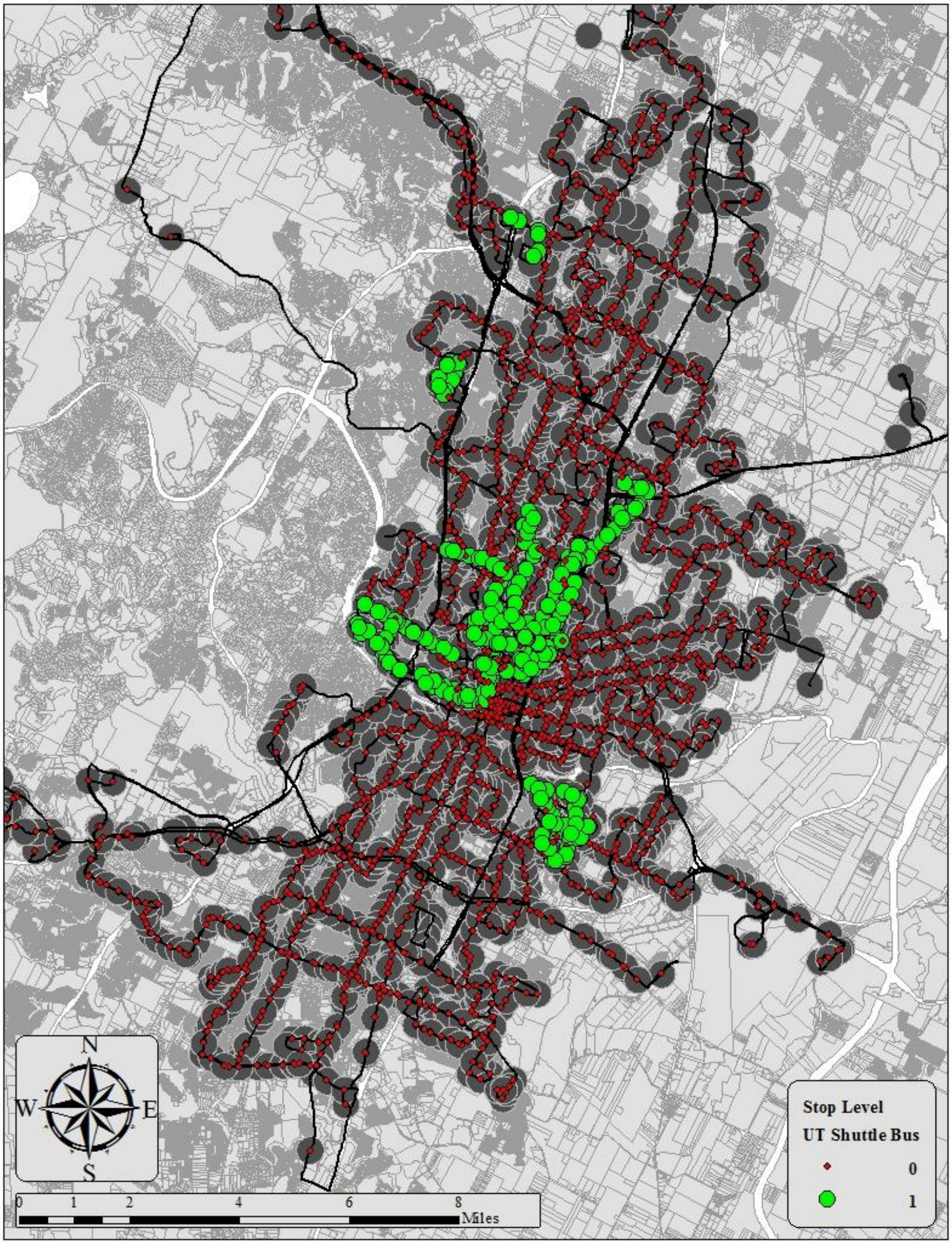

Figure 23: Stop Level Analysis Independent Variable 6 - UT Shuttle Bus

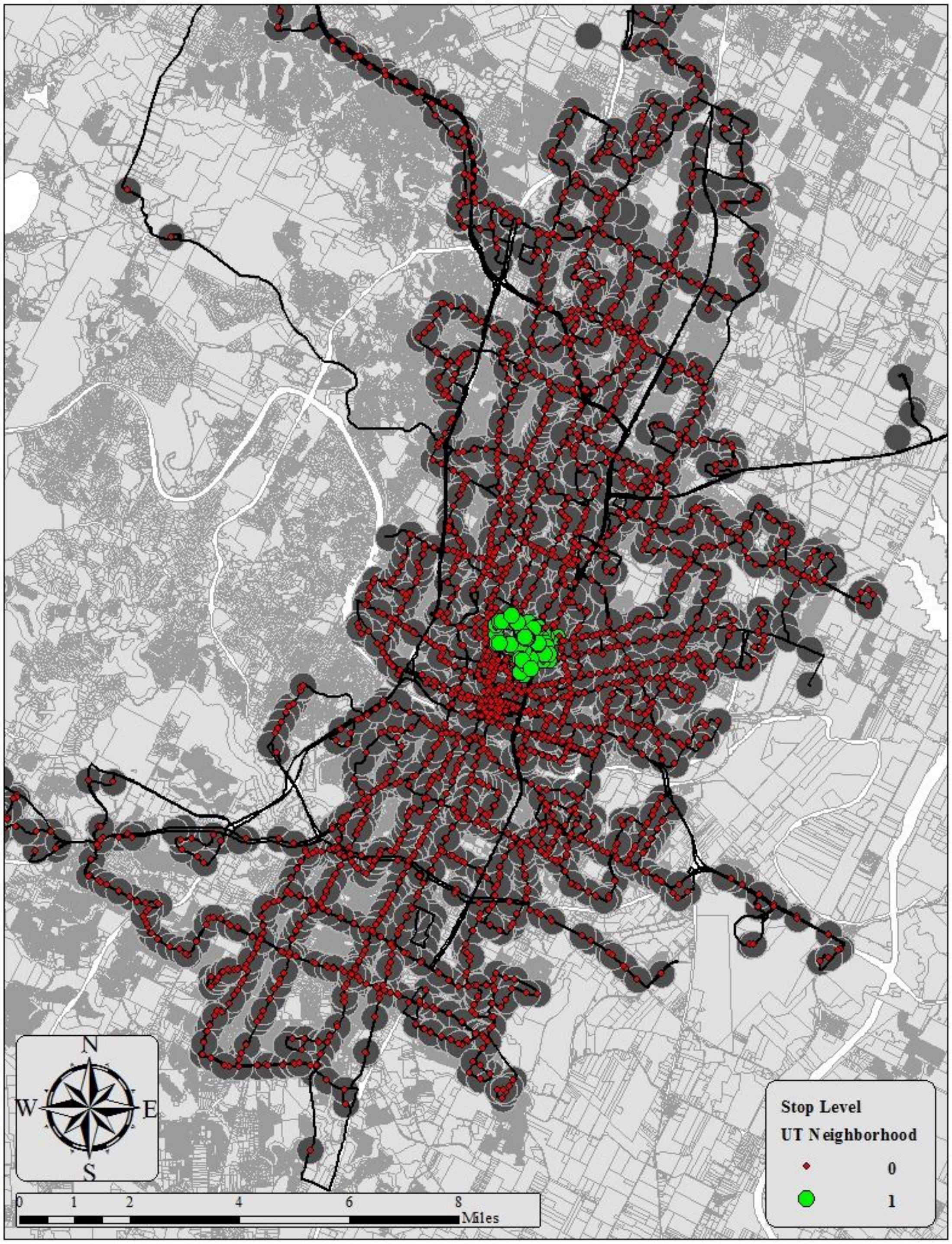

Figure 24: Stop Level Analysis Independent Variable 7 - UT Neighborhood

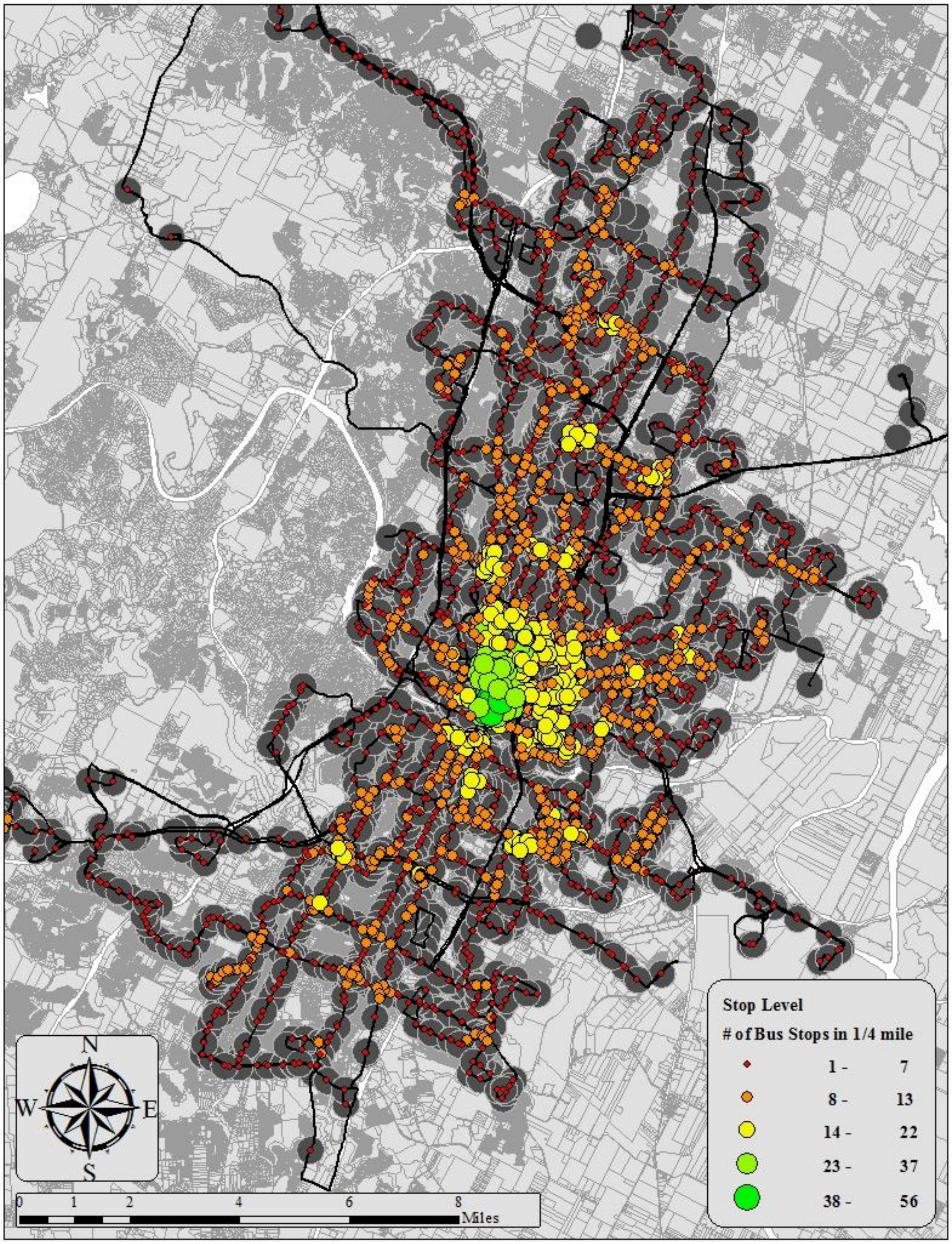

Figure 25: Stop Level Analysis Independent Variable 8 - Count

3) Stop-Level Correlation Analysis

Correlation analysis was conducted prior to regression analysis. There was no strong relationship between any pair of independent variables to be used in the regression model at the next step. Although other variables could be included and analyzed in the regression model, more reasonably explainable but still important variables were chosen between similar independent variables because the model does not permit correlated independent variables to be used at the same time. It was also necessary to compare the results with previous regression models, so that the independent variables used in the previous route/segmented route level analyses were included if there was no correlation found with new independent variables. Subsequently, 8 independent variables were used in the stop-level analysis: White_Percentage, Income_less_than_25,000_Percentage, Multi-Family_acreage, Parking_acreage, Freqeuncy_per_day, UT Shuttle Bus Dummy, UT neighborhood Dummy, and The number of Clustered Bus Stops.

Category	White_Prct	INC_LS_25000_PC	MF_acreage	Parking_acreage	Frequency_per_day	UTline_D	UT_NBHD	Count
White_Prct	1.000	-0.336	-0.072	0.003	-0.070	0.128	0.012	0.127
		<.0001	0.000	0.870	0.000	<.0001	0.511	<.0001
INC_LS_25000_PC	-0.336	1.000	0.234	0.097	0.202	0.308	0.233	0.330
	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
MF_acreage	-0.072	0.234	1.000	0.337	0.082	0.247	-0.017	-0.028
	0.000	<.0001		<.0001	<.0001	<.0001	0.359	0.137
Parking_acreage	0.003	0.097	0.337	1.000	0.030	0.120	0.004	0.040
	0.870	<.0001	<.0001		0.119	<.0001	0.845	0.032
Frequency_per_day	-0.070	0.202	0.082	0.030	1.000	0.212	0.131	-0.085
	0.000	<.0001	<.0001	0.119		<.0001	<.0001	<.0001
Headway	0.128	0.308	0.247	0.120	0.212	1.000	0.367	0.200
	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001
UT_NBHD	0.012	0.233	-0.017	0.004	0.131	0.367	1.000	0.206
	0.511	<.0001	0.359	0.845	<.0001	<.0001		<.0001
Count	0.127	0.330	-0.028	0.040	-0.085	0.200	0.206	1.000
	<.0001	<.0001	0.137	0.032	<.0001	<.0001	<.0001	

- Upper Cell: Coefficient between independent variables
- Lower Cell: Significance Level

Table 14: Stop-Level Correlation Analysis

4) Stop-Level Regression Result

The goodness of fit of the two models was almost the same, slightly in favor of the semi-log model. In the linear model, although the White_Percentage variable was not statistically significant, INC_LS_25000_PC was positive and statistically significant at the 5% level. Multi-family acreage was not statistically significant but was positive. Parking acreage had a positive coefficient of 1.75, which was significant at the 1% level. Bus frequency at each stop was also statistically significant at the 1% level (coefficient 1.04). This means one more time of bus services results in increase of 1.04 riders in the end. The results indicate that a daily increase of 1 bus departure per stop is associated with a daily increase of 1.76 passengers. UT shuttle line stops have an advantage, drawing as many as 41 passengers. Also, bus stops in the UT neighborhood had a positive influence on ridership (coefficient 638), interpreted as an additional 638 riders using a stop located within the UT neighborhood.

The relative importance of independent variables on the result was verified by the result of standardized coefficients. The UT_NBHD dummy variable was the most influential variable, followed by Count, Parking, Frequency, UT shuttle bus, and Income.

According to the result of the semi-log model, every independent variable was statistically significant and the sign of every variable was same as the hypothesis of this study. A one percent increase in Caucasian residents within the quarter mile buffer from a bus stop results in a 0.6 percent decrease in ridership. A one percent increase in low-income households is related to a 1.14 percent increase of ridership. A one acre increase in multi-family land use and parking space results in 0.92 and 2.97 percent increases in ridership, respectively. An increase of 1 daily bus departure service at each stop is responsible for a 1.76 percent growth in ridership. If a stop is located on the route of at least one of the UT shuttle buses, then it contributes to a 26.97 percent increase in bus ridership compared with 97.74 percent if it is located within the UT neighborhood. Lastly, the addition of one stop to a cluster of stops within the quarter mile buffer from each bus stop contributes to a 0.04 percent increase in ridership. These values of coefficient within the semi-log model analysis were all significant at the 1% level.

Model		Linear Model		
		Coefficient		Standardized Coefficient
	Intercept	-81.8836	**	0.0000
Traveler Variables	White_Prct	0.0389		0.0021
	INC_LS_25000_PC	0.9328	**	0.0505
Land Use Variable	MF_acrerage	0.5216		0.0307
	Parking_acreage	1.7594	***	0.0761
Service Variable	Frequency_per_day	1.0466	***	0.0650
	UTline_D	41.5934	**	0.0503
	UT_NBHD	638.0795	***	0.3736
Contextual Variable	Count	4.2681	***	0.1126
R-Square		0.2283		
Observation		2630		
***: 1% significance level, **: 5% significance level, *: 10% significance level				

Table 15: Linear Model (Stop-Level Ridership Analysis)

Model		Linear Model		
		Coefficient		Standardized Coefficient
	Intercept	1.5152	***	0.0000
Traveler Variables	White_Prct	-0.0065	***	-0.0645
	INC_LS_25000_PC	0.0114	***	0.1126
Land Use Variable	MF_acrerage	0.0092	***	0.0981
	Parking_acreage	0.0297	***	0.2347
Service Variable	Frequency_per_day	0.0176	***	0.1989
	UTline_D	0.2609	***	0.0576
	UT_NBHD	0.9774	***	0.1044
Contextual Variable	Count	0.0448	***	0.2158
R-Square		0.2779		
Observation		2630		
***: 1% significance level, **: 5% significance level, *: 10% significance level				

Table 16: Semi-Log Model (Stop-Level Ridership Analysis)

Chapter 4: Conclusion

This study analyzed bus ridership in the City of Austin area, Texas. The research could assist organizations, including the City of Austin and the Cap Metro, in informing transportation policies or economic policies, especially those targeting low-income families.

The study examined four categories of variables affecting bus ridership: Demographic characteristics, Land Use, Service levels and qualities, and Contextual factors. These variables were somewhat related to each other, so that some variables could not be simultaneously included in regression models. To sum up, at the route-level analysis, it was found that the proportion of low-income households, longer revenue hours, and reduced headway attracted riders to buses. At the segregated route-level, a lower percentage of Caucasian residents, a higher proportion of low-income households, greater multi-family acreage, longer revenue hours, shorter headway, and location on UT shuttle bus routes have a positive influence on increasing ridership. In the case of stop-level analysis, fewer Caucasian residents, a higher proportion of low-income households, greater multi-family acreage, greater parking acreage, more frequent average bus services by routes at bus stops, location on UT shuttle bus routes, location within the UT neighborhood, and additional stops clustered adjacent to the target stop all encouraged bus ridership.

These three different types of models have their own advantages and disadvantages. First, the route-level analysis clearly explains every service variable including revenue hours, headway, and a specific route dummy-variable. It is clear that a bus service with longer revenue hours attracts more riders in the route-level model. This revenue-hours variable was the most significant factor for increase in ridership in the route level analysis. Also, longer headway contributes to reduced ridership at the route level. This headway is originally defined by bus routes, so that it clearly affects the ridership analysis result. In the case of the route number dummy-variable, it suggests how many people can be affected by a certain route bus. In this study, the Route #1 bus is one of the major routes in Austin, and it clearly shows significant influence on ridership.

On the other hand, the segmented route level analysis shows weakened influence of service variables, while showing stronger influence of traveler variables and land use variables compared to the route level analysis. In this segmented route model, the ethnic origin of residents in the route catchment and household income variables were found to have significant effects on ridership. Multi-family acreage, which is one of the land use variables, was also found to be a positive factor, whereas it was not significant in the route level analysis.

Lastly, the stop-level analysis shows the strongest result of every individual variable including traveler, land use, service, and contextual variables. Especially in the semi-log model, all of the independent variables show significant results at 1% significance level. It is important to know how strongly each individual independent variable influences ridership. In this sense, the stop-level analysis clearly shows the level of influence of each independent variable within statistically significant level. Although it was difficult to measure the clustering effect of stops, the clustering effect has a positive influence on the ridership overall.

More detailed studies would require examination of the same ridership data both at each route and each stop, which was connected to the analysis of ridership characteristics. Also, in terms of data acquisition, it would be preferable to use recent data in order to make the annual data similar between dependent and independent variables, including population, employment, income, and the ethnic makeup of transit catchment areas. In addition, it might be an appropriate to conduct time-series analysis using accumulated information with five-year or ten-year term data sets. Such a time-series analysis could make it possible to explain dynamic changes in the use of buses, especially in a dynamically growing city such as Austin, thereby providing a more detailed basis for public policy development. Lastly, analysis of both supply and demand factors, as well as the level of influence on each other, can provide more valuable information for use by public sector organizations.

Bibliography

Alperovich, G., Kemp, M. A., and Goodman, K. M. (1977) "An Econometric Model of Bus Transit Demand and Supply," *The Urban Institute Working Paper* No. 5032-1-4, Washington, D.C.

Arrington, G. B. and Cervero, R. (2009) *Transit Cooperative Research Program Report*, 128, Transportation Research Board.

Cervero, R., Murakami, J., and Miller, M. (2010) "Direct Ridership Model of Bus Rapid Transit in Los Angeles County, California," *Transportation Research Board: Journal of the Transportation Research Board*, 2145: 1-7.

Gaudry, M. (1975) "An aggregate time-series analysis of urban transit demand: the Montreal case," *Transportation Research*, 9: 249-258.

Horowitz, J. L., Koppelman, F. S., and Lerman, S. R. (1986) "A Self-instructing Course in Disaggregate Mode Choice Modeling (No. Report DOT-T-93-18)," US Department of Transportation.

Huang, Z. and Zhang, M. (2011) "Accessibility-based Approach to Forecasting Bus Trip Production at the Stop Level - An Application Example from Wuhan, China," 90th Transportation Research Board Annual Meeting, Washington, D. C.

Kikuchi, S. and Miljkovic, D. (2001) "Use of Fuzzy Inference for Modeling Prediction of Transit Ridership at Individual Stops," *Transportation Research Record: Journal of the Transportation Research Board*, 1774: 25-35.

Lam, W. and Morrall, J. (1982) "Bus Passenger Walking Distances and Waiting Times: A Summer-Winter Comparison," *Transportation Quarterly*, 36(3): 407-421.

Levinson, H. S. and Brown-West, O. (1984) "Estimating Bus Ridership," *Transportation Research Record*, 994: 8-12.

Lin, J. J. and Shin, T. Y. (2008) "Does Transit-Oriented Development Affect Metro Ridership?" *Transportation Research Board: Journal of the Transportation Research Board*, 2063: 149-158.

Multisystems, Inc. (1982) *Route-Level Demand Models: A review*, Urban Mass Transportation Administration, US Department of Transportation, DOT-1-82-6.

Peng, Z. and Dueker, J. K. (1995) "Spatial data integration in route-level transit demand modeling," *Journal of the Urban and Regional Information Systems Association*, 7:26-37.

Peng, Z., Dueker, J. K., Strathman, J., and Hopper, J. (1997) "A simultaneous route-level transit patronage model: demand, supply, and inter-route relationship," *Transportation*, 24: 159-181.

Small, K. A. (1992) *Urban Transportation Economics*, Harwood Academic, Philadelphia, PA.

Small, K. A., Noland, R., Chu, X., and Lewis, D. (1999) "Valuation of travel-time savings and predictability in congested conditions for highway user-cost estimation," *National Cooperative Highway Research Program Report*, 431.

Taylor, B. D., Mueller, D., Iseki, H., and Fink, C. (2008) Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. *Transportation Research Part A.*

TranSystems, Planners Collaborative, Inc., and Tom Crikelair Associates (2009) *Elements Needed to Create High-Ridership Transit Systems: Interim Guidebook*, Transportation Research Board.

Tri-Met (1989) Tri-Met Service Standards. Portland, Oregon.

Wardman, M. (2001) "Review of British evidence on time and service quality valuations. *Transportation Research Part E: Logistic and Transportation Review*, (37) 107-128.

Zhao, F., Chow, L. F., Li, M. T., and Liu, X. (2005) *A transit ridership model Based on geographically weighted regression and service quality variables*, Lehman Center for Transportation Research, Florida International University.

Zhao, F., Chow, L. F., Li, M. T., Ubaka, I., and Gan, A. (2003) Forecasting Transit Walk Accessibility: Regression Alternative to Buffer method, *Transportation Research Record*, 1835: 34-41.