Assessment of Mobility and Transit Access to Captive Riders in Suburban and Rural Areas

DRAFT FINAL REPORT

Prepared By:

Sabya Mishra, Ph.D., P.E.
Assistant Professor, Dept. of Civil Engineering, University of Memphis, 112D Engr. Sc. Bldg., 3815 Central Avenue, Memphis, TN 38152, Tel: 901-678-5043, Fax: 901-678-3026, Email: smishra3@memphis.edu

Christopher R. Cherry, Ph.D., P.E.
Associate Professor, Department of Civil and Environmental Engineering, The University of Tennessee – Knoxville, 321 John D. Tickle Building, Knoxville, TN 37996, Phone: 865-974-7710, Email: cherry@utk.edu

Mihalis M. Golias, Ph.D. A.M. ASCE
Associate Professor, Dept. of Civil Engineering, Associate Director for Research Intermodal Freight Transportation Institute, University of Memphis, 104C Engr. Sc. Bldg., 3815 Central Avenue, Memphis, TN 38152, Tel: 901-678-3048, Fax: 901-678-3026, Email: mgkolias@memphis.edu

Zohora Sultana
Research Assistant, Dept. of Civil Engineering, University of Memphis, 302A Engineering Admin. Bldg., 3815 Central Ave Memphis, TN 38152, USA, Phone: Email: zsultana@memphis.edu

Samuel Jordan
Research Assistant, Dept. of Civil Engineering, University of Memphis, 302A Engineering Admin. Bldg., 3815 Central Ave Memphis, TN 38152, USA, Email: swjordan@memphis.edu

Saman Tabrizizadeh
Research Assistant, Department of Civil and Environmental Engineering, The University of Tennessee – Knoxville, TN 37996, Email: stabrizi@vols.utk.edu

Technical Report Documentation Page

1. Report No. FHWA-CFL/TD-0x-00x	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle *Assessment of Mobility and Transit Access to Captive Riders in Suburban and Rural Areas*		5. Report Date March 2018 6. Performing Organization Code
7. Author(s) Mishra, Sabya; Cherry, Christopher; Golias, Mihalis; Sultana, Zohora; Jordan, Samuel; and Tabrizizadeh, Saman.		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Memphis 3815 Central Avenue Memphis, TN 38152		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Tennessee Department of Transportation James K. Polk Building, Suite 900 505 Deaderick St., Nashville, TN 37243-0344		13. Type of Report and Period Covered 14. Sponsoring Agency Code HFTS-16.4

15. Supplementary Notes

COTR: First & Last name, FHWA CFLHD; Advisory Panel Members: List First & last names. This project was funded under the FHWA Federal Lands Highway Technology Deployment Initiatives and Partnership Program (TDIPP). Or the Coordinated Technology Implementation Program (CTIP).

16. Abstract

This project addresses mobility and transit access in urban and rural areas. For urban areas, we identify the challenge to the provision of equitable access is quantifying equity-oriented metrics for public transit service, and utilize an approach that relies on available transit network characteristics and spatial demographic data to quantify multimodal transit connectivity and equity. This method is illustrated by analyzing transit connectivity for three metropolitan cities in the state of Tennessee in the United States and overlaying that connectivity on demographic data. Results indicate that the distribution of transit services among vulnerable populations varies within and between cities. For rural area, we analyze demand response transit (DRT) services are provided to meet transit needs. However, state public agencies face the dilemma of whether to increase or decrease the service availability for on-call services. To enhance decision making of identifying what the causal factors related to DRT trips, the authors present a set of econometric models by integrating a sample DRT data with other explanatory variables such as land use, socio-economic, and demographic characteristics. We developed various econometric models to understand the factors that affect DRT trips. A number of factors showed significant relationship with DRT trip frequency including distance, population density, elderly population, average income, and others. Further, the elasticity effects of these different factors were computed to quantify the magnitude of their impact on DRT. The proposed model can be helpful for transit agencies to predict the frequency of DRT trips and to provide adequate services in rural areas.

17. Key Words	18. Distribution Statement
CONNECTIVITY, EQUITY, DEMAND RESPONSE TRIPS, ECONOMETRIC MODELS	No restriction. This document is available to the public from the sponsoring agency at the website http://www.cflhd.gov.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	104	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

DISCLAIMER

This research was funded through the State Research and Planning (SPR) Program by the Tennessee Department of Transportation and the Federal Highway Administration under RES #: 2016-16, Research Project Title: Assessment of Mobility and Transit Access to Captive Riders in Suburban and Rural Areas .

This document is disseminated under the sponsorship of the Tennessee Department of Transportation and the United States Department of Transportation in the interest of information exchange. The State of Tennessee and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the views of the author(s) who are solely responsible for the facts and accuracy of the material presented. The contents do not necessarily reflect the official views of the Tennessee Department of Transportation or the United States Department of Transportation.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Public Transit is a critical component of Tennessee Department of Transport's (TDOT) Long-Range Transportation Plan. The demographic data and trends in the state of Tennessee point to a potential increase in need for public transit services in cities and rural communities. The role of TDOT in providing the mobility and accessibility options to the residents, especially captive riders, is critical for the future quality of life and economic competitiveness of Tennessee. Also, the trend of increased percentages of household income spent on transportation and increased commuting distances are going to be major contributing factors behind increased transit demand in the future. This study aims to identify potential demand for transit services, in urban and rural areas, in the context of "captive ridership" and develop a methodology that will assist TDOT to monitor such transit needs and examine the provision of transit services in a cost-effective manner. This research will be crucial in identifying areas in needs of service, developing a methodology to address the accessibility and mobility issues and formulating a cost-effective plan to provide transit services. The results can be potentially tied to the Tennessee statewide mobility report.

First, transit connectivity of urban areas is determined. The three major cities considered are Knoxville, Memphis, and Nashville. Transit connectivity is a multidimensional problem involving various service quality factors that include walking distance, in-vehicle travel time, waiting time, number of destinations served and number of transfers to reach destinations. Further adding to this complexity is (usually) the high number of available routes with distinct characteristics within a network. Based on network graph properties, this paper proposes connectivity indices at stop, route, and zonal level by considering various factors such as speed, frequency, operational capacity, fare, route origins and destination, and urban form characteristics that serves the transit system. The connectivity indices are applied to three metropolitan cities (Knoxville, Memphis, and Nashville) of Tennessee by using open access Generalized Transit Feed Service (GTFS) data. The models and data processes developed in this paper can be used to (i) determine the performance transit system with no additional data purchase, (ii) use of transit performance measures along with other data sources (such as vehicle ownership, income etc.) to assess future service needs, (iii) use of geographic information systems capabilities to disseminate transit performance measures for potential future users, and to further induce demand, and (iv) seamless re-estimation of transit performance measures both in alternate dimensions of time and space. Further, the transit connectivity measure is used to determine equity by various socio-economic factors such as household income, vehicle ownership, employment, and population. Transit connectivity equity is estimated by the GINI index. All three cities have both strengths and weaknesses in serving the captive riders when various socioeconomic factors are considered. For instance, Knoxville provides more equitable transit service when household income is considered, while Memphis based on population, and Nashville based on vehicle ownership.

Second, mobility and transit access of rural areas are determined using demand response transit (DRT) service data using count data models. A hybrid dataset was prepared that include DTR trips at zip-code level along with various socio-economic and demographic data. A set of count models were developed including Poisson, negative binomial, zero inflated Poisson, zero inflated negative binomial, hurdle Poisson, hurdle negative binomial, and zero-inflated negative binomial mixed effect. DRT trip forecasting model is validated with 20% of the data not used for model training. Further, elasticity of variables is determined to assess the importance of various factors' influence on rural transit and mobility needs.

CHAPTER 1: INTRODUCTION

Past studies have shown that the needs for transit services among vulnerable groups are higher than average (Grengs, 2002). Income is directly related to such service needs, as lower household income indicates higher potential to limited or zero car ownership, which in turn, suggests higher need for transit (captive demand). In this context, federal and state agencies are urged to focus on providing access to transit resources for such individuals, which frequently belong to vulnerable population groups (e.g., minority or low-income citizens). Major public and private facilities that provide personal and professional services are commonly located within urban boundaries. Employment centers, educational institutions, medical facilities, as well as retail and entertainment venues are some examples. Captive riders have fewer travel choices and increased transportation barriers to such locations and the challenge is to determine: (i) the number of people in a given geographic area likely to require passenger transportation services as well as (ii) the number of trips likely to be made by those persons if they had minimal limitations on their personal mobility.

1.1 Urban Transit

Many measures of transit service and accessibility have been put forth in the literature, but few offer a metric to measure the quality of service and performance of a large multi-modal regional transit system. The literature that does purport to offer such insight requires significant amounts of data not only about the transit system, but also of the complete demographics of the service area (Modarres, 2003). Other methods require a full transportation demand and transit assignment models, tools that are prohibitively expensive for many localities (Lam & Schuler, 1982).

Measuring transit system performance and the level of service at many different levels is vital to funding decisions (Dajani & Gilbert, 1978). Agencies with the objective to improve the transit system using external funds must make the case that the project will be a worthwhile improvement to the system. At the same time, agencies interested in investigating the potential effect of removing a stop, group of stops or transit line from service must know the potential effect it will have on the performance of the system. In the absence of complex transportation demand models, this information is nearly impossible to obtain (Baughan et al. 2009). A methodology that reduces the need for large amount of data, yet provides essential information on system performance is critical to the decision-making process. A simple, yet comprehensive, measure will be the determination of transit connectivity index. Beyond Transit Oriented Development (TOD) style plans, the connectivity index provides a way for planners to measure passenger acceptance rates and accessibility for a single node based on its access within an entire multi-modal regional transportation network.

A transit network represents complex interactions of nodes (stops), and links (routes) with unique characteristics serving various origins and destinations. Headway, frequency, speed, and capacity are critical terms that define the characteristics of a stop or transit route. The evaluation of transit supply and demand requires a systematic representation of all the network elements (e.g., stop, route features) and service level (e.g., headway, capacity, fare) characteristics. A number of performance measures are available in the literature including degree centrality, eigenvector centrality, closeness centrality, betweenness centrality, etc. However, such measures only consider network level characteristics and often ignore service level characteristics. This study proposes a transit connectivity measure that captures basic graph theoretic properties of the network and in addition uses a connecting power of individual stops serving each route in the system. Further connectivity measure is proposed at stop, route and zone level to provide a both micro and macro level performance of the transit system. Equity of transit connectivity is analyzed by GINI index. Equity of transit connectivity is determined using various socioeconomic data including household income, vehicle ownership, employment, and population groups. Both the transit connectivity and equity measures were computed using open access General Transit Feed Specification (GTFS) data, and census data. Transit connectivity and equity of three metropolitan cities (Knoxville, Memphis and Nashville) in Tennessee (TN) are analyzed and the findings are discussed.

1.2 Sub-urban and rural Transit

In rural areas, captive riders need to travel towards (sub)urban areas for financial, health, shopping and other needs. Such travel needs are crucial and unavoidable in most cases. This is because sub(urban) areas contain major public and private facilities for personal and professional services. With greater transit needs and fewer travel choices per capita, public transit is an important mode of transportation for rural residents who do not own or operate a car, albeit they do not have immediate access to private transportation or they are bound to use public transportation in order to meet their travel needs. In rural areas, travel demand density is lower and more dispersed, diminishing the effectiveness of traditional forms of fixed route bus-based public transport systems. Because of the low population density and dispersed origins and destinations, rural transit services usually have a very low fare box recovery rate, which results in abandonment of fixed route public transports after short period of operation. Alternatively, demand response transit (DRT) systems in rural areas can be more cost-effective by reducing frequencies and providing smaller vehicles. DRT service can adapt the changes in demand by either shifting its timetable and/or altering its route. The fare charged is very low or free depending on passenger socioeconomic characteristics and the route being served. In many places, DRT remains an effective service which may only be available for specific groups of users like the elderly and/or mobility impaired. However, certainly there are other user groups who need DRT for non-discretionary trips.

Planning a new service demands the assessment of the ‘user needs’. Davison et al.(2014) showed DRT as the most cost effective way of ensuring the transit of rural communities without a conventional bus service. Enoch et al. (2006) evaluated DRT service in the rural area, but did not focus on demand distribution. This indicates a need of study to identify what destinations are essential for rural residents, as well as how the frequently rural residents will access these services.

In this research effort, the objectives are to identify transit demand in rural areas, exploring socio-economic and demographics patterns on DRT, and develop a method to assist state Departments of Transportation (DOTs) and transit providers to identify where transit connections and investments should be made. In this context, the authors have assembled a comprehensive dataset for analyzing DRT trip frequency, and developed count models to explore the effects of potential factors on DRT trip frequency.

The methodology is based on census data and data collected from travel agencies providing demand-response service in rural area of Tennessee. The census data record demographic characteristics, household attributes, etc. The travel data record comprises of trip attributes such as origin and destination region, ZIP Code, County, and trip purpose. Details of these travel diaries along with the demographic characteristics such as age, gender, relationship in household, vehicle ownership, employment type, household size and structure, and household income are available for predicting the travel demand patterns for new modes of transportation.

1.3 Objectives and Report Organization

The primary objectives of this project are to:

- provide an extensive review of current and past literature from other public and private sources on providing transit services to sub-urban and rural areas
- develop a methodology to address the accessibility and mobility issues in rural and suburban areas
- determine the urban locations (or destinations such as schools, hospitals, shopping malls, etc.) which are essential for rural residents and are visited frequently
- formulate a cost-effective plan to provide transit services to the suburban and rural residents by coordinating the regional transit services
- develop a needs assessment matrix that TDOT can utilize for long-term planning and to provide desired level of accessibility and mobility to the captive riders.

The rest of the report is organized as follows. Chapter 2 presents the literature review urban transit connectivity and rural transit mobility access. Chapter 3 discusses data requirement found in the literature and briefly introduces the reader to the data sources used in this research. Chapter 4 presents methodology and analyses result of urban area transit connectivity. Chapter 5 describes methodology for modeling DRT trips. Chapter 6 concludes the report with summary of findings and proposes scope of future research.

CHAPTER 2: LITERATURE REVIEW

The relevant studies on accessibility and mobility issues and captive ridership in urban, suburban, and rural areas are discussed in this chapter.

2.1 Connectivity in Urban Transit Networks

Measures of transit connectivity are significant in transportation planning, in the context of overall system performance, as well as for distinct system parts evaluation. Such measures provide a consistent basis for rationalizing public spending through identifying the critical (in terms of connectivity and general mobility) parts of the transit network, and for developing service strategies (Sarker, Mishra, Welch, Golias, & Torrens, 2015). From a different perspective, an agency might be interested in evaluating the potential effect on the overall performance of the transit system, when removing a single or multiple stops or even a whole transit line from a network and so forth. Thus, measuring system performance at various levels (e.g., node, link or line) is vital for supporting such decisions. TDOT has been strongly motivated to improve the quality of life throughout the State, by improving connectivity of its transit system and the mobility options, aiming to provide viable alternatives to the single occupancy vehicles (SOV). In this direction, TDOT's Transportation Demand Management (TDM) Programs aim to "increase travel by alternative modes and at alternative times to reduce total trips, reduce congestion, and decrease the use of single-occupant modes" (Tennessee Department of Transportation (TDOT), 2015). In the following paragraphs of this section, we present pertinent literature that is related to service provision to urban, suburban, and rural areas as well as methods for evaluating connectivity of transit networks in the context of providing the decision maker with investment decision tools.

2.1.1 Critical Transit Network Elements Assessment

The authors have reviewed literature on accessibility and mobility issues and captive ridership in urban, suburban, and rural areas. The review includes published journal articles, past research in Tennessee, as well as in other states in order to decide which data is the most appropriate for the specific study needs and to provide a methodological framework that will allow transit agency officials and decision makers to easily identify critical investment locations. As past research indicated, destinations of interest are well connected while origins in certain cases are not. The cornerstone of this study has been to combine transit network connectivity performance measures (of various levels like node, link, etc.) with socioeconomic data (such as household income, car ownership, etc.) in order to locate critical areas with low connectivity and increased captive ridership, that may serve as potential generators of transit demand. Past literature on transit network performance can be grouped into two main categories: (i) rural/intercity transit related studies and (ii) urban transit related studies. A concise review of past work is presented for both categories in the following paragraphs.

Yang and Cherry (2012), examined the characteristics of intercity bus riders (trips between 30 and 170 miles) and proposed methods for service gap identification and network investment (expansion) prioritization. The study concluded that while bus stations are well connected to destinations, they are poorly connected to demand locations. Yang (2013), studied the rural transit rider characteristics and proposed an Intercity Bus (ICB) system evaluation method as well as route design directions for Deviated Fixed Route Transit (DFRT) services for the State of Tennessee. The study found that DFRT and DRT passengers are likely to be female, of minority races, of low personal/household income and low or zero car ownership, etc. The author also proposed a methodology for locating high ICB demand areas and for ICB proper stop design, suggesting that stops should be located in areas with high population density (since those were identified as high demand areas). The study also proposed a methodology for cost effective/optimum design rural DFRT network. Yang et al.(2016), presented a methodology to model trip generation and proposed an optimal (DFRT) route(Yang et al., 2016). To fit the data, they developed a zero-inflated negative binomial regression model that was later used to estimate trip generations in other parts of the state. They then proposed a methodology to identify all possible DFRT routes and selected the optimum ones by generating an operating cost per passenger dataset for routes of different length in order to cost-effective routes. The modeling framework was applied throughout the state of Tennessee.

In 2010, Park and Kang (2010) developed a quantitative model for multimodal urban transit network connectivity evaluation. Selecting length, speed, and capacity as measures of a transit line's efficiency, they then defined its connecting power as the product of those measures for Centrality and Connectivity are available in the literature (Ahmed et al., 2005; Bell, Atkinson, & Carlson, 1999; Bonacich, 2007; Bonacich & Lloyd, 2001; Carrington, Scott, & Wasserman, 2005; Estrada & Rodriguez-Velazquez, 2005; Freeman, 1978; Garroway, Bowman, Carr, & Wilson, 2008; Guimera, Mossa, Turtschi, & Amaral, 2005; Junker, Koschützki, & Schreiber, 2006; Liu, Bollen, Nelson, & Van de Sompel, 2005; Martınez, Dimitriadis, Rubia, Gómez, & De La Fuente, 2003; Moore, Eng, & Daniel, 2003; Newman, 2004; Ruhnau, 2000) . Degree of centrality, as defined for social networks, was then appropriately modified to suit transit networks. The study finally derived connectivity indices for transit stops and from them, a connectivity index of the transit line under study as well as of the area of a multimodal transit network. This approach was the first to correlate transit characteristics with a connectivity index (Park & Gang, 2010). Welch and Mishra (2013) proposed a methodology to quantify and evaluate transit equity (i.e., provision of transit services in the fairest and least possible discriminatory manner and according to the relevant directives of 1964 Civil Rights Act). The proposed estimates were designed to offer a "before & after" evaluation of equity, as in cases where an agency is interested in making changes to the transit system. They further introduced an Inequality Index, as a measure of the distribution of transit services quality among the population. The proposed methodology could be useful for transit and relevant transportation agencies, to measure the distribution of transit services among specific population groups in order to provide better access to groups that need transit services the most (i.e., captive riders). From a graph theoretical approach, transit networks have different characteristics than road networks. While a link in a road network is a physical segment that connects adjacent nodes, a link in a multi-modal network is a part of a transit line that serves a sequence of transit stops (nodes) and since a stop can be served by multiple transit lines, multiple transit links (Welch & Mishra, 2013) may cross each stop. Mishra, Welch, and Jha (2012), proposed a graph theoretical approach to evaluate connectivity and assess and prioritize potential locations for transit service funding (Mishra, Welch, & Jha, 2012). The proposed approach aimed to determine the performance of large scale, multi-modal, transit networks and suggested a methodology that formulates connectivity indices as evaluation measures for nodal, line, transfer center, and regional level. Those indices incorporate unique transit line qualities (as opposed to the connectivity indices derived for social networks) and accessibility measures in order to provide a more solid evaluation for transit systems. In a similar concept publication, Mishra et al., (2015) presented a visualization tool for stops, routes and transfer zones connectivity of a multimodal transit network (Mishra et al., 2015). This approach was extended by Sarker et al. (2015) to evaluate transit connectivity through a unique connectivity index, using GTFS data. Table 1 presents formulations on centrality measures in social networks and transportation found in the relevant published literature.

Chakraborty and Mishra, (2013) indicated that land use, socioeconomic variables, and transit ridership are strongly connected. They found that land use type, transit accessibility, income, and density are strongly significant predictors of transit ridership even if their coefficients may vary across urban, suburban, and rural areas. In their study, they developed a framework to assist decision making at higher planning (i.e., State level), which, according to the authors, is the only option in capturing and eliminating possible interdependencies, due to local interests and biases. The key measure of the approach is transit ridership, which, under different agency choices/scenarios, is then projected in future. Using two scenarios (Business as usual and high energy prices) the study shows how a state agency can consider multiple choices in making decisions.

The TCRP Report 161 (Vanasse Hangen Brustlin et al., 2013), provides planners with a methodological framework for assessing the need for public transit services within a geographic area, as well as the potential annual demand of such a transit service. The methods described in TCRP 161 (Vanasse Hangen Brustlin et al., 2013) are applicable for rural counties assessment (and in cases where the area under study is not currently served by passenger transportation), and are not intended for specifying the needs and potential demand of individual routes or neighborhoods. In 2015 and from a similar perspective RSG prepared for Federal Transit Administration (FTA) a software package called STOPS (i.e., Simplified Trips-On-Project Software). STOPS is a series of programs developed to estimate transit ridership through a set of estimation procedures which skip the process of time-demanding and complex Regional Travel Demand Forecasting Modeling. While similar to regional models, STOPS is much simpler as it estimates total origin-to destination data from census data rather than trip generation and destination modeling procedures, does not require detailed transit network development in the planning environment as transit levels of service are derived form timetable information, the model is self-calibrated to represent current conditions (RSG, 2015).

2.1.2 Transportation Equity

An equitable transit system can cater to the needs of captive riders and maximize transit service coverage and all federal agencies must distribute federal resources equitably in such a way as to provide services in the fairest and least discriminatory manner. Typically, two definitions of equity are in use: vertical and horizontal. In the context of transit vertical equity (perhaps the broadest definition), indicates that those paying the most should receive the most benefit. On the other hand, horizontal equity is concerned with the equal treatment of those with equal means. Equity is difficult to quantify in many transportation applications, but emerging methods aim to include equity explicitly (and qualitatively) in the transportation planning process (Bills, Sall, & Walker, 2012; Joshi & Lambert, 2007).

2.2 Demand Responsive Transit Service in Rural Areas

Over the decades, DRT has developed as one of the most effective methods to provide transportation services to captive riders in rural areas. Many relevant studies in this area examine the effectiveness of DRT and explore the social and economic factors affecting transit trips. Bakker (1999) explained paratransit (DRT) as a “transportation option that falls between private car and conventional public bus services. It is usually considered to be an option only for less developed countries and for niches like elderly and disabled people”. Ambrosino et al. (2004) described DRT as an “intermediate form of transport, somewhere between the bus and taxi, which covers a wide range of transport services, ranging from less formal community transport through to area-wide service networks”. Wang and Winter (2010) showed that DRT has the potential to solve the challenges of the public transportation in low density urban areas. Braun and Winter (2009) have demonstrated that the collaborative transport can effectively solve classical transport planning problems in real-time. Ad-hoc DRT does not have pre-defined schedules and flexible routes but provides point-to-point transportation by reacting on demand in real-time. The fare of DRT is usually very low compared to taxis as it offers shared forms of transport and in some cases government subsidized costs. On the other hand, DRT has a long list of failure cases around the world. Enoch et al.(2006) listed several cases of DRT failed projects along with lessons that each provides. Their findings on the cause of DRT projects failure is that DRT often were not realistically designed with a full understanding of the demand of serving area and proper future plan. In many places, DRT cost is subsided by government considering this as a service for captive and low-income travelers and performance metrics should focus on the effectiveness as a social service.

Paratransit microsimulation patron accessibility analysis tool has been developed by LaMondia and Bhat (2010) by combining paratransit trip data with census data to explore variables associated with paratransit trips in Brownsville, Texas. From the analyses of the data, the authors revealed that paratransit trips are higher in census block groups with larger population, older populations, larger households, and close proximity to fixed route transit**.** TCRP Report 161 (2013) developed a model for forecasting transit demand for general public , non-program related services based on 2009 rural National Transit Database (NTD). This model proposed that the demand can be forecasted based on the size of the demographic groups such as the older adults, people with disabilities, and people without access to a vehicle because they are the dominant riders of these services according to 2009 rural NTD data. The demand model with estimated coefficients is as follows:

Non-program Demand (trips per year) = (2.20 × Population Age 60+) + (5.21 × Mobility Limited Population age 18-64) + (1.52 × Residents of Household Having No Vehicle)

TCRP 161 (2013) also recommended demand model for program related trips. These trips are only produced with the existence of a specific social-service program or activity. The developed models for program trips is given bellow:

Number of Program Participants × Program Events per Week × the Proportion of Program Participants who attend the Program on an Average Day × the Proportion of Program Participants that are Transit Dependent or Likely to Use the Transit Service provided/funded by the Agency × the Number of Weeks per Year the Program is Offered × 2 (trips per participant per event)

TCRP 161 (2013) also developed models for estimating demand for small city fixed-route service, where population is less than 50,000 and commuter transit from a rural area to an urban center. The main significant factors were revenue hours of service provided, population of service area, and college/university enrollment. In addition, the number of workers commuting, the commute distance, and if the urban place is a state capital have been used to estimate commuter trips of the proportion of workers using transit. Simple regression models have been developed for estimating ridership based on service characteristics of DRT service providers and demographic characteristics for rural demand response transit service (Mattson, 2017). He explored potential service characteristics as geographic coverage, span of services, fares, reservation requirements, and demographic characteristics as percentage of the population comprised of older adults or people without access to a vehicle etc. Multilevel models were developed to examine the effects of DRT supply-oriented factors and socio-economic attributes to estimate the demand for DRT services. The models predict that DRT users are higher in areas with higher levels of poverty, lower car ownership, lower population density, lower proportion of people working from home (C. Wang, Quddus, Enoch, Ryley, & Davison, 2014). Lerman et al. (1980) identified that vehicle ownership is negatively associated with service coverage of DRT. From a study of DRT services in Belgium, it was found that female, retired, homebound persons, and students are dominant users of DRT (Mageean & Nelson, 2003). A report from Active Age (solution for an ageing society) (2008) showed that DRT reduces the dependency on private vehicles and can be used to support mobility of disabled riders. Female and retired persons are identified as more than 50% of the users of DRT services from another study of DRT services in Tyne and Wear in the UK (Nelson & Phonphitakchai, 2012). Yang and Cherry (2017) studied the rural transit rider characteristics of Deviated Fixed Route Transit (DFRT) and DRT services for the State of Tennessee. The study found that DFRT and DRT passengers are likely to be female, of minority races, of low personal/household income and low or zero car ownership, etc. The Telebus Mobility and Access Benefits Project was done by Maddern and Jenner (2007) in Melbourne, Australia and they revealed that people aged 15-24 years and over 55 years were 74% of the DRT Telebus users and 31% passengers used the service for shopping purposes. They also conclude that 78% of passengers had no driving license and 74% of users are female.

A DRT demand model was developed by Nguyen-Hoang and Yeung (2010) at the national level in the U.S. They identified that disabled and elderly people are positively related to the unlinked passenger trips but poor households decrease the demand for unlinked passenger trips. Methodologies for paratransit service demand and a new tool for forecasting demand for transportation-disadvantaged services has been developed by Goodwill & Joslin (2013). Paratransit service demand can be estimated based on data from 2009 National Household Travel Survey by calculating trips rate for households without access to a vehicle. People with disabilities, older adults, children defined as “high-risk” or “at-risk,” and low-income persons are defined as transportation-disadvantaged (TD) population according to this study. Kattiyapornpong and Miller (2006) revealed that passenger’s travel decisions were significantly influenced by the potential demographic characteristics such as income, age, and life cycle . They also found that the travelers aged 20 to 24 years are different in their travel behavior, and the short trip planning behavior is closely associated with the income and life stage interaction, and the income and age interaction. In addition, Piatkowski and Marshal (2015) and Jain et al.(2017) point out that various socio-economic characteristics and trip characteristics of travelers affect travel behavior. TCRP (1995) and TCRP (2004) studies found the elderly, mobility limited, and those on low incomes as potential markets for DRT in rural areas. Enoch et al.(2006) found target markets for DRT: people who cannot access public transport, people without personal transport, unemployed people, single pension households, individuals with a limiting long-term illness, ethnic minority households, and people aged 14-19 years. The various demographic characteristics of the population and trip characteristics affect the travel decisions.

2.3 Literature review summary

From the review of previous studies, it can be understood that majority of models developed for DRT demand are specific to certain user groups like elderly or disable. But the prediction of DRT trip frequency that is dependent on various factors has not been addressed. This research intends to review all those identified potential variables affecting DRT ridership, identify new variables and build and compare a comprehensive set of statistical models to predict future traffic trends of DRT. According to our knowledge, this is the first such comparison of a comprehensive set of statistical models predicting DRT trip frequency.

Next chapter will describe the data which was used for this study and the data source, collection method, and its cleaning procedure.

CHAPTER 3: DATA COLLECTION

Data requirements for transit network assessment differ, depending whether the focus is on urban or rural areas. Two primary sources of data are used for determining transit connectivity measures in urban areas: (i) GTFS data for transit network characteristics, (ii) Tennessee socio-economic data from census, and Tennessee statewide transportation travel demand model. On the other hand, ACS data and DRT service data are used to develop statistical model for rural areas. In the following paragraphs the reader will be briefly introduced to the data types that have been selected and used for the specific project.

3.1 Urban Areas - GTFS Data

GTFS data for the cities of Nashville, Memphis, and Knoxville urban areas, was requested by the relevant transit authorities and has been analyzed in order to generate transit connectivity thematic maps (in ArcGIS environment). GTFS data is a standardized transit data format that incorporates public transit schedules and transit associated geographic data (transit stops, routes, etc.) To analyze the given datasets in the context of connectivity, as described earlier in the literature review, a code has been developed in R environment enabling data processing in order to monitor performance at node, line, transfer center, and zone level. The result of this type of analysis was the generation of shape files illustrating routes (and route stops) colored and sized accordingly to account for the level of connectivity that they provide. More on GTFS data is provided in Appendix A through D.

3.2 Urban Areas - Tennessee Traffic Analysis Zone (TAZ) Statewide model data

As mentioned in the literature review, captive ridership is correlated with low income, low or zero car ownership, etc. Tennessee TAZ Statewide model was used to indicate TAZs that are more likely to generate such transit demand (i.e., TAZs with low household income or low car ownership levels etc.). Combining the results of this data analysis along with GTFS data, comprehensive thematic maps were created to support the visualization of areas most likely to produce transit demand, because of the existence of potential captive riders.

3.3 Rural Areas – American Community Survey (ACS) data and DRT service data

The data for the empirical analysis was compiled from three different data sources which are shown in figure 3-1. DRT services provided in the state of Tennessee are considered as the case study in this research. All DRT trip occurrences for the year 2012 were collected from Tennessee Department of Transportation (TDOT). Each trip record includes trip attributes such as origin and destination ZIP Code, County, and trip purpose. The data is provided by TDOT is at the ZIP Code level to maintain anonymity of the traveler. For each DRT trip corresponding demographic data was collected from American Community Survey (ACS) for each of the ZIP Codes in Tennessee. The demographic characteristics include age, gender split, vehicle ownership, household size and structure, household income etc. Combining socio-economic data for each of the ZIP Codes from ACS 2011 with DRT trip data, a comprehensive dataset was developed. Further, service variables such as distance and travel time between ZIP Codes are determined using shortest path method and added to the dataset. All the trips from a specific origin ZIP Code to a destination ZIP Code have been accumulated to find total trip count for that pair. The final dataset contains number of trips between two ZIP Codes, the origin and destination ZIP Codes along with *DRT trip* features*, socio-economic and demographic* characteristics, and *level of service* measures.

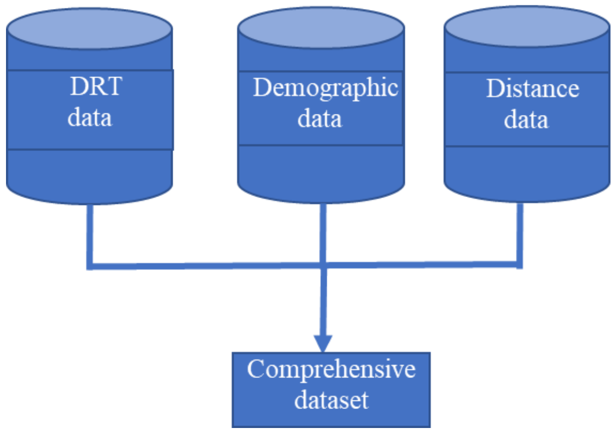

Figure 1: Data components.

Demographic data attributes of ZIP Codes are expected to be correlated to each other. To avoid multicollinearity problem, a correlation matrix is calculated consisting all continuous independent variables and one of the two highly correlated variables were dropped for inclusion in the final model dataset. There are total 640 ZIP Codes in the state of Tennessee which creates a total of 409,600 (640*640) origin-destination pairs. The number of *trips produced* from each ZIP Code to another ZIP Code is almost equal to the number of *trips attracted* by each ZIP Code from another ZIP Code. This is because almost all the trips reported in the travel diaries are round trips. Hence only production trips are considered for our study and a total of 205,120 (640*641/2) observations were found. By eliminating missing data for DRT trips, 185,500 records were kept for further analysis. 148,454 observations (80%) are used in model training and 37,046 (20%) observations are used for model validation.

In the next chapter, we will discuss about the data analysis, methodology, and result of urban transit study.

CHAPTER 4: METHODOLOGY AND RESULT-URBAN AREAS

4.1 Node and Line Connectivity

For the purposes of this study, data analysis has been based on the concept of connectivity index as described in the study of Welch and Mishra (2013). Instead of measuring connectivity as the transit frequency to a specific stop, Mishra et al (2012), addressed the shortcomings of this measure by developing a Node Connecting Power that considers information such as the opportunities accessible by transit, the time it takes to reach those opportunities, or the ability to transfer to different routes and modes to reach a broader array of activities. In a brief, the measure uses frequency, speed, distance, capacity, required transfers, and activity density of the underlying land use served by a transit node, for all modes including buses, light rail, bus rapid transit, and other similar transit facilities. The inbound and outbound connecting power of a transit line (on a specific node n) can be defined as follows:

$$P^{o}_{l,n}=\alpha C_l \times \beta V_l \times \gamma D^{o}_{l,n} \times \vartheta A_{l,n} \times \varphi T_{l,n} \tag{4.1}$$

$$P^{i}_{l,n}=\alpha C_l \times \beta V_l \times \gamma D^{o}_{l,n} \times \vartheta A_{l,n} \times \varphi T_{l,n} \tag{4.2}$$

where C_l is the average vehicle capacity of line l, F_l is the frequency on line l (60 is divided by F_l to determine the number of operation per hour), H is the daily hours of operation of line l, V_l is the speed of line l, and $D^{o}_{l,n}$ is the distance of line l, from node n to the destination. Parameters: α is the scaling factor coefficient for capacity which is the reciprocal of the average capacity of the system multiplied by the average number of daily operations of each line, β is the scaling factor coefficient for speed represented by the reciprocal of the average speed on each line, and γ is the scaling factor coefficient for distance which is the reciprocal of the average network-route distance and θ and φ are scaling factors. $A_{l,n}$ is the activity density which represents the development pattern (as a ratio of households and employment in a zone to the unit area) based on both land use and transportation characteristics and incorporates the quantity of opportunities accessible at each node in the system. Activity density is defined as:

$$A_{l,n}=\frac{H^{z}_{l,n}+E^{z}_{l,n}}{\Theta^{z}_{l,n}} \tag{4.3}$$

where, $H_{zl,n}$ is the number of households in zone z, $E_{zl,n}$ is the number of jobs, and $\Theta_{zl,n}$ is the area of zone z.

The node connectivity index, is then defined in to measure the aggregate connecting power of all lines, accessible to a given node:

$$T_{l,n}=\frac{\sum P^{t}_{l,n}}{\Theta^{n}_{l}} \tag{4.4}$$

The total connecting power of a line, is calculated as the sum of the averages of inbound and outbound connecting powers for all transit nodes on the line (equation (4.5)) scaled by the number of stops (S_l) on each line. The line connectivity can be defined as:

$$\theta_l=(|S_l|-1)^{-1}\sum P^{t}_{l,n} \tag{4.5}$$

This approach of connectivity was incorporated in a code developed in *R*, using as inputs GTFS and activity density data. Through processing, shape files that store connectivity related data of transit nodes and lines were generated. These files were combined with shape files containing socioeconomic data (at a TAZ level), and are, briefly, analyzed in the next section, in order to develop thematic maps that can depict both network connectivity and TAZs that may host potential captive riders.

4.2 Transit Catchment and Accessibility

To determine accessibility to transit stops transit catchment is defined as the buffer distance around a housing unit (e.g., half mile) in which at least one transit stop may or may not exist. Using this definition, a distance decay function may be formulated in order to the connectivity of transit nodes within these catchment (Euclidian) distances, of each housing unit (based on the centroid of a residential parcel in which they are concentrated). Equation (4.6) is used to calculate the pro-rated connectivity of a station within the catchment area.

$$\rho_{z_1,n}=\alpha\times e^{-bt_{h_1,n}} \quad (4.6)$$

where, $\rho_{z1,n}$ is the pro-rated connectivity, *a* and *b* are the parameters (based on empirical data) of pro-rated connectivity and $t_{h1,n}$ is the walk time to travel from housing unit h_1 to node transit stop *n*. For nodes outside the catchment $\rho_{z1,n}$ takes a value of zero. To obtain the connectivity index of a zone, the connecting power of each node in the catchment area is scaled by the number of transit nodes within the catchment area of each zone, and summed afterwards. Thus, a zone in a very dense transit area is made comparable to a zone in a less dense area. Connectivity index for a zone is given by equation (4.7):

$$\theta_{zu}=(|S_\omega|-1)\sum P^t_{l,n}(\rho_{n_1,n}) \quad (4.7)$$

4.3 Inequity Index

Inequity is a measure of the geographic concentration of a certain phenomenon (commonly used to describe the distribution of income among populations. The most common measure for this inequity is the *GINI index*, used to estimate the distribution of wealth among a population. The index measure is the difference between a perfect equity line (a straight 45-degree line) and a Lorenz curve, which measures the real distribution. GINI index values of 0 indicate perfect equity (coincidence of Lorenz curve with equity line), while zero values indicate perfect inequity. This principle can be applied to transit service quality, where it can be a measure of the cumulative proportion of population and the cumulative proportion of transit connectivity that is immediately accessible to the population. Generally, to estimate the GINI index integration is necessary to find the (area) difference between Lorenz curves and the equity line. An approximating approach to avoid this complex task is given by the formula of equation (4.8):

$$G_a=1-\sum_{k=1}^{n}(X_k-X_{k-1})(Y_k-Y_{k-1}) \quad (4.8)$$

where, G_a is the GINI index value for a population or sample *a*, X_k is the cumulative proportion of the population endowed with attribute *k* (in this case transit connectivity) for *k* = (0,...,*n*), and Y_k is the cumulative proportion of attribute *k*. For more information on the concepts of Catchment/Accessibility and inequity we refer to Welch and Mishra (T. F. Welch & Mishra, 2013).

4.4 Tennessee Statewide Model – Socioeconomic Data Analysis

In order to obtain socioeconomic data at the TAZ level on Average Household Income and Car Ownership, the Tennessee Statewide model was used.

4.5 Memphis Transit Network Connectivity Example

Utilizing the data files from connectivity and socioeconomic analysis, thematic maps have been developed for the cities of Memphis, Knoxville, and Nashville. Within the context of this study, an example of such thematic maps will be presented for the three cities and the case of the city of Memphis will be briefly discussed as an example.

In Figure 2, the line connectivity of the three cities transit network is presented, whereas the line numbering is following the codification of lines, as given by the local transit authority. The number shown on top of each line shows the bus route number. The connectivity (in terms of connectivity index as described earlier) of each line is represented by colored lines that are gradually moving from shades of red to shades of green and are also of increasing width as the connectivity index increases. At the same time the thematic maps include a colored TAZ layer that presents a specific socioeconomic value in each TAZ (here for example, this value is household income) where low incomes are represented by lighter green shaded TAZs and vice versa. This analysis can support agency decision makers when having to choose which areas of the city can be potential locations of investment in new lines or how existing transit lines could be modified in order to incorporate more areas that may be hosting potential captive riders. Taking Memphis as an example, network performance appears to be relatively good for locations and TAZs near the city center, while distant locations especially in West Memphis, but also in North and Northeast Memphis appear to be poorly connected. Figure 2, also suggests that central business districts (i.e., Downtown Memphis) that mostly serve as destinations rather than origins show high network density, while areas located at the outskirts and which are, most likely, commuting origins, are served by a less dense transit network.

Similarly, Figure 3 presents transit stops connectivity in comparison to household income. Transit network stops appear as blue dots of increasing size for increased values of connectivity indices. Taking Memphis as an example, stops located on major corridors (e.g., Poplar Avenue), have higher connectivity than others. Again, one can easily identify stops with low connectivity indices that reside in areas of also low incomes. Similar maps have been developed to assess low connectivity versus areas of low car ownership. By cross-examining such thematic maps simultaneously, the decision maker can, inductively, locate stops that could be further supported if needed, by including them to neighboring lines, or including them in new lines during planning processes and so forth.

4.6 Income and Connectivity

Results of connectivity at various income levels are shown in Table 1. Nine groups were defined based on the distribution of income and these groups were held constant for connectivity comparison in three cities. For each city three performance measures were selected, namely percent population, average connectivity, and percent connected. Percent population refers to share of overall population residing in a city respective to a defined income group. Connectivity of a zone is estimated by using the formula shown in equation 7. Connectivity of a specific income group is further estimated by average of connectivity indices for all zones within the group. Percent connected for each income group is the ratio of connectivity of each group out of the total connectivity for each city. Three top most connected income groups are shaded in gray for each income group and for each city. For instance, in Memphis, the top most connected segment is the highest income group followed by two other groups of incomes more than $100,000. The highest income group only contains population of 1.02 percent enjoys the 34.75% of connectivity. In contrast, the lowest income group only receives 3.91 percent of the overall connectivity. The highest percent of population is in the income group of $20,000 and $40,000. The 40.66% of population only receives 1.33% of the overall connectivity. This refers to the need for better transit connectivity for lower income group population where further investment may aid to satisfy the basic travel needs for captive riders. Connectivity of transit in Memphis is attributed to its land use, urban infrastructure, history of service, and sprawl. Connectivity is highest in the CBD or in the neighborhood area that serves the medical district, large scale establishments, and tourist attractions close to the Mississippi river. Some of the wealthiest neighborhoods are also close to the downtown, though these income segments do not need transit per say, but live in areas that serves as a pathway to the CBD.

Transit connectivity in Nashville is very similar to Memphis though there exist, distinct differences. First, the lower income groups are relatively better connected in compared with Memphis. For example, the lowest income group receives 10.39% of overall connectivity as compared to 3.91% in Memphis. However, still, low-income areas which may be hosting potential captive riders, could be a be potential locations of investment in new lines or modifying existing transit lines in order to serve people more efficient. Second, the largest population segment is contained in the income group of $40,000-$60,000 with 40.87% of the population which receives the connectivity with 6.84%. Average per capita income is higher in Nashville compared to Memphis. Third, the highest connected group is the income group of $140,000-$160,000, and 70% of connectivity is for 8.81% of the population with income higher than $100,000.

In Knoxville, the trend is somewhat different – the lowest income group is one of the top three categories receiving percent share of overall connectivity of the city, 16.59% of overall connectivity as compared to 3.91% in Memphis and 10.39 in Nashville. Also, the largest population segment with 47.93% of the population is contained in the income group of $40,000-$60,000 which receives the lowest connectivity with 5.13%. On the other hand, 51.17% of connectivity is for 3.5% of the population with more than $100,000 income. In addition, there are not enough population centers in the group of incomes more than $140,000. This does not mean that individuals do not earn more in Knoxville, but rather the average income of zones considered in this analysis does not portray any share of income groups higher than $140,000. Overall, the connectivity of all three represents some similarity but there exist unique characteristics of each.

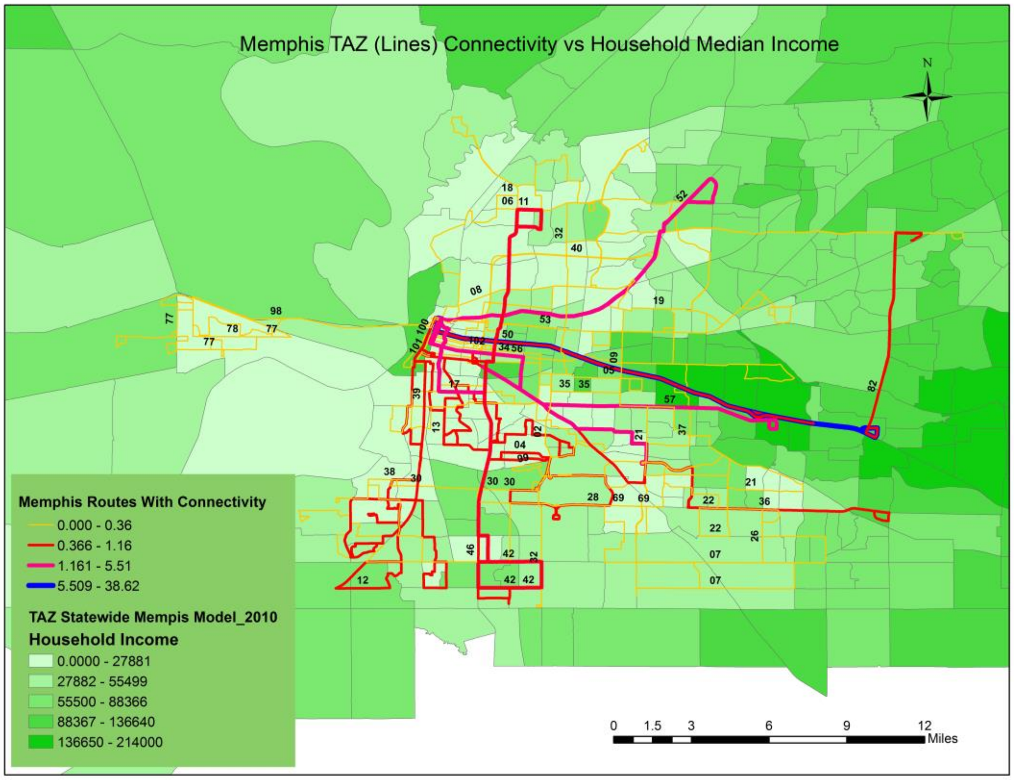

Figure 2a: Memphis transit line connectivity and TAZ household median income.

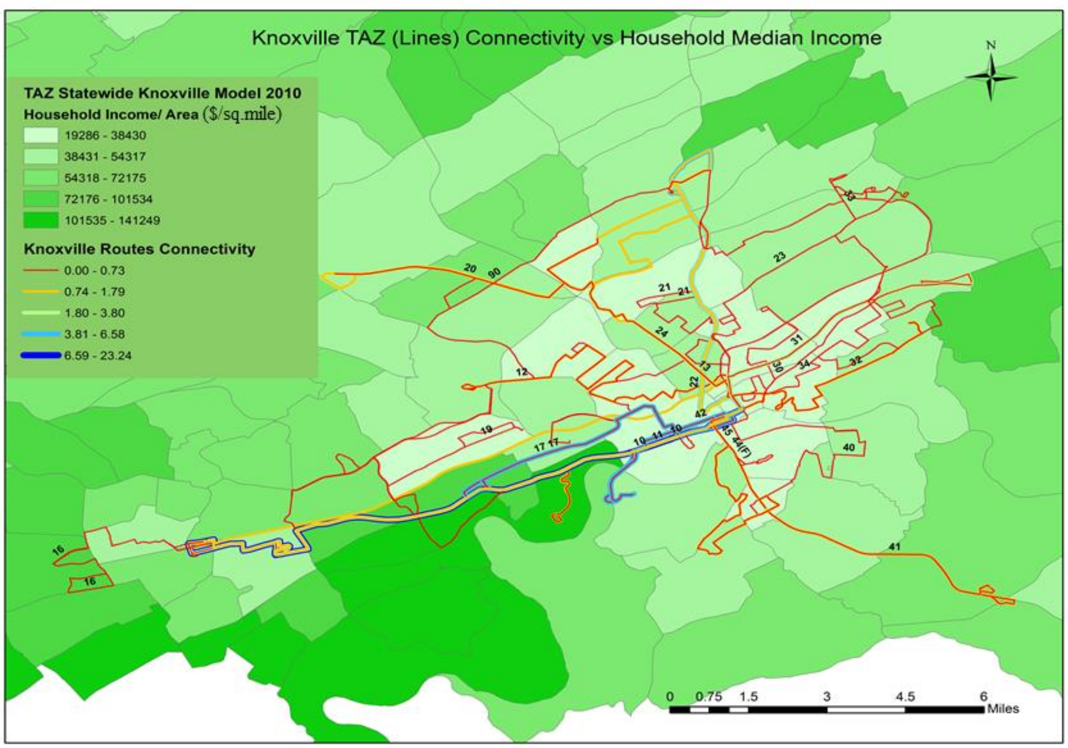

Figure 2b: Knoxville transit line connectivity and TAZ household median income.

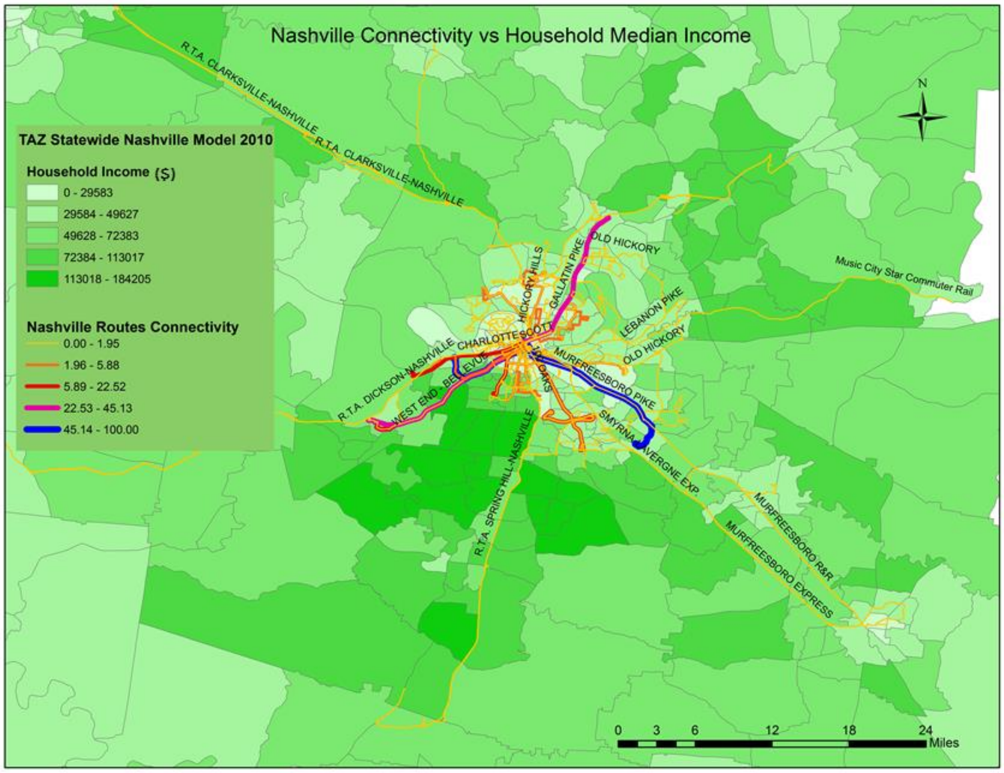

Figure 2c: Nashville transit line connectivity and TAZ household median income.

Figure 2: Transit line connectivity and TAZ household Median income.

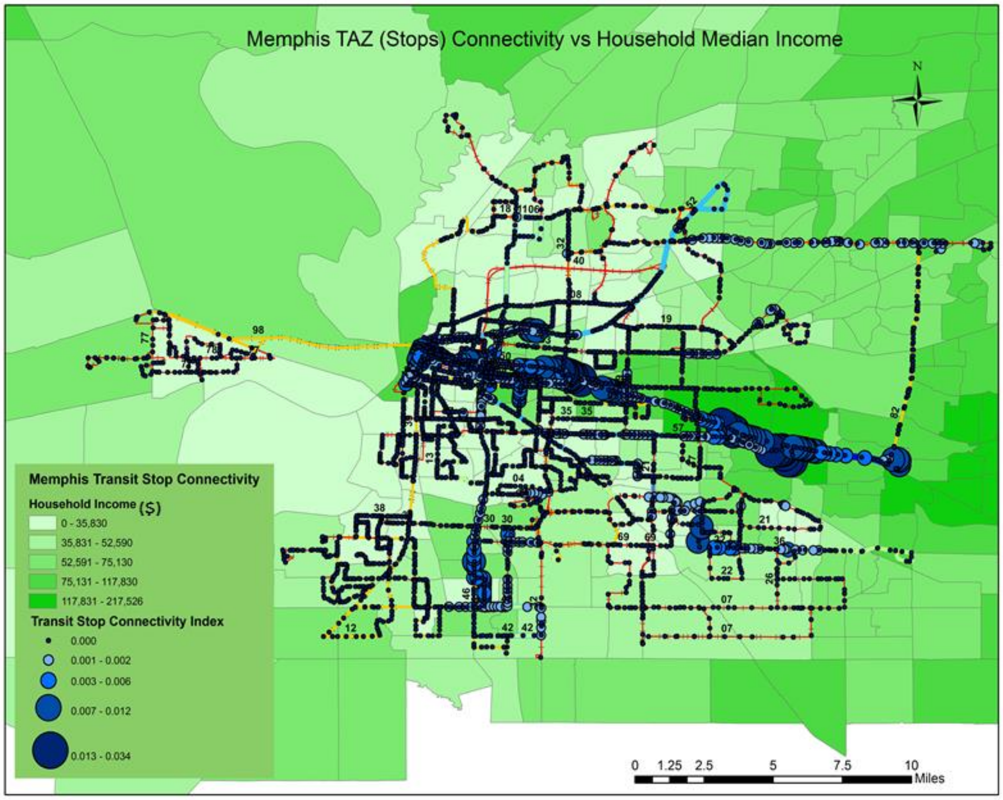

Figure 3a: Memphis transit stops connectivity and TAZ household median income.

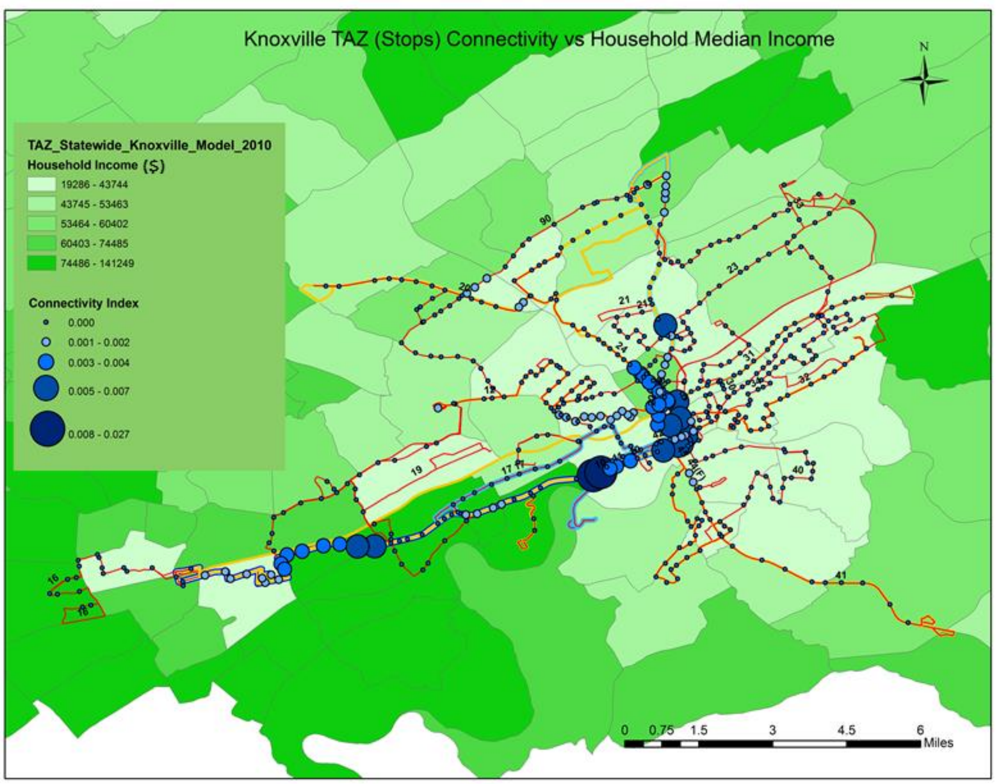

Figure 3b: Knoxville transit stops connectivity and TAZ household median income.

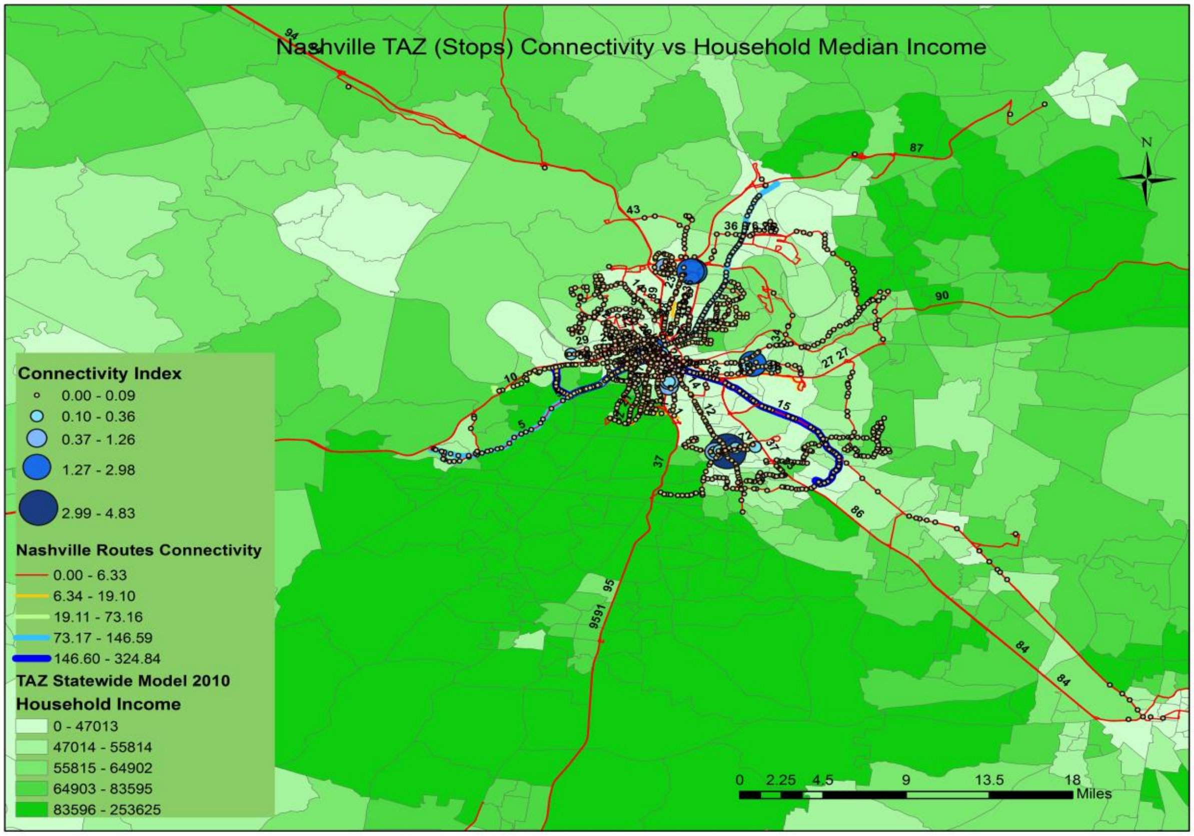

Figure 3c: Nashville transit stops connectivity and TAZ household median income.

Figure 3: Transit stops connectivity and TAZ household median income (Note: the numeric value above each line on the map refers to bus route number).

4.7 Vehicle Ownership and Connectivity

Vehicle ownership is a critical factor that represents need for transit connectivity. Considering distribution of vehicle ownership in three cities, three groups were formed namely less than equal to one vehicle, greater than one and less than or equal to two vehicles, and more than two vehicles. Percent population, connectivity, and percent connected are estimated for each vehicle ownership group and for each city. The results are shown in Table 2. One aspect common in all three cities is that the average vehicle ownership group is one to two vehicles per household. Both in Memphis and Knoxville the highest connected group is one to two vehicles per household. In Knoxville, the share of connectivity is 92.27% while in Memphis 56.63%. In Nashville, the highest connectivity is received by the lowest vehicle ownership group. The lowest vehicle ownership group receives 52.05% of overall connectivity. Certainly, a transit system like in Nashville is desirable from the needs of the users, and from the view point of the transit agency.

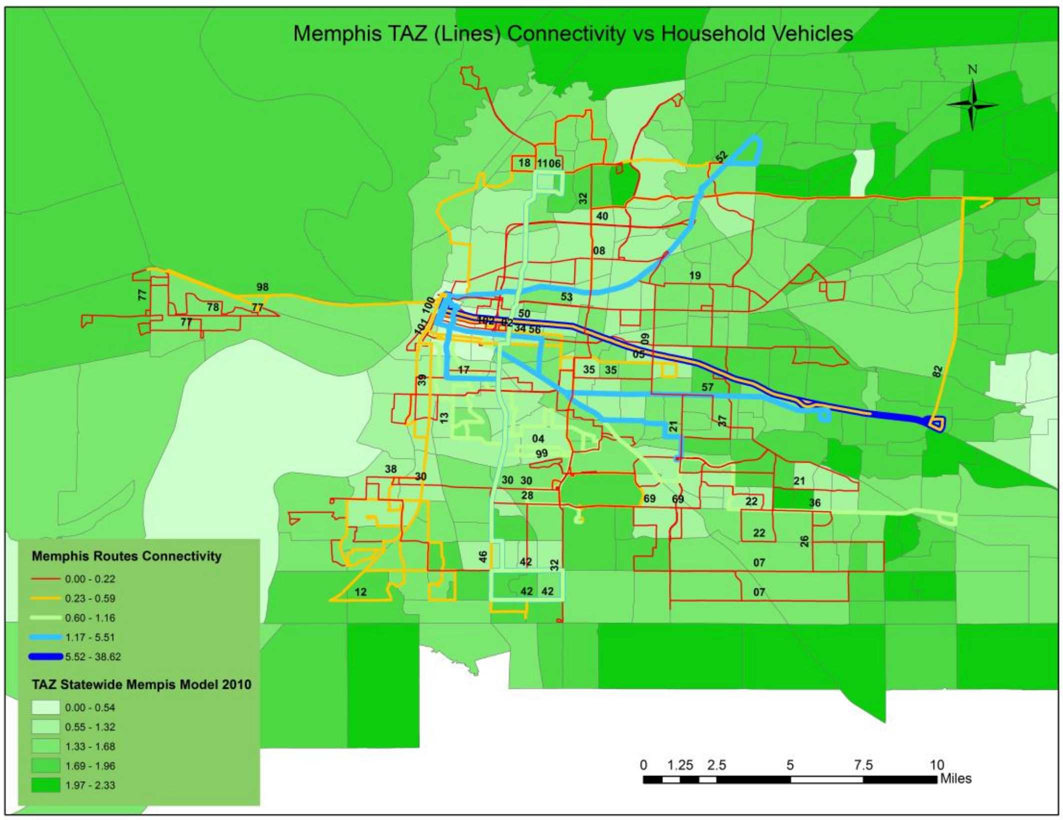

Figure 4a: Memphis transit lines connectivity and TAZ vehicle ownership.

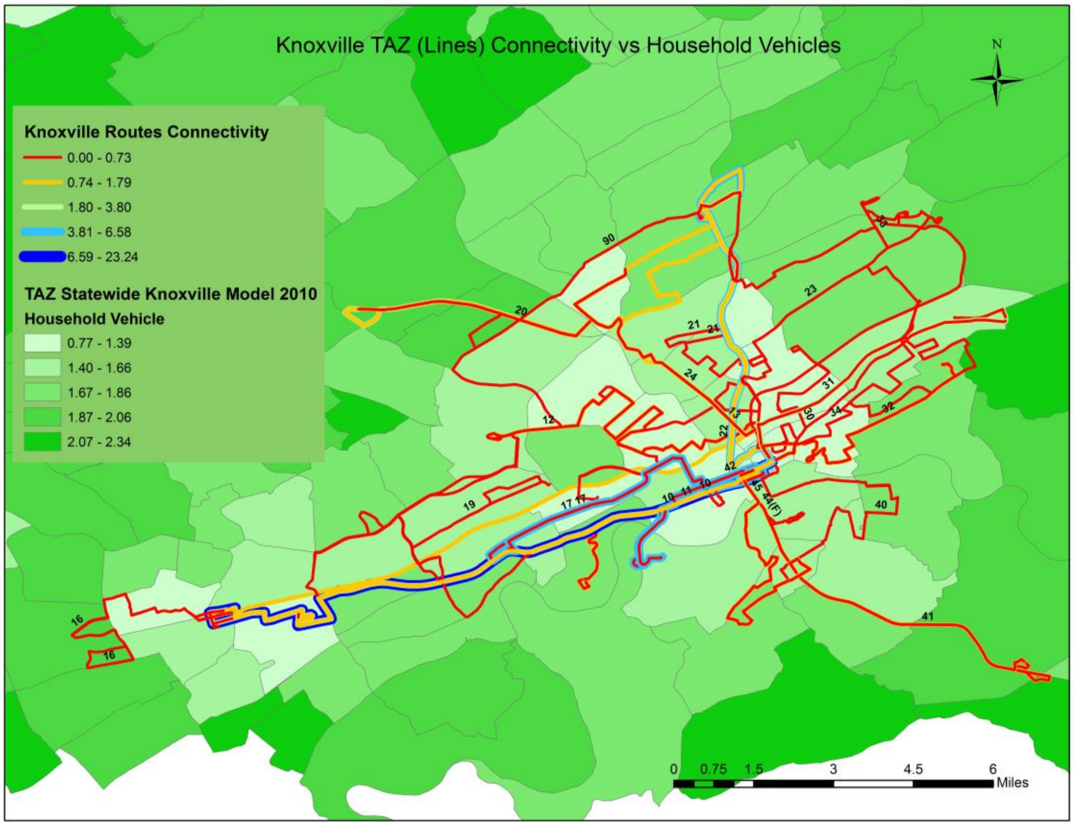

Figure 4b: Knoxville transit lines connectivity and TAZ vehicle ownership.

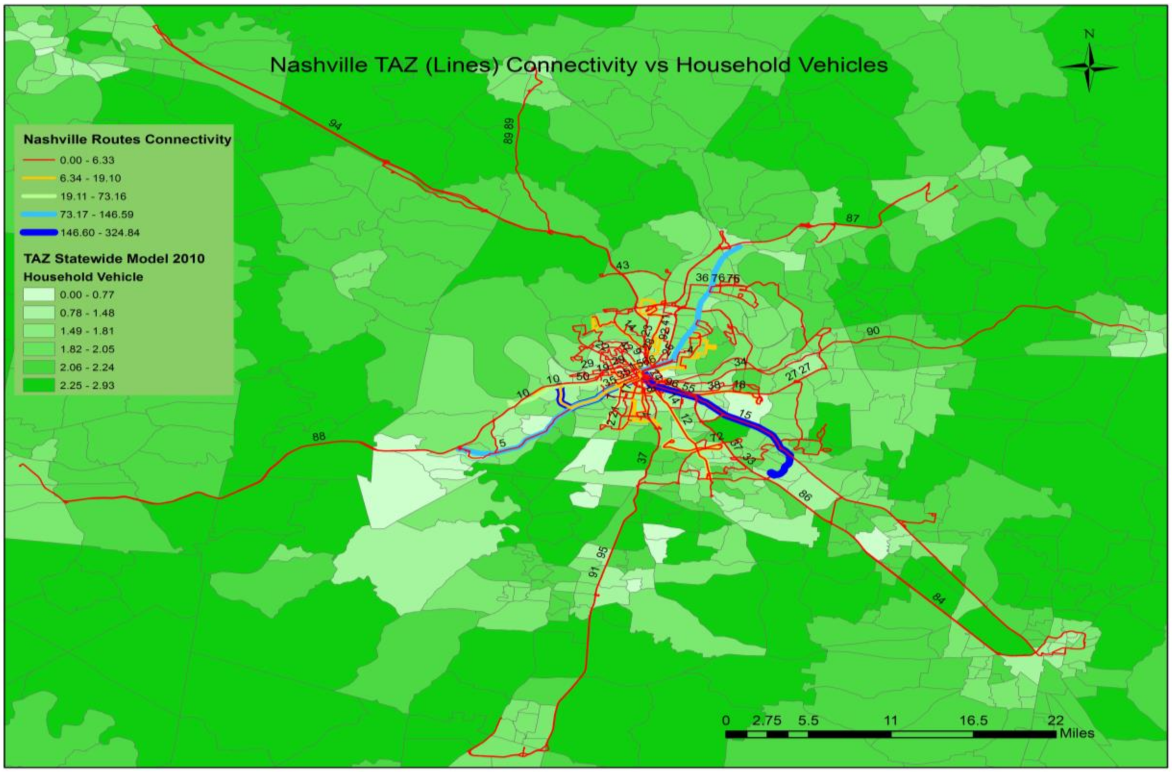

Figure 4c: Nashville transit lines connectivity and TAZ vehicle ownership.

Figure 4: Transit lines connectivity and TAZ vehicle ownership (Note: the numeric value above each line on the map refers to bus route number).

4.8 Employment and Connectivity

Employment by workplace for each TAZ is estimated, and based on its distribution seven categories are constructed for all three cities as shown in Table 3. The highest and lowest zonal employment is less than 500 and more than 2,500 respectively. Memphis is unique in one sense that the highest connected segment is the employment group 2,000 to 2,500. This group receives 62.85% of connectivity which shows that larger employment group is well connected, and transit service is available to place of work. Another interesting observation is that the highest employment group contains the second highest percent population the percent connectivity is third highest. In contrast, the top most employment group contains 21% population and is not in top three categories of percent connected. Nashville's transit connectivity does not predominantly serve to any specific employment group. The highest share of connectivity is received by the zonal employment group 1,500 to 2,000. Knoxville's transit connectivity also attributed to the higher zonal employment groups. Similar to Nashville, Knoxville also does not predominantly serve any specific employment group. The highest share of connectivity (27.4%) is received by the top most employment group. The highest share of population is also contained within the highest employment group. Such a population and employment distribution in Knoxville demonstrates adequate transit service is provided to higher zonal employment and population.

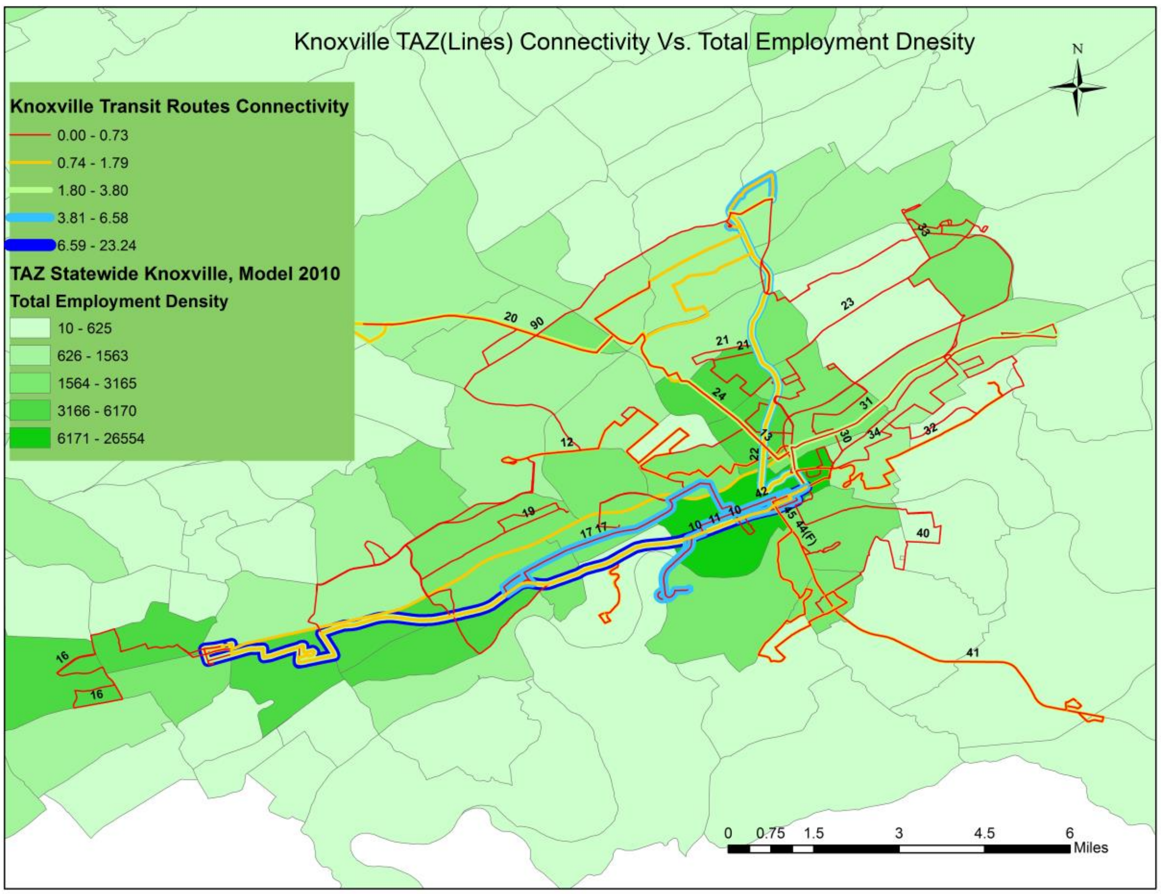

Figure 5a: Knoxville transit lines connectivity and TAZ total employment.

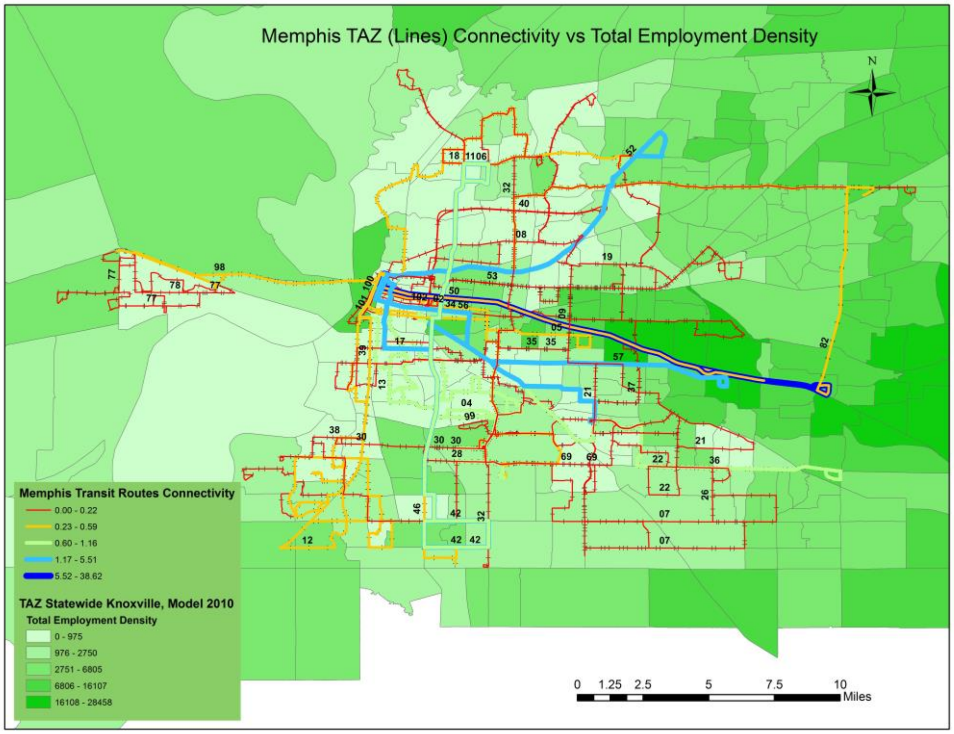

Figure 5b: Memphis transit lines connectivity and TAZ total employment.

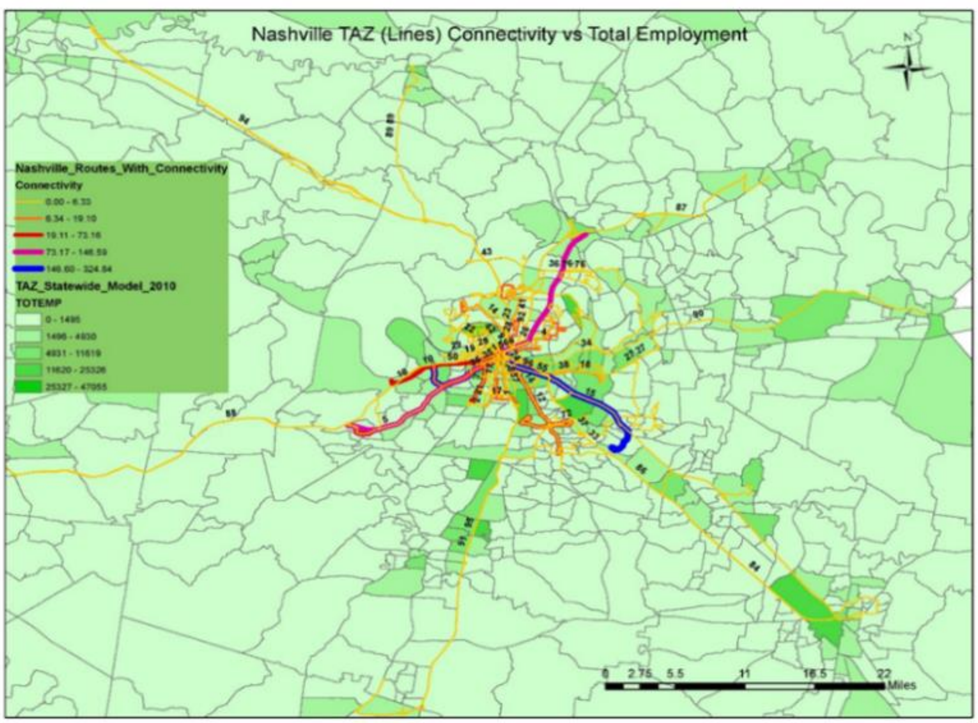

Figure 5a: Nashville transit lines connectivity and TAZ total employment.

Figure 5: Transit line connectivity and TAZ total employment (Note: the numeric value above each line on the map refers to bus route number).

4.9 Population and Connectivity

Table 4 shows connectivity for all three cities based on zonal population. Nine categories are constructed based on population distribution. In Memphis, approximately the top three population groups receive connectivity accordingly. Similar transit connectivity distribution is not observed in Nashville. The highest percent population group is not among the top three connected categories. Similar to Nashville, Knoxville's highest population group is not among the top recipient of transit connectivity. However, the population distribution alone does not portray other imperative attributes such as household income, vehicle ownership, and employment. Since transit service is decided upon a number of other factors besides just population results in Table 1 through Table 3 should be viewed in conjunction with population distribution.

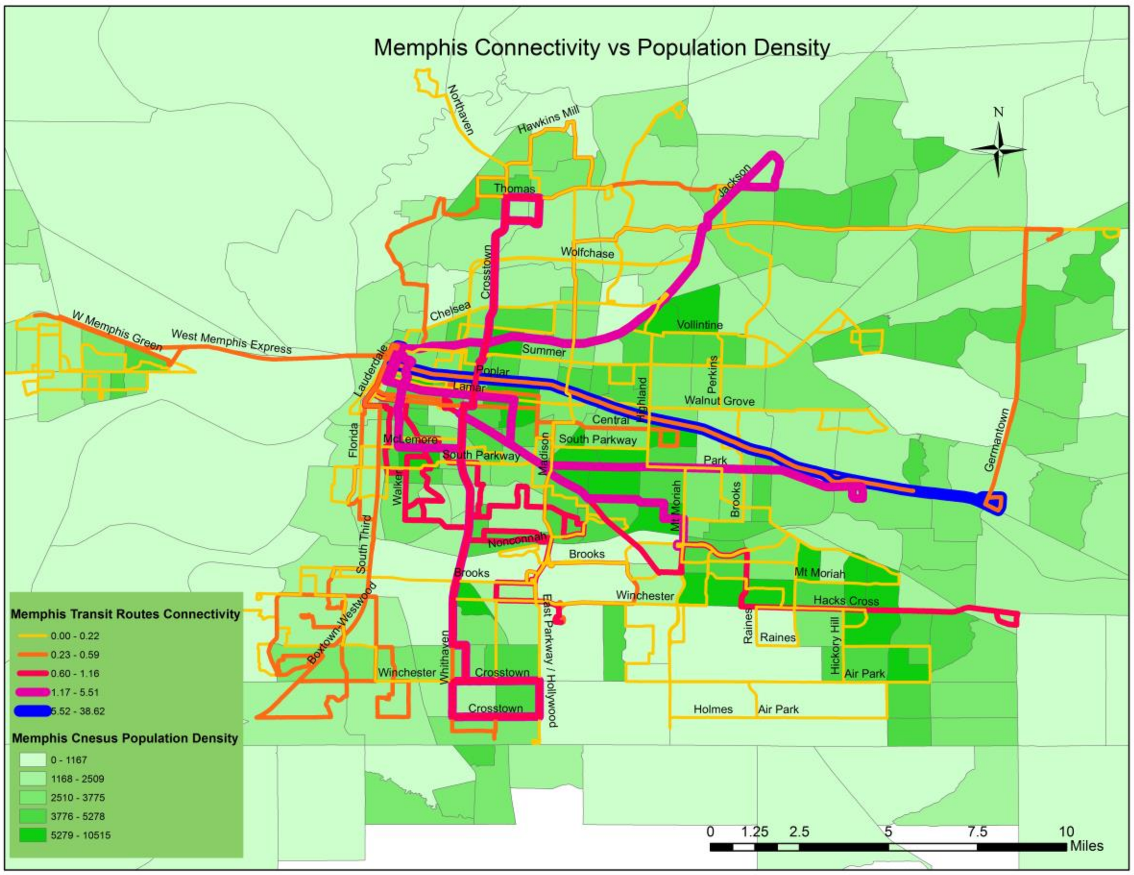

Figure 6a: Memphis transit lines connectivity and population density.

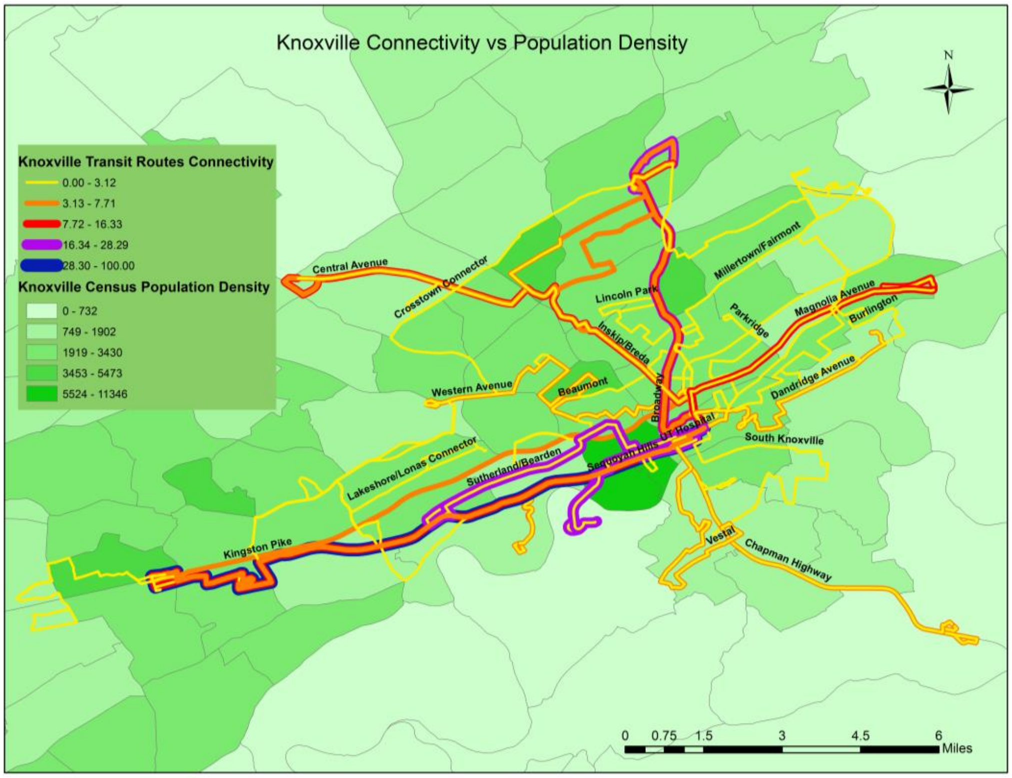

Figure 6b: Knoxville transit lines connectivity and population density.

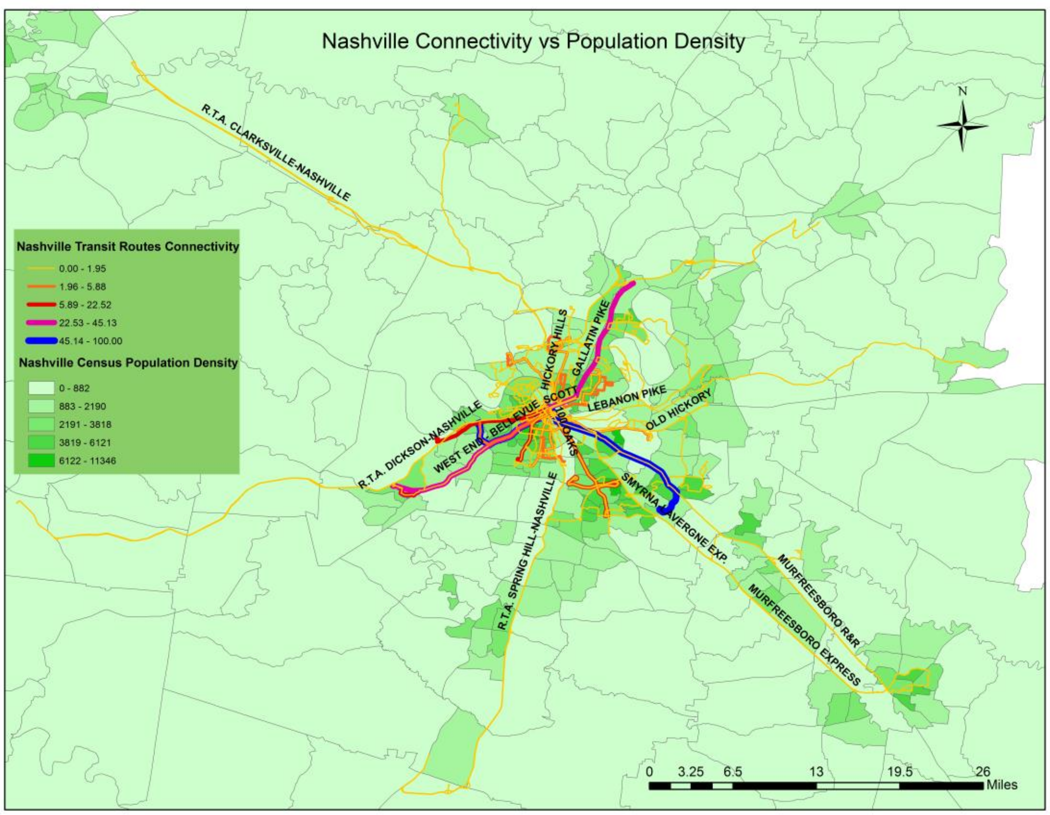

Figure 6c: Nashville transit lines connectivity and population density.

Figure 6: Connectivity vs Population Density.

4.10 Equity

Table 5 provides the GINI index values for the three cities (based on household income, vehicle ownership, employment and population). Results showcase a mix of equity conditions for all three cities. For example, the population based GINI index shows that none of the three cities provide equitable transit connectivity while Knoxville and Nashville do a better job on providing the equitable connectivity to transit riders if income or vehicle ownership is used as the population grouping factor. Unfortunately, Memphis (and always based on the GINI index) provides inequitable transit connectivity except by population criteria. If an agency's goal is to spread high-quality transit service among all households, the scores should be evaluated with a goal of reducing the GINI index towards zero. On the other hand, should an agency wish to provide very high-quality transit service to a highly concentrated geographic area, a score moving towards a value of one would be the goal. In either case, the framework provides a tool to measure distribution at several levels of aggregation.

Table 1: Transit connectivity by household income.

Income Group ($)	MEMPHIS			NASHVILLE			KNOXVILLE		
	% Population	Connectivity	% Connected	% Population	Connectivity	% Connected	% Population	Connectivity	% Connected
<20,000	1.14	1.90	3.91	0.76	13.82	10.39	3.02	1.95	16.59
20,000-40,000	40.66	0.64	1.33	15.16	10.32	7.76	25.49	0.87	7.42
40,000-60,000	29.07	0.71	1.45	40.87	9.10	6.84	47.93	0.60	5.13
60,000-80,000	19.21	2.17	4.48	23.20	3.66	2.75	13.68	1.61	13.74
80,000-100,000	4.54	1.59	3.29	11.20	3.01	2.26	6.38	0.70	5.96
100,000-120,000	2.91	9.58	19.76	3.93	23.32	17.53	2.09	1.91	16.31
120,000-140,000	0.74	8.99	18.54	2.86	2.63	1.98	1.41	4.09	34.86
140,000-160,000	0.72	6.06	12.49	0.71	43.39	32.62	0.00	0.00	0.00
>160,000	1.02	16.85	34.75	1.31	23.75	17.86	0.00	0.00	0.00

Note: Shaded area represents top three income groups by percent connected

Table 2: Transit connectivity by vehicle ownership.

Vehicle Ownership Group	MEMPHIS			NASHVILLE			KNOXVILLE		
	% Population	Connectivity	% Connected	% Population	Connectivity	% Connected	% Population	Connectivity	% Connected
<=1	7.91	1.07	38.59	4.35	13.49	52.05	2.71	0.64	2.71
>1 and <=2	89.79	1.56	56.36	77.00	9.51	36.67	92.27	0.96	92.27
>2	2.29	0.14	5.05	18.65	2.93	11.28	5.02	0.96	5.02

Note: Shaded area represents the top most vehicle ownership group by percent connected

Table 3: Transit connectivity by employment density.

Employment Group	MEMPHIS			NASHVILLE			KNOXVILLE		
	% Population	Connectivity	% Connected	% Population	Connectivity	% Connected	% Population	Connectivity	% Connected
<500	40.40	0.6	3.60	31.83	5.3	8.45	7.27	0.4	8.19
500-1,000	16.19	0.4	2.25	20.47	8.5	13.59	11.51	0.4	8.38
1,000-1,500	10.40	0.6	3.25	10.59	8.1	13.00	23.30	1.0	20.93
1,500-2,000	4.53	2.7	15.11	8.78	18.4	29.44	7.70	0.5	10.21
2,000-2,500	3.95	11.2	62.85	7.33	15.3	24.52	7.65	1.1	24.89
>2,500	24.53	2.3	12.94	21.00	6.9	11.00	42.56	1.3	27.40

Note: Shaded area represents top three employment groups by percent connected

Table 4: Transit connectivity by population density.

Population Group	MEMPHIS			NASHVILLE			KNOXVILLE		
	% Population	Connectivity	% Connected	% Population	Connectivity	% Connected	% Population	Connectivity	% Connected
<500	0.02	0.13	1.29	0.10	12.17	15.52	0.00	0.00	0.00
500-1,000	0.73	0.47	4.79	1.00	0.98	1.24	0.00	0.00	0.00
1,000-1,500	2.00	0.43	4.37	3.41	3.28	4.19	0.49	0.60	9.59
1,500-2,000	5.61	0.72	7.33	9.23	12.72	16.23	4.20	1.05	16.68
2,000-2,500	14.79	1.62	16.52	18.25	6.77	8.64	12.58	1.00	15.95
2,500-3,000	14.77	1.10	11.17	17.77	5.07	6.47	19.60	1.13	17.90
3,000-3,500	20.99	2.05	20.92	22.56	4.59	5.85	13.96	0.74	11.71
3,500-4,000	26.22	2.03	20.69	19.09	16.37	20.88	20.93	0.80	12.68
>4,000	14.87	1.27	12.93	8.59	16.46	20.99	28.24	0.98	15.48

Note: Shaded area represents top three population groups by percent connectivity

Table 5: GINI index by socioeconomic criteria.

Criteria	City		
	Knoxville	Memphis	Nashville
Household Income	0.21*	0.68	0.38
Vehicle Ownership	0.46	0.67	0.14*
Employment Density	0.86	0.91	0.79*
Population Density	0.85	0.84*	0.88

*Note: * represents most equitable transit connectivity by criteria*

CHAPTER 5: METHODOLOGY AND RESULT-RURAL AREAS

5.1 Data Description

This section provides a comprehensive view of the DRT data along with the descriptive statistics of the complete data set (combination of ACS and DRT data).

For this project, we have collected demographic data from ACS and travel impedance data (such as travel distance, time etc.) from TDOT which are the primary data source for trip related information. The final dataset contains number of trips between two ZIP Codes, the origin and destination ZIP Codes demographic profile, and the travel cost (distance, time, etc.) between them. Figure 7 shows proportion of all trip purposes of DRT. The highest proportion of trips was for medical purposes (52.37 %). Hence, medical trip is the most important cause of making demand response trip in rural areas of Tennessee. Second largest trip purpose was for work related activities (employment, work, and customer home) which combines to 15.17 % of total travels. Other significant causes of trip request were recreation (3.62 %), senior center (3.37 %), and shopping (3.97 %).

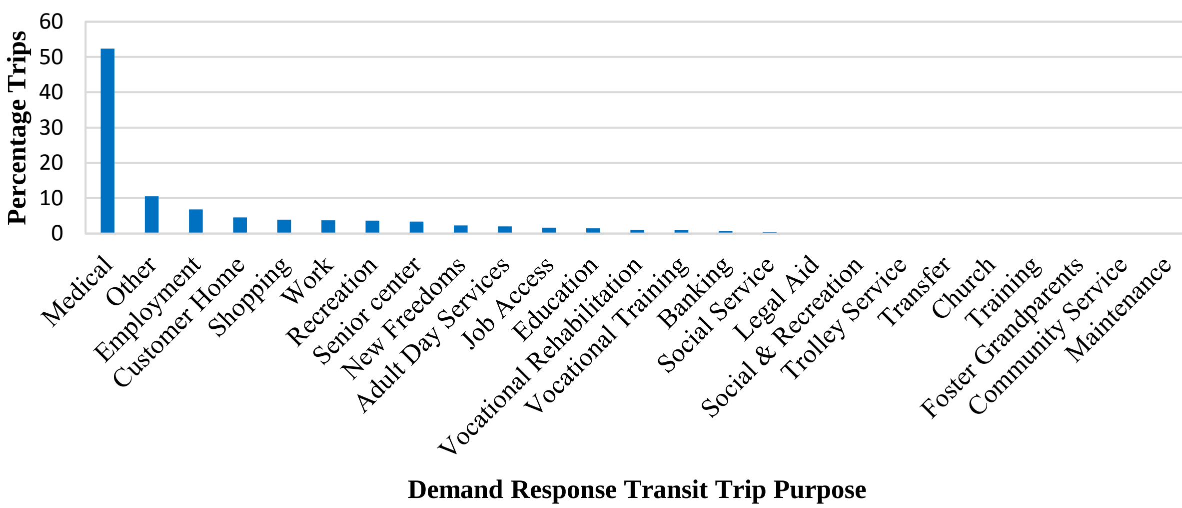

Figure 7 : DRT trip purpose frequency.

Figure 8 shows trip production for each ZIP code in the state of Tennessee. For illustration, trip count is divided into six quintile levels (0, 1-4, 5-71, 72-346, 347-1099, 1100 and above). In the figure, deep green colored areas are highlighted as most trip production region. It is clear that most of the smaller cities closer to big cities are the main source of demand-response traffic generation. As example, Columbia city (ZIP Code 38401) is smaller sub-urban city, which is 44 miles away from Nashville, and produced highest number of trips in the whole state (13.10% of total trips). The second highest (6.31%) trip generating region was Tullahoma city (ZIP Code 37388) which is around 74 miles away from Nashville. Another significant trip generating area is Shelbyville city (ZIP Code 37160) which is 57 miles from Nashville. These information give insight of selecting covariates which may influence DRT trips.

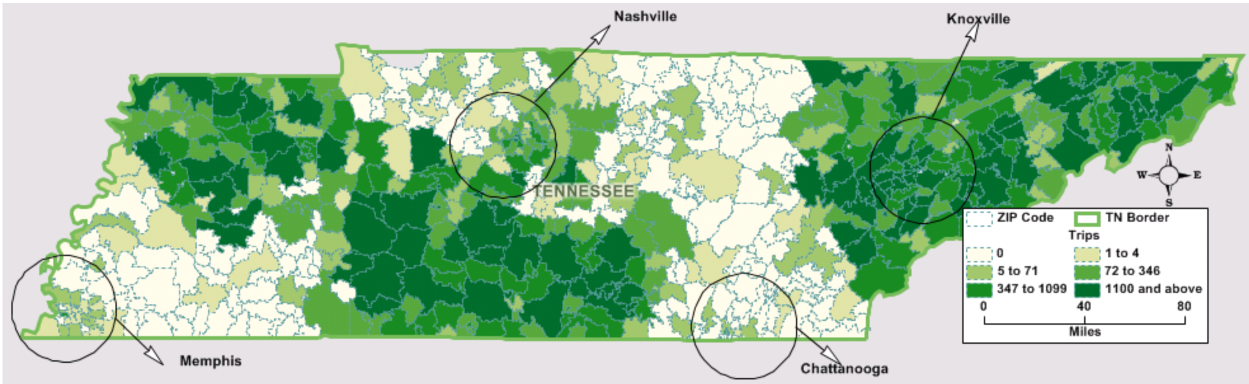

Figure 8: Trip production at ZIP Code level.

The most trip generating ZIP Codes and their purposes of trips in the map are shown in Figure 9. Most of the travels made in high trip generating ZIP Code are due to medical purpose. There are few ZIP Codes, near Knoxville, have remarkable number of trips for employment purpose. Another important thing to note, out of 640 ZIP Codes, 337 have trip count zero which is probably because of the fact that all ZIP Code may not have DRT service or people of those ZIP Code are reluctant to use that service.

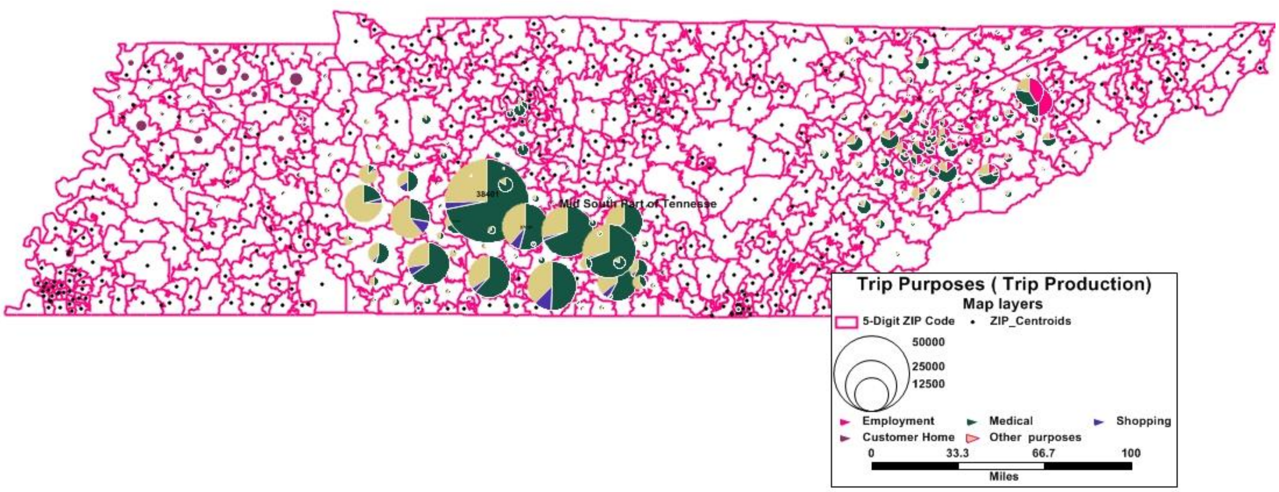

Figure 9: Trip Purposes (Trip production) at ZIP Code level.

Figure 10 shows the total amount of trip count and its purpose based on destination ZIP Code. There is no noticeable difference between trip production and trip attraction map based on the travel dataset because majority of DRT trips return to the origin ZIP code.

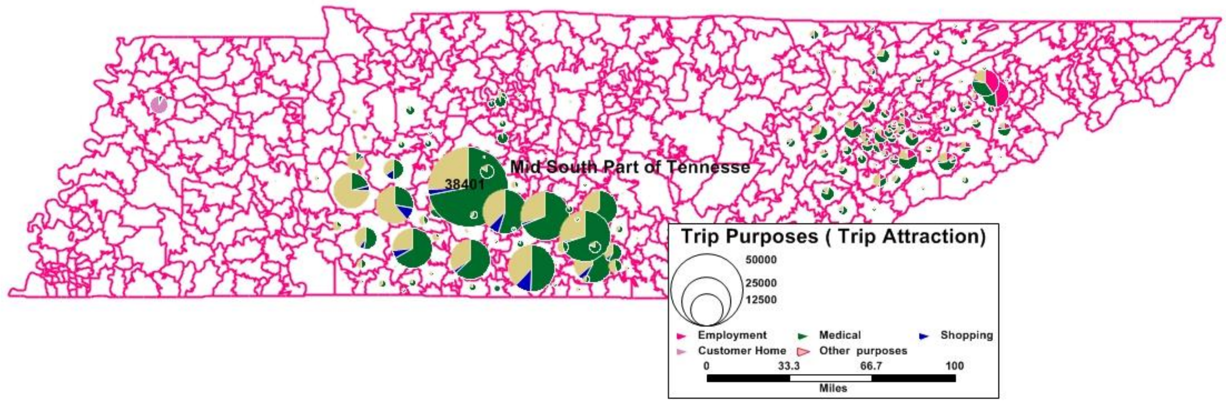

Figure 10: Trip Purposes (Trip Attraction) at ZIP Code level.

Flow network map of ninety-five counties is presented (shown in Figure 11). The highest number of trips are produced or attracted from Maury County to other counties. There are no transit trips has been produced or attracted by Macon, Pickett, Clay, Fentress, Overton, DeKalb, Van Buren, Bledsoe and Fayette. This analysis can support agency decision makers when having to choose which areas of the counties can be potential locations of investment in new demand response services or how existing transit services could be modified in order to incorporate more areas that may be hosting potential captive riders.

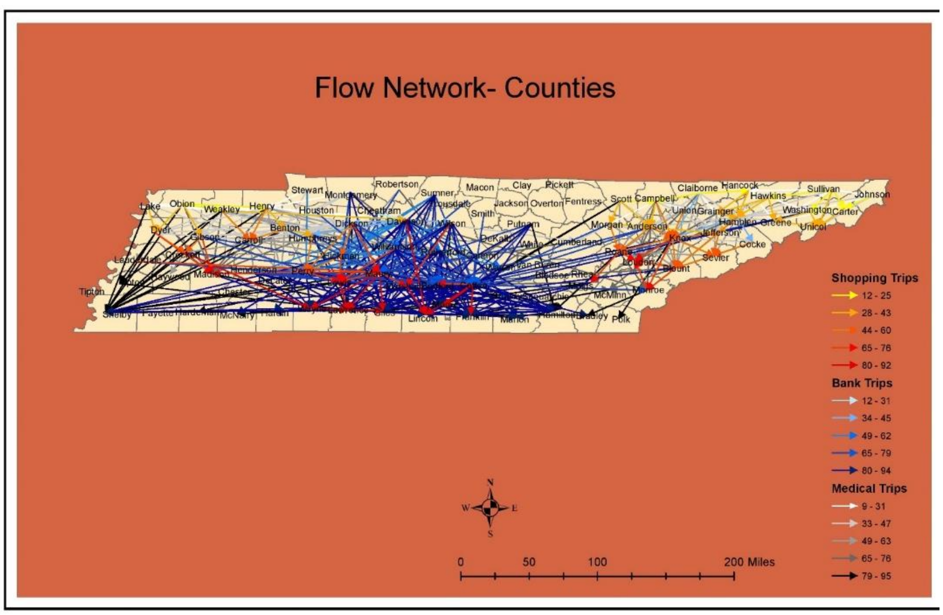

Figure 11: Flow network at county level.

Figure 12 presents trips production of each ZIP Code in comparison to trips attraction. Trips production and trips attraction of each ZIP Code are presented by bars of increasing height for increased values of trips attraction and trips production. Number of trips produced and attracted of each ZIP Code is almost same. Transit users normally take this service for medical, shopping, work, employment purposes, hence they probably like to go to their nearest destination to serve their needs from their origin points and they return to their origin. Taking ZIP Code 38401 as an example has the highest number of produced and attracted trips than others. By this map, the decision maker can locate areas that could be further supported if needed, by including them to neighboring demand response transit service providers, or including them in new transit service providers during planning processes and so forth.

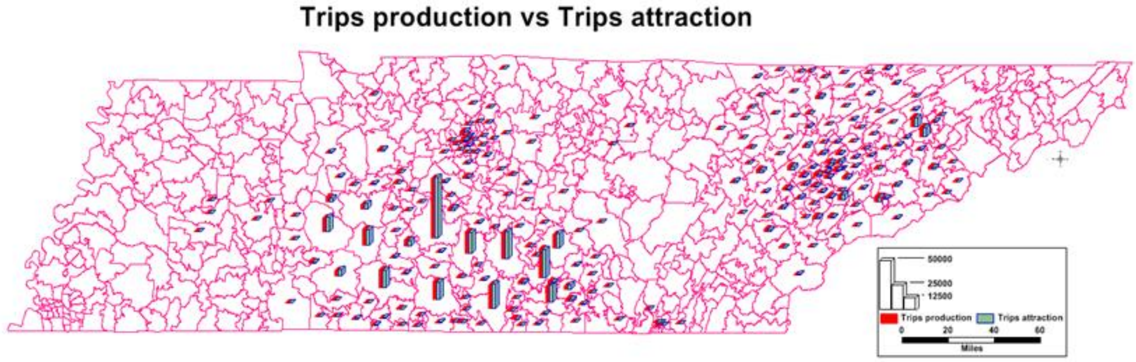

Figure 12: Trips production vs trips attraction at ZIP Code level.

Figure 13 illustrates trips production of each zip code in metropolitan areas of Knoxville and Nashville. As it is shown the map, even in the metropolitan areas, most of the trip are concentrated in central counties and zip codes.

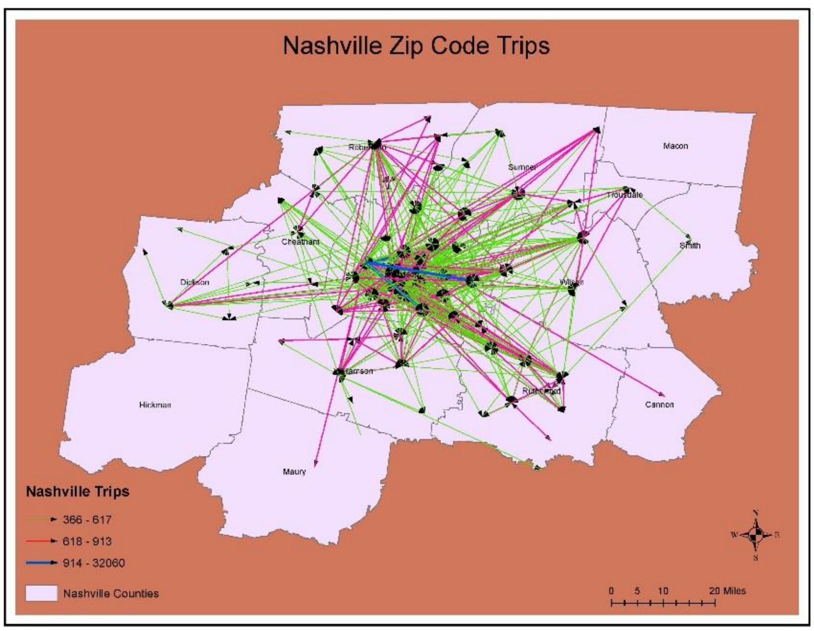

Figure 13a: Nashville zip code trips.

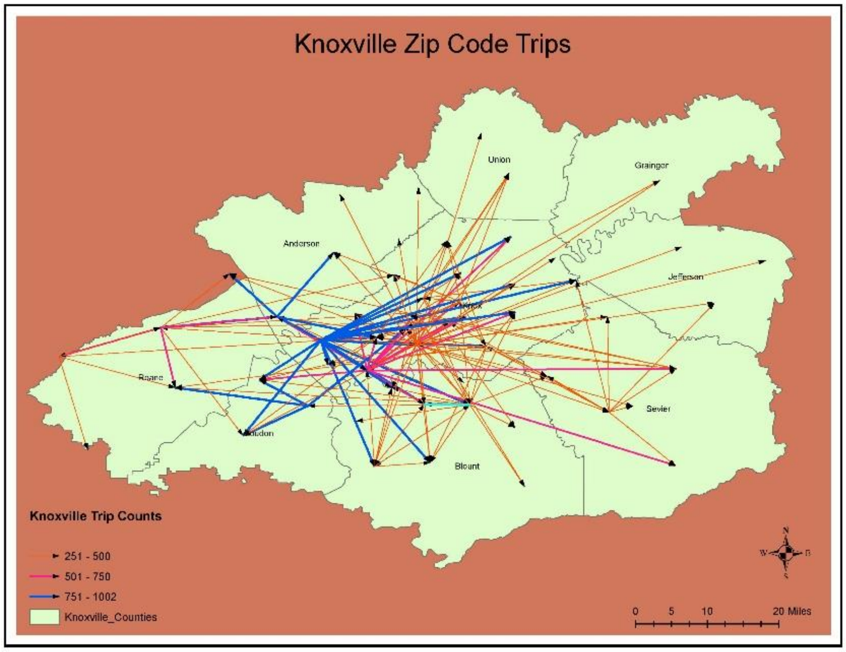

Figure 13b: Knoxville zip code trips.

Figure 13: ZIP Codes trip production- Metropolitan areas.

In society, individuals who may not have the option of traveling by car are seniors, teenagers, persons with disabilities, those with low incomes, and those without access to a car. Analyzing these populations can be helpful in understanding the potential for transit use in the area. A review of 2007-2011 ACS data provided the following findings:

- A total of 15.0% of the population with disabilities, and can be considered as a primary source of DRT trips.
- Senior population (age 65 and over) share is 13.286%. Few ZIP Codes (i.e. 38558, 37326) have this proportion more than 50%.
- 6.202% of the occupied housing units, do not own a vehicle, and 32.551% own only one vehicle.
- Children who are old enough to travel alone, but not yet old enough to drive are also a good source of DRT traffic. In Tennessee, 6.654% of the population is between 10 and 14.
- 28.304% of households have an income lower than $25,000 per year and approximately 8.965% have an income below $10,000.

Table 6 represents the potential independent variables selected for models along with brief description. The descriptive statistics of those variables is presented below in Table 7.

TABLE 6: Potential independent variables.

Variable Description	Type	Categories (if applicable)
Destination ZIP Code Retail Trade related businesses	Continuous	
Destination ZIP Code Wholesale Trade related businesses	Continuous	
Distance between two ZIP Code (mi)	Continuous	
Origin ZIP Code Population density (/)	Continuous	
Origin ZIP Code Proportion of white population	Continuous	
Origin ZIP Code Household median income	Continuous	
Origin ZIP Code Average household Size	Continuous	
Origin ZIP Code Homeowner vacancy rate	Continuous	
Origin ZIP Code Renter occupied housing Unit	Continuous	
Origin ZIP Code Disabled population	Continuous	
Origin ZIP Code Population aged 14 years or less	Continuous	
Origin ZIP Code Population aged 65 years or over	Continuous	
Origin ZIP Code Household income 200K or more	Continuous	
Rural urban commuting area type of Origin ZIP Code	Categorical	1- Metropolitan 2- Micropolitan 3- Small town 4- Rural

Rural urban commuting area type of Destination ZIP Code	Categorical	1- Metropolitan 2- Micropolitan 3- Small town 4- Rural
Dominant Sex of Origin ZIP Code	Categorical	0-Female 1-Male
Dominant Race of Origin ZIP Code	Categorical	0-Black 1-White

TABLE 7: Descriptive statistics.

Variable	**Min**	**Mean**	**Max**	**Standard deviation**
Destination ZIP Code Retail Trade related businesses	0	40.55	401	63.01
Destination ZIP Code Wholesale Trade related businesses	0	11.94	275.00	24.70
Distance between two ZIP Code (mi)	0	176.60	544.50	106.26
Origin ZIP Code Population density (/)	4.72	475.10	17,840	1,235.69
Origin ZIP Code Proportion of white population	0.02	0.859	1.00	0.19
Origin ZIP Code Household median income	8,524	40,000	136,200	13,652.29
Origin ZIP Code Average Household Size	1.46	2.62	19.96	0.96
Origin ZIP Code Homeowner vacancy rate	0	1.49	37.82	2.31
Origin ZIP Code Renter occupied housing Unit	0	1,119	14,530	1,748.65
Origin ZIP Code Disabled population	0	1,444	9,259	1,654.53
Origin ZIP Code Population aged 14 years or less	0	1,781	16,800	2,399.16
Origin ZIP Code Population aged 65 years or over	1	1,284	7,975	1,502.29

Origin ZIP Code Household income 200K or more	0	93.93	4,448	242.51
Rural urban commuting area type of Origin ZIP Code	1	2.24	4	1.16
Rural urban commuting area type of Destination ZIP Code	1	1.84	4	1.10
Dominant Sex of Origin ZIP Code	0	0.35	1	0.47
Dominant Race of Origin ZIP Code	0	0.92	1	0.26

Figure 14 represents the relationship between trip count (dependent variable) and various ZIP Code related socio-economic factors (independent variables). It is difficult to understand by directly looking at the graph whether there is a positive or negative relationship between trip count and these socio-economic factors. This may be due to the fact that all factors affecting DRT demand need to be considered when developing a relationship between them. Hence, a set of econometric models were necessary to conclude how these socio-economic factors affect DRT trips.

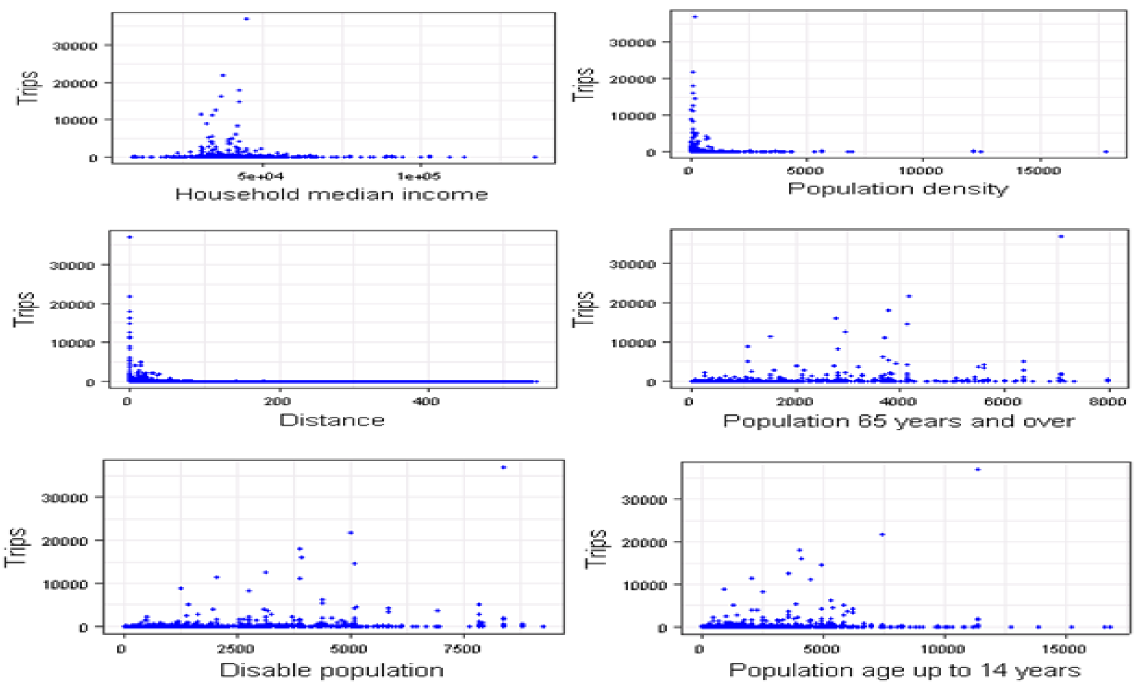

Figure 14 : Relationship between dependent variable (DRT trip) and other potential independent variables.

5.2. DRT Model Descriptions:

Count or frequency models are usually considered as a parametric model where the model parameters are estimated from count observations. The parameters of the underlying distribution are specified as a function of different covariates to capture their influence on count dependent variable. Count variable has non-negative integer value which implies that a log-linear model is better fit for a count variable. A linear regression model generally produces negative predicted outcomes and there is a substantial problem of heteroscedasticity. Another advantage of using the log-linear specification is that, with count data, the effects of predictors are often multiplicative rather than additive. That is, one typically observes small effects for small counts, and large effects for large counts. If the effect is in fact proportional to the count, working in the log scale leads to a much simpler model. Poisson model is a good choice in this case. The Poisson and Negative Binomial (NB) log-linear models are the two most commonly implemented parametric model in the literature for count data modelling (Washington, Karlaftis, & Mannering, 2010). The Poisson model has a restrictive assumption of equi-dispersion property i.e., the expected mean parameter of the Poisson distribution is equal to the variance. The NB model overcomes that assumption, which makes it suitable for cases when there is over-dispersion in the count data being modeled. Another aspect of considerable importance while modeling count data is over-representation of zeroes beyond the probability mass implied by the standard count models – a property referred to as the excess zeroes problem. Several variants of standard models including the zero-inflated count models, hurdle count models, and zero inflated mixed effect models were developed to address the excess zeroes problem (Fang et al., 2014; Gurmu, 1998; Hu et al., 2011; Hur et al., 2002; Moghimbeigi et al., 2008; Yang et al., 2016; Yau et al., 2003; Yau and Lee, 2001). In this section, we briefly present specification of each model type for analyzing DRT trip frequency. A brief discussion of alternate modeling methods are follows.

5.2.1 Poisson regression model:

In Poisson model, the probability of an event count y_i, given the vector of covariates X_i, is given by the Poisson distribution:

$$P(Y_i=y_i|X_i)=\frac{e^{-\lambda_i}\times\lambda_i^{\ y_i}}{y_i!},\ y_i=0,1,2,\ \ldots\ldots \tag{5.1}$$

The mean parameter λ_i is a function of the vector of covariates in period:

$$E(y_i|X_i)=\lambda_i=exp(X_i'\beta) \tag{5.2}$$

where β is a (k+1) x1 parameter vector. The intercept is β_0, and the coefficients for the k covariates are $\beta_1,\ldots.,\beta_k$.

In Poisson distribution, predictor variables are linked to the outcome via a natural log transformation, and this log transformation guarantees that the regression model predicted values are never negative. The general form of Poisson regression model to predict trip count is as follows

$$\log(y)=\beta0+\beta_1x_1+\beta_2x_2+\ldots+\beta_kx_k=x_i^T\beta \tag{5.3}$$

Where, y is the expected count of trips (mean) given a set of explanatory variables X = (X_1, X_2, X_k). In Poisson distribution, predictor variables are linked to the outcome via a natural log transformation, and this log transformation guarantees that the regression model predicted values are never negative.

5.2.2 Negative Binomial model

In the NB model, the probability of observing count outcome y_i conditional on the expected mean parameter λ and dispersion parameter $\theta>0$ is given by:

$$P(Y = y) = \left(\frac{\theta}{\theta+\lambda}\right)^{\theta} \times \frac{\Gamma(\theta+y)}{\Gamma(y+1)\Gamma(\theta)} \times \left(\frac{\lambda}{\theta+\lambda}\right)^{y} \tag{5.4}$$

Where Γ is the gamma function defined as follows:

$$\Gamma(t)=\begin{cases} \int_{x=0}^{\infty} x^{t-1}e^{-x}dx & \text{for positive non-integer t} \\ (t-1)! & \text{for positive integer t} \end{cases} \tag{5.5}$$

The variance of the NB model is $v=\lambda+\frac{\lambda^2}{\theta}$. Here, θ is an over-dispersion parameter and λ is the expected mean.

5.2.3 Zero-inflated models

Zero-inflated count models (assuming either the Poisson or NB distribution of the count outcome) provide a way of modeling the excess zeros in addition to allowing for over-dispersion (negative binomial) or without (Poisson distribution). In particular, for each observation, there are two possible data generation processes. For each observation, Process 1 is chosen with probability p_i and Process 2 with probability 1- p_i. Process 1 generates only zero counts, whereas Process 2, $P(Y_i=y_i|X_i)$, generates counts from either a Poisson or a NB model. In general:

$$y_i = \begin{cases} 0 & \text{with probability } p_i \\ P(Y_i = y_i|X_i) & \text{with probability } 1 - p_i \end{cases} \tag{5.6}$$

The probability $P(Y_i=y_i|X_i)$ depends on the process where it is zero-inflated Poisson (ZIP) or zero-inflated negative binomial (ZINB). Zero-inflated model consists of binary logit model and counts models. Binary logit model is commonly used to predict a behavior's occurrence, but with ZIP /ZINB, the logistic regression part of the model predicts non-occurrence (i.e., it predicts the zeros). The count models predict how frequently the behavior occurred.

The expected count is function of the two processes. In this study, the expected trip count is defined as follows:

$$E(y_i) = p_i * 0 + (1 - p_i) * e^{a_i} \tag{5.7}$$

p_i is the predicted probability that trip count is zero, e^{a_i} is the expected trip count given it is not zero and it is modeled using Poisson/NB regression.

The probability whether the trip is not possible (zero part), p_i is modeled by a logistic regression. Its form is:

$$log\left(\frac{p_i}{1-p_i}\right)=\beta_0+\beta_1X_1+\beta_2X_2+\dots \tag{5.8}$$

β_i is the parameter that will be estimated and X_i is the feature of the ZIP Code, such as population density, household income, and trip distance. e^{a_i} is modeled using Poisson/NB regression. Its form is:

$$e^{a_i} = e^{(\alpha_0+\alpha_1X_1+\alpha_2X_2+\ldots)} \quad (5.9)$$

α_i is the parameter that will be estimated and again, X_i is the feature of the ZIP Code.

5.2.4 Hurdle Models

In hurdle models, the count data generating process is controlled by Bernoulli probability that governs the binary outcome of whether a count variable has a zero or non-zero value. If the value is positive, the hurdle is crossed, and the conditional distribution of the non-zero outcome is governed by a Poisson/NB count data model. Hence Hurdle models can take shape of various count structures such as: Hurdle Poisson (HP) or Hurdle NB (HNB). In general, the hurdle model has two parts:

1. Zero count generating model.
2. Value (positive) generating model.

These two models are not considered to be the same. Hence, the difference from zero-inflated model is that the value generating part is not allowed to create zero outcomes. If the predicted variable $y_i>0$, the hurdle is crossed, the conditional distribution of the count value is governed by value generating model part. The zero-generating model can be considered as a logit model:

$$P(y_i = 0|X_i) = \frac{\exp(X_i'\gamma)}{1+\exp(X_i'\gamma)} = p_i \quad (5.10)$$

The value generating part of the model has conditional probability of count value given that the number is greater than zero. If we consider that the value generating model is Poisson model:

$$P(y_i = j|y_i > 0, X_i) = \frac{P(y_i=j \,\&\, y_i>0|X_i)}{P(y_i>0|X_i)} = \frac{\exp(X_i'\beta)\hat{}j\ \mathrm{e}^{-(X_j'\beta)}}{j![1-e^{-(X_i'\beta)}]}, j = 1,2, \ldots \quad (5.11)$$

So, the expected value of y_i is

$$E[y_i|X_i] = p_i * 0 + (1-p_i) * E[y_i|y_i > 0, X_i] \quad (5.12)$$

If there is over-dispersion, the estimate of the parameters from HP will be biased and inconsistent. In that case, the NB is a good substitute as a value generating model. For a HNB model, a dependent variable Y_i (i=1, 2, …, n) has the distribution

$$Pr(Y_i = y_i) = \begin{cases} p_i, & y_i = 0, \\ (1-p_i)\frac{\Gamma(y_i+\theta^{-1})}{\Gamma(y_i+1)\Gamma(\theta^{-1})}\frac{(1+\theta\lambda_i)^{-\theta^{-1}-y_i}\theta^{y_i}\lambda_i^{y_i}}{1-(1+\theta\lambda_i)^{-\theta^{-1}}}, & y_i > 0, \end{cases} \quad (5.13)$$

Where, θ (≥0) is dispersion parameter that is assumed not to be dependent on independent variables. p_i is a non-negative function that is modeled via logit link function,

$$logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \sum_{j=1}^{m} z_{ij}\delta_j \quad (5.14)$$

Where, $z_{ij}$is i-th row of covariate matrix Z and δ_j are unknown m-dimensional column vector of parameters.

5.2.5 Zero-inflated Negative Binomial Mixed-effects Model

Zero-inflated negative binomial mixed effect models (ZINB Mixed Effect) were developed to address over-dispersed count data with excess number of zeros (Fang et al., 2014; Moghimbeigi et al., 2008; Yau et al., 2003). This mixed model contains extra parameters to model the probability of excess zero values and the variability in non-zero values, allowing for repeated measures incorporating independent random effects for these two parts. ZINB Mixed Effect model can be expressed as follows:

$$\log(\lambda_{ij}) = X_{ij}\beta + a_i \tag{5.15}$$

$$logit(p_{ij}) = Z_{ij}\gamma + b_i \tag{5.16}$$

Here, X_{ij} represents the matrix of covariates and β is their respective regression coefficient for the negative binomial part, Z_{ij} represents the covariate matrix and the respective vector of regression coefficient γ for the logistic part, a_i and b_i are the random intercepts and they follow normal distribution with mean zero. For simplicity those intercepts are assumed to be independent. This assumption is also used in the literature of ZINB/ZIP with random effects (Fang et al., 2014; Hur et al., 2002; Yau and Lee, 2001).

5.3 Model Estimation Results

A comparison of the estimation results of seven count data models: Poisson (Model 1), NB (Model 2), ZIP (Model 3), ZINB (Model 4), HP (Model 5), HNB (Model 6) and ZINB Mixed Effect (Model 7) is presented in Table 8. The statistically significant explanatory variables along with their estimated coefficients and t-statistics (in parenthesis) for each of the developed models are shown in Table 8. Only ZI and hurdle models have estimates of the parameters for zero counts. The log-likelihood value at convergence, the Bayesian Information Criterion value (BIC), and the total number of observations are also included for each model. Poisson regression is one of the most basic count regression models. The explicit assumption used for Poisson model is that the mean and variance of count variable are statistically equal. Given that there is no a priori reason for the mean and variance in any practical context to be equal, the use of a NB distribution for Model 2, 4, 6, and 7 is an important empirical generalization over the Poisson distribution. The NB model is considered as a generalization of Poisson model since it has the same mean structure as Poisson regression and it has an extra parameter ("θ ") to model over-dispersion. If the conditional distribution of the outcome variable is over-dispersed, the confidence intervals for the NB regression are likely to be narrower as compared to that of a Poisson model. In the NB model, the dispersion parameter properly captures the difference between mean and variance. However, the NB model needs to be further examined to model DRT trip frequency due to the presence of excessive zeros in this dataset. Zero-inflated models (Model 3 through 7) accounts for presence of excess zeros in the trip frequency. The distribution of dependent variable (Trips) is extremely skewed because of excess number of zero trip occurrences (97.78% of origin-destination pairs) in trip count data. Zero inflated and hurdle models are good candidates for this data which can address over-dispersion for the excess zeroes problem effectively. ZIP model has been developed for this dataset. The zero-estimation part is a binary model which examines if the trips ever occurred by using a logistic regression. The second model is the normal Poisson model (value estimation part) that predicts the frequency of the trip if that is non-zero. This model was able to estimate excess amount of zero but failed to capture variability due to dispersion. However, this model gave us a set of significant independent variables along with a good starting estimate value for ZINB and hurdle models. The difference between mean and variance is still high even if all zero trips occurrences are not taken into consideration which has standard deviation (678.03) that is much higher than mean (73.51). The results from ZINB model demonstrate that we can indeed reject the hypothesis that the trip generation process is Poisson, since log ( ) =-2.470 with p-value < 0.0001, and thus the variance of the process is much larger than the mean. The estimate of significant positive intercept in logistic model part proves that there is excess number of zeros in the data.

Hurdle models including HP and HNB have also been developed to accommodate over dispersion and excess zeroes problem. Hurdle model in case of zero estimation is different from zero-inflated models. The sign of estimated parameters in the hurdle model is not opposite for value-estimation and zero-estimation parts because these two processes are independent and likely to follow similar effect over trip count. In addition, to achieve inter-ZIP Code trip variability, which is not captured well by covariates, the origin and destination ZIP Codes are introduced in the zero-estimation part of the model as random effect parameters. The estimated standard deviations of those random effects are significantly large. Another random effect variable incorporated in the value estimate part of the ZINB Mixed Effect model is, the rural urban commuting area (RUCA), which indicates the type of ZIP Code area based on the size and direction of the primary commuting flows. The estimated standard deviation of this random effect variable indicates that the trip count has variability across different types of ZIP Codes. The estimated standard deviations (σ) of the random effects are presented in Table 8.

The variables that have significant effect on DRT trip frequency includes origin ZIP Code population density, distance between two ZIP Codes, population aged 14 years or less, population aged 65 years or over, the number of disabled people, household median income , homeowner vacancy rate , average household size, the number of renter occupied housing unit, dominant sex (male), proportion of white people, the number of wholesale trade establishments in destination Zip and the number of retail trade establishments in destination ZIP Code. The estimated parameter signs are similar across the models which means the effect of variables are consistent. The results indicate that lower population density is likely to increase the overall trip count. The similar relationship between this variable and DRT demand is also found in the demand model developed by Wang et al.(2014). This is intuitive because of unavailability of demand response service in an urban area where the population density is higher and lower the density means the ZIP Code area is in rural area. It is more likely to have fixed route public transportation services in an urban area. Moreover, people living in higher population density areas can coordinate with others to make a trip. The distance between ZIP Codes has an opposite effect over trip count. The results indicate that with increasing distance the likelihood of occurrence of a DRT trip decreases. This is intuitive because DRT serves trips that are relatively short and not supporting inter-city type services that tend to cover long distance. Trip count is likely to decrease with the increase of younger population (age 14 or less) in the origin ZIP Code area. The presence of children of less than 14 years reduces DRT trips as parents are typically not elderly and may own a car in such households. On the other hand, older age group (age 65 or over) population has opposite effect on trip count because they likely rely on DRT for medical services and increasingly, the baby boomer generation is “aging in place”. Moreover, the aged population might not own a car or be unable to drive. The similar effect observed from disabled population in the origin ZIP Code where the trip count increases with the disabled population size increases. This result is consistent with the research conducted by Mattson (2017). The disabled population tends to be most captive to transit services and may need additional medical services. The variable Household median income has a negative impact on trip count because people like to get their own vehicle when they have higher income level. This finding coincides with the research conducted by Yang and Cherry (2017). The Homeowner vacancy rate in origin ZIP Code are likely to increase DRT trip in the sense that we have higher homeowner vacancy rate in rural area. The Average household size is also likely to increase the trip count. The number of Renter occupied housing unit in origin ZIP Code has positive impact over trip count. This is because the renter occupied people in rural area is less likely to own and operate a vehicle. The variable Sex indicates that women are the primary user of DRT service. If the Origin ZIP Code with higher number of female compared to male, it is more likely to induce demand for DRT trip. This similar relation is also observed in the DRT and DFRT study of Yang and Cherry (2017). In case of variable Race, white people are most likely to use DRT service in rural areas. This finding coincides with the research conducted by Wang et al.(2014).When destinations are based on Retail trade, they are likely to attract more DRT trips as population from neighboring areas will likely to make trips for retail goods. However, Whole sale trade shows an inverse relationship with DRT frequency.

In zero estimation part, the parameters for zero estimation indicate which variables will predominantly describe likelihood of DRT to be zero. Especially Model 3 through 7 have zero estimation parameters. For zero-inflated models, the sign of the estimated parameters for zero estimation part is opposite to the sign of the estimated parameters for value estimation part. Also, the sign of the variables remains consistent across all zero-inflated models. The variable in zero estimation part like *Origin ZIP Code Average Household size* has positive regression coefficient implies that the probability of zero DRT trip increases with the increase of average household size. The variable *Origin ZIP Code Population age up to 14 years* increases the zero occurrence of trip count whereas the variable *Origin ZIP Code Population age 65 and over years* decrease the probability of DRP trip count being zero. The variable distance has greater impact on zero trip count probability which increase with the increase of distance. But the sign of estimated parameters is not opposite in the hurdle model for two processes like zero inflated models because these two parts are independent in hurdle model. Variables indicating zero DRT trips include Destination ZIP Code Retail Trade related businesses, Log (Origin ZIP Code Household Median income), Log (Origin ZIP Code Population age up to 14 years), Log (Origin ZIP Code Population age 65 and over years), Log (Origin ZIP Code Disable Population), and Distance between Origin and Destination ZIP Code.

TABLE 8: Model estimation results

Variables	**Poisson (Model 1)**	**NB (Model 2)**	**ZIP (Model 3)**	**ZINB (Model 4)**	**HP (Model 5)**	**HNB (Model 6)**	**ZINB Mixed Effect (Model 7)**
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Value estimation part							
Intercept	6.003 (58.785)	-11.013 (-7.76)	2.486 (25.980)	1.142 (2.725)	-1.169 (-50.77)	-9.003 (-0.612)	0.512 (.223)
Distance between two ZIP Code	-0.112 (-939.448)		-0.090 (-762.450)	-0.031 (-35.488)	-0.087 (-725.68)	-0.031 (-32.717)	-0.030 (-39.610)
Origin ZIP Code Population density	-1.192 (-524.096)						
Log (Origin ZIP Code Population density)			-1.043 (-466.470)	-0.652 (-18.624)	-1.223 (-477.01)	-0.522 (-11.407)	-0.403 (-12.300)
Log (Origin ZIP Code Population aged 14 years or less)		-.665 (-7.674)		-0.387 (-5.784)	-0.215 (-30.940)	-0.329 (-3.936)	-0.346 (-5.960)
Proportion of white people in Origin ZIP Code		1.717 (8.156)					
Log (Origin ZIP Code Population aged 65 years or over)	0.317 (37.179)	0.713 (6.516)					
Log (Origin ZIP Code Household Median income)	-1.150 (-116.075)	-0.386 (-2.889)	-0.602 (-65.750)				
Destination ZIP Code Retail Trade related businesses	0.009 (459.281)						

Log (Destination ZIP Code Retail Trade related businesses)		1.439 (32.024)	0.727 (385.870)	0.731 (17.065)	0.823 (279.04)	0.843 (15.862)	0.562 (13.340)
Log (Destination ZIP Code Whole Sale Trade related businesses)		-0.698 (-13.724)		-0.336 (-6.533)	-0.180 (-63.780)	-0.415 (-6.160)	-0.004 (-.080)
Origin ZIP Code Homeowner vacancy rate	0.187 (185.638)		0.161 (165.060)				
Origin ZIP Code Average Household size	0.016 (2.134)		0.668 (113.060)				
Log (Origin ZIP Code Disabled population)	1.551 (172.618)	1.423 (17.111)	1.424 (491.550)	0.205 (2.441)	1.069 (151.3)	0.336 (3.065)	0.460 (5.720)
Log (Origin ZIP Code Renter occupied housing unit)				0.806 (10.546)	0.636 (96.42)	0.638 (6.416)	0.502 (7.230)
Origin ZIP Code Dominant Sex (male)	-0.344 (-68.014)						
Origin ZIP Code Dominant Race (white)	2.814 (114.842)						
Log (θ)		-5.149 (-50.902)		-2.47 (-63.796)		-11.690 (-0.796)	-1.437 (-46.477)
Random effects parameters:							
σ (RUCA type of Origin ZIP Code)							0.00013

σ (RUCA type of Destination ZIP Code)							0.502
Zero estimation part							
Intercept			4.592 (4.729)	4.841 (18.334)	-4.598 (-6.502)	-2.065 (-3.088)	9.662 (10.060)
Origin ZIP Code Average Household size			0.533 (7.369)	0.205 (5.513)	-0.297 (-4.528)	-0.236 (-4.503)	0.438 (3.060)
Destination ZIP Code Retail Trade related businesses			-0.284 (-19.866)				
Log (Destination ZIP Code Retail Trade related businesses)				-0.668 (-25.879)			-0.954 (-11.830)
Log (Origin ZIP Code Household Median income)			-0.217 (-2.321)		0.153 (2.375)	-0.197 (-3.128)	
Log (Origin ZIP Code Population aged 14 years or less)			0.522 (10.110)	0.343 (3.767)	-0.320 (-8.459)	-0.348 (-8.988)	1.143 (3.990)
Log (Origin ZIP Code Population aged 65 years or over)				-0.311 (-3.129)		0.530 (9.129)	-1.440 (-3.950)
Log (Origin ZIP Code Disabled Population)					0.274 (5.166)	0.196 (3.492)	
Log (Origin ZIP Code Renter occupied housing unit)			-0.781 (-14.048)	-0.730 (-13.171)	0.733 (15.619)	0.211 (6.857)	-0.656 (-3.170)

Origin ZIP Code Number of Households with income 200K or more			-0.001 (-5.404)				
Log (Origin ZIP Code Population density)			0.059 (2.717)		-0.289 (-13.949)		
Distance between two ZIP Codes				0.039 (37.368)	-0.034 (-59.382	-0.033 (-59.506)	0.053 (32.240)
Random effects parameters:							
σ (Origin ZIP Code)							2.636
σ (Destination ZIP Code)							2.509
Measures of fit							
Log-Likelihood at convergence	-613,048	-27,933	-36,8191	-22,421	-376413	-23,222	-19,585
BIC	1,226,228	55,973	736,573	45,033	753,015	46,648	39,408
Number of observations	148,454	148,454	148,454	148,454	148,454	148,454	148,454
Number of parameters estimated	11	9	16	16	16	17	20

5.4 Model Selection and Statistical Fit

Several criteria can be used to select the best performing model among non-nested models. Here, two goodness-of-fit indices were used to evaluate the fitness of the model, log-likelihood, and BIC. Goodness-of-fit indices for the seven models are shown in Table 8. The log-likelihood value of ZINB Mixed Effect model is lowest. To facilitate comparison across different models estimated in this study, BIC value was computed as: $-2 \times LL + K \times LN(N)$, where K is the number of model parameters and N is the number of observations in the estimation sample. The BIC statistic penalizes model that attain higher LL values using more parameters to the estimated model (Akaike, 1987; Schwarz, 1978). According to the BIC criterion, a model with lower BIC value is preferred over model with higher BIC value. It can be seen from the table that the ZINB Mixed Effect model had the lowest BIC value among all models for the dataset used in the analysis. Among all the models considered, the ZINB Mixed Effect model with spatial effects has the highest LL value and the least BIC value suggesting superior data fit.

5.5 Elasticity Effects

The parameter estimates in the count models (shown in Table 8) do not directly indicate the magnitude of impact of different independent variables on expected DRT trip frequency. In order to determine the magnitude of effects of the different independent variables on DRT trip frequency, it is necessary to compute their corresponding elasticity effects. The elasticity effect represents the percentage change in the response variable due to a unit percentage change in an explanatory variable (Castro et al., 2012). Table 9 presents the elasticity effect of the best performing ZINB Mixed Effect model. From the Table 10, it can be observed that the elasticity effects are consistent with the coefficient estimates of the model variables. The elasticity parameter of population density indicates that doubling the log of population density in the origin ZIP Code will cause the expected trip counts to be decreased by 0.716%, on average if everything else remains the same. The highest elasticity effect was observed on distance variable. It indicates that, on average if everything else remains the same, the trip generation will be decreased by 2.026% with one unit increase of the distance between origin and destination ZIP Code. Other elasticity values in the table can be interpreted similarly.

TABLE 9: Elasticity effects of the ZINB mixed effect model.

Variables	**ZINB Mixed Effect**
Value estimation part	
Log (Origin ZIP Code Population density)	-0.716
Log (Destination ZIP Code Retail Trade related businesses)	0.562
Log (Destination ZIP Code Whole Sale Trade related businesses)	-0.0024
Distance between two ZIP Code	-2.026
Log (Origin ZIP Code Population aged 14 years or less)	-0.869
Log (Origin ZIP Code Disabled population)	1.154
Log (Origin ZIP Code Renter occupied housing unit)	1.127
Zero estimation part	
Log (Origin ZIP Code Population aged 65 years or over)	-3.524
Log (Origin ZIP Code Renter occupied housing unit)	-1.437
Origin ZIP Code Average Household size	0.439
Log (Origin ZIP Code Population aged 14 years or less)	2.871
Distance between two ZIP Code	3.585
Log (Destination ZIP Code Retail Trade related businesses)	-0.955

Figure 15 shows the effects of socio-economic variables on the frequency of DRT trip. The vertical axis shows percentage changes in frequency of DRT trip. The effects of 10, 20, and 30% increase in retail trade, the number of disabled people, population aged 14 years or less, the number of renter occupied housing unit and population density on DRT trip count are shown. As expected, the figure shows that the DRT trip frequency decreases with the population density and population aged 14 years or less increases. On the other hand, DRT trip frequency increases with the increased number of disabled people, and population aged 14 years or less increases. Number of renter occupied housing unit, retail trade related establishment which are clearly visible from the figure 15.

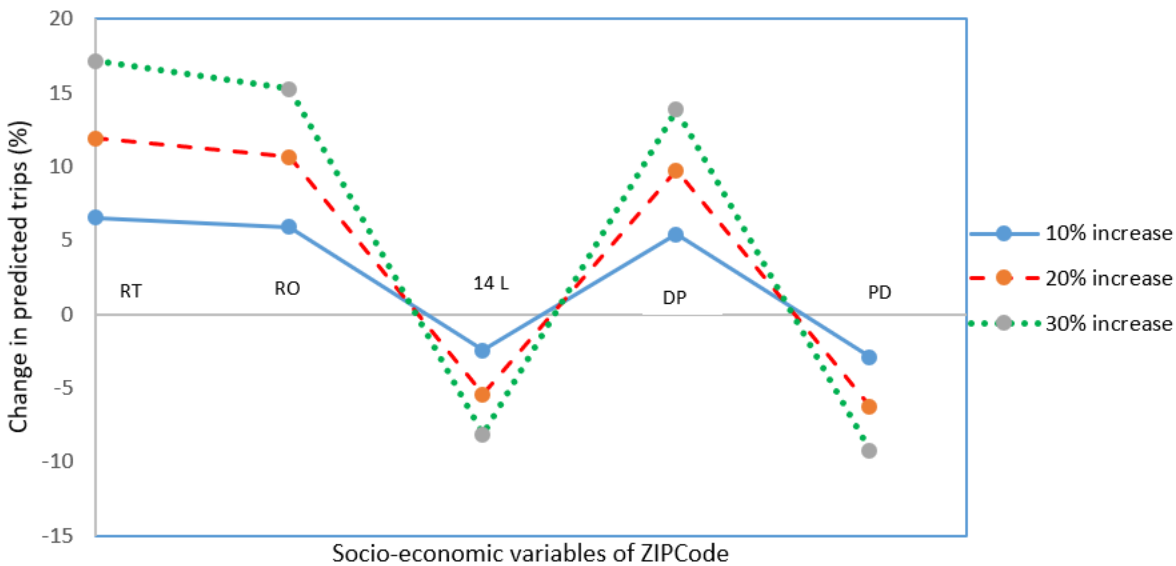

Figure 15 : Changes in predicted trips based on Independent variables; RT, RO, 14L, DP and PD mean Retail Trade, Renter occupied, Population aged 14 years or less, Disabled population and Population density respectively.

5.6 Model Validation

To test the predictive power of these models, a validation exercise was undertaken in which the predicted demand trip counts were compared with the observed counts in the data (Table 10). The dataset is divided into training set (80%) and test set (20%) by randomly taking data points. Absolute Percentage Difference (APD) between predicted and observed shares for each count outcome was computed. Next, Average Absolute Percentage Difference (AAPD) across all count outcomes was computed and used as a metric of predictive performance. Models with lower AAPD value are preferred over models with higher AAPD values. Table 10 represents the result of the prediction analyses. It is observed from the table that ZINB Mixed Effect model better suited to capture dispersion in count data among all models for DRT trips in rural areas.

TABLE 10: Model validation based on AAPD.

Trips	Observed Count	Expected Count													
		Poisson		NB		ZIP		HP		HNB		ZINB		ZINB Mixed Effect	
		Count	APD (%)	Count	APD (%)	Count	APD (%)	Count	APD (%)	Count	APD (%)	Count	APD (%)	Count	APD (%)
0	36,250	34,743	4.15	25,625	29.31	34,991	3.47	35,158	3.01	33,378	7.92	33,221	8.35	34,829	3.92
1-10	511	143	72.01	219	57.14	163	68.10	132	74.16	259	49.32	241	52.83	187	63.41
11-100	202	54	59.00	59	61	61	69.80	55	72.77	84	58.41	109	46.04	158	21.78
>100	83	24	71.08	0	100	10	87.95	20	75.90	4	93.38	4	95.18	24	71.08
AAPD (%)			52.13		64.31		57.33		56.46		52.26		50.60		40.05

CHAPTER 6: DISCUSSION AND CONCLUSIONS

6.1 Urban connectivity

Because of budget constraints most suburban transit agencies do not have the capacity to (1) routinely collect transit ridership, boarding, alighting, and (2) maintain a comprehensive and well-designed transit assignment module in a travel demand model or from an advanced transit system where smart cards are used to keep track of transit demand performance. Transit connectivity is a multidimensional problem involving various service quality factors that include walking distance, in-vehicle travel time, waiting time, number of destinations served and number of transfers to reach destinations. Further adding to this complexity is the (usually) high number of available routes with distinct characteristics within a network. Based on network graph properties this paper proposes connectivity indices at stop, route, and zonal level by considering various factors such as speed, frequency, operational capacity, fare, route origins and destination, and urban form characteristics that serves the transit system. The connectivity indices are applied to three metropolitan cities (Knoxville, Memphis, and Nashville) of Tennessee by using open access GTFS data. The models and data processes developed in this paper can be used to (i) determine the performance transit system with no additional data purchase, (ii) use of transit performance measures along with other data sources (such as vehicle ownership, income etc.) to assess future service needs, (iii) use of geographic information systems capabilities to disseminate transit performance measures for potential future users, and to further induce demand, and (iv) seamless re-estimation of transit performance measures both in alternate dimensions of time and space.

Public Transit is a critical component of TDOT's Long-Range Transportation Plan. The demographic data and trends in the state of Tennessee point to a potential increase in need for public transit services in cities and rural communities. The role of TDOT in providing the mobility and accessibility options to the residents, especially captive riders, is critical for the future quality of life and economic competitiveness of Tennessee. Also, the trend of increased percentages of household income spent on transportation and increased commuting distances are going to be major contributing factors behind increased transit demand in the future. This research is crucial in identifying areas in needs of service, developing a methodology to address the accessibility and mobility issues and formulating a cost-effective plan to provide transit services. The results will serve as components of Tennessee statewide mobility report. The demand model developed in TCRP Report 161 includes some service characteristics, such as size of service area and service miles, but it lacks other service characteristics. This study came up with a better demand-response trip predictive model which considers few other factors like transportation-disadvantaged groups of societies play an important role in demand ridership. Performance of this model can be further improved if we add Land use data to capture real picture of trip attraction in the locality. The label of ZIP Code whether that belongs to urban or rural area can be another good predictor for zero inflated part of the ZINB model. Other factors such as fares span of service, reservation requirements, and other service characteristics of DRT providers will likely impact ridership.

Further, the transit connectivity measure is used to determine equity by various socio-economic factors such as household income, vehicle ownership, employment, and population. Transit connectivity equity is estimated by the GINI index. All three cities have both strengths and weaknesses in serving the captive riders when various socioeconomic factors are considered. For instance, Knoxville provides more equitable transit service when household income is considered, while Memphis based on population, and Nashville based on vehicle ownership. This paper presents results from a broader study that aims to provide TDOT with valuable information on captive ridership in urban, suburban, and rural areas, and identify transit needs in these areas.

Methods and results presented in this study can provide input to a base framework for state DOTs to maintain a five-year transit plans, as well for identifying changing service impacts in correlation with changing demographics in order to assess the transportation needs of metropolitan and local communities. However, the generalized data set and its integration with the proposed model can be adopted by any public agency to assess connectivity and equity.

6.2 Rural demand-responsive transit

The primary objective of this research was to develop a set of econometric models that can predict DRT trip frequency as a function of land use, socio-economic and demographic characteristics. We test these models on DRT trip data for rural areas in the state of Tennessee. To be specific, seven count data models; Poisson, Negative Binomial (NB), Zero-inflated Poisson (ZIP), Zero-inflated Negative Binomial (ZINB), Hurdle Poisson (HP), Hurdle Negative Binomial (HNB), and ZINB Mixed Effect were developed to determine the causal factors related to DRT trips. BIC and Log-likelihood was computed to compare different models. In addition, the predicted number of DRT trips was used for model validation. The ZINB Mixed Effect model performed better compared to all other models on model fit statistics and on the validation exercise. The results of statistical models revealed that the significant contributing factors that lead to DRT trip frequency are: trip distance, population density, population aged 14 years or less, population aged 65 years or over, average household size, average income, retail and wholesale trade related establishments and others. The elasticity effects of all variables entered ZINB Mixed Effect model were also computed to understand clearly the impacts of those variables. The analyses of the elasticity effect revealed that the variables with the largest effect were trip distance, population aged 65 years or over, disabled population etc.

In terms of future research, characteristics of DRT service providers should be taken into consideration while developing models for better prediction of DRT trip frequency. These characteristics (i.e. reservation requirements, fare, days of operation per week etc.) may impact the trip count in their serving area. Inclusion of theses service characteristics information with demographic and land use data of ZIP Codes should provide better predictive outcome. In addition, if more attributes of the trip makers were available (Yang and Cherry, 2017), the models could have developed at a finer geographic level or even at individual level rather than ZIP Codes. The models can be strengthened if time-of-day travel information is available to predict DRT trips by various times of the day.

REFERENCES

ActiveAge, 2008. An introduction to Demand Responsive Transport as a Mobility Solution in an Ageing Society. Retrieved from http://www.activeage.org/publications/doc_download/13-An-introduction-to-demand-responsive-transport-DRT (accessed 8.21.15)

Ahmed, A., Dwyer, T., Forster, M., Fu, X., Ho, J., Hong, S.-H., … Taib, R. (2005). Geomi: Geometry for maximum insight (pp. 468–479). Presented at the International Symposium on Graph Drawing, Springer.

Akaike, H., 1987. Factor analysis and AIC. *Psychometrika* 52, 317–332. doi: 10.1007/BF02294359.

Ambrosino, G., Nelson, J.D., Romanazzo, M., 2004. Demand Responsive Transport Services: Towards the Flexible Mobility Agency, Italian National Agency for New Technologies. Energy and the Environment, Rom.

Bakker, P., 1999. Large scale demand responsive transit systems - A local suburban transport solution for the next millennium? Working Paper, AVV Transport Research Centre, Dutch Ministry of Transport, Public Works and Water Management, Rotterdam, Netherlands

Bell, D. C., Atkinson, J. S., & Carlson, J. W. (1999). Centrality measures for disease transmission networks. *Social Networks*, *21*(1), 1–21.

Bills, T., Sall, E., & Walker, J. (2012). Activity-based travel models and transportation equity analysis: Research directions and exploration of model performance. *Transportation Research Record: Journal of the Transportation Research Board*, (2320), 18–27.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564.

Bonacich, P., & Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, *23*(3), 191–201.

Braun, M., Winter, S., 2009. Ad hoc solution of the multicommodity-flow-over-time problem. *IEEE Transactions on Intelligent Transportation Systems* 10(4), 658–667.doi: 10.1109/TITS.2009.2026443.

Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation Research Part B: Methodological* 46, 253–272. doi: 10.1016/j.trb.2011.09.007.

Carrington, P. J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis* (Vol. 28). Cambridge university press.

Chakraborty, A., & Mishra, S. (2013). Land use and transit ridership connections: Implications for state-level planning agencies. *Land Use Policy*, *30*(1), 458–469.

Dajani, J. S., & Gilbert, G. (1978). Measuring the performance of transit systems. *Transportation Planning and Technology*, *4*(2), 97–103. https://doi.org/10.1080/03081067808717125

Davison, L., Enoch, M., Ryley, T., Quddus, M., Wang, C., 2014. *A survey of demand responsive transport in Great Britain. Transport Policy* 31, 47–54. doi: 10.1016/j.tranpol.2013.11.004

Enoch, M.P., Ison, S.G., Laws, R., Zhang, L., 2006. Evaluation study of demand responsive transport services in Wiltshire. Final report. Wiltshire County Council, Loughborough University, Loughborough.

Enoch, M., Potter, S., Parkhurst, G., Smith, M., 2006. Why do demand responsive transport systems fail? Presented at the Transportation Research Board 85th Annual Meeting.

Estrada, E., & Rodriguez-Velazquez, J. A. (2005). Subgraph centrality in complex networks. *Physical Review E*, *71*(5), 056103.

Fang, R., Wagner, B., Harris, J.K., Fillon, S.A., 2014. Application of zero-inflated negative binomial mixed model to human microbiota sequence data (No. e215v1). PeerJ PrePrints.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Garroway, C. J., Bowman, J., Carr, D., & Wilson, P. J. (2008). Applications of graph theory to landscape genetics. *Evolutionary Applications*, *1*(4), 620–630.

Goodwill, J.A., Joslin, A., 2013. Forecasting Paratransit Services Demand–Review and Recommendations. National Center for Transit Research, University of South Florida.

Grengs, J. (2002). Community-based planning as a source of political change: The transit equity movement of Los Angeles' Bus Riders Union. *Journal of the American Planning Association*, *68*(2), 165–178.

Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. A. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(22), 7794–7799.

Gurmu, S. (1998). Generalized hurdle count data regression models. *Economics Letters 58*. 263–268. doi: 10.1016/S0165-1765(97)00295-4.

Hu, M.C., Pavlicova, M., Nunes, E.V. (2011). Zero-Inflated and Hurdle Models of Count Data with Extra Zeros: examples from an HIV-Risk Reduction Intervention Trial. *The American Journal of Drug and Alcohol Abuse 37*, 367–375. doi:10.3109/00952990.2011.597280.

Hur, K., Hedeker, D., Henderson, W., Khuri, S., Daley, J. (2002). Modeling clustered count data with excess zeros in health care outcomes research. *Health Services and Outcomes Research Methodology 3*, 5–20. doi: 10.1023/A:1021594923546.

Jain, S., Ronald, N., Thompson, R., Winter, S. (2017). Predicting susceptibility to use demand responsive transport using demographic and trip characteristics of the population. *Travel Behaviour and Society 6*, 44–56.
doi: 10.1016/j.tbs.2016.06.001.

Joshi, N. N., & Lambert, J. H. (2007). Equity metrics with risk, performance, and cost objectives for the prioritization of transportation projects. *IEEE Transactions on Engineering Management*, *54*(3), 539–547.

Junker, B. H., Koschützki, D., & Schreiber, F. (2006). Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics*, *7*(1), 219.

Kattiyapornpong, U., Miller, K.E. (2006). Understanding travel behavior using demographic and socioeconomic variables as travel constraints. Presented at the Australian & New Zealand Marketing Academy Conference, Queensland University of Technology, School of Advertising, Marketing and Public Relations.

Lam, T. N., & Schuler, H. J. (1982). Connectivity index for systemwide transit route and schedule performance. *Transportation Research Record*, *854*(854), 17–23.

LaMondia, J.J., Bhat, C.R. (2010). Development of a paratransit microsimulation patron accessibility analysis tool for small and medium sized communities. *Transportation Research Record: Journal of the Transportation Research Board 2174*, 29–38.

Lerman, S.R., Flusberg, M., Pecknold, W.M., Nestle, R.E., Wilson, N.H. (1980). A model system for forecasting patronage on demand responsive transportation systems. *Transportation Research Part A: General 14*(1), 13–23.
doi: 10.1016/0191-2607(80)90090-4.

Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing & Management*, *41*(6), 1462–1480.

Maddern, C., Jenner, D. (2007). Telebus mobility and accessibility benefits: final report. Presented at the 12th International Conference on Mobility and Transport for Elderly and Disabled transport (TRANSED), Hong Kong.

Martınez, A., Dimitriadis, Y., Rubia, B., Gómez, E., & De La Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education*, *41*(4), 353–368.

Mattson, J., 2017. Estimating Ridership of Rural Demand-Response Transit Services for the General Public. Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C.

Mishra, S., Welch, T. F., & Jha, M. K. (2012). Performance indicators for public transit connectivity in multi-modal transportation networks. *Transportation Research Part A: Policy and Practice*, *46*(7), 1066–1085.

Mishra, S., Welch, T. F., Torrens, P. M., Fu, C., Zhu, H., & Knaap, E. (2015). A tool for measuring and visualizing connectivity of transit stop, route and transfer center in a multimodal transportation network. *Public Transport*, *7*(1), 77–99.

Modarres, A. (2003). Polycentricity and transit service. *Transportation Research Part A: Policy and Practice*, *37*(10), 841–864. https://doi.org/10.1016/S0965-8564(03)00059-4

Moghimbeigi, A., Eshraghian, M.R., Mohammad, K., Mcardle, B. (2008). Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros. Journal of Applied Statistics *35*(10), 1193–1202.

Moore, S., Eng, E., & Daniel, M. (2003). International NGOs and the role of network centrality in humanitarian aid operations: a case study of coordination during the 2000 Mozambique floods. *Disasters*, *27*(4), 305–318.

Newman, M. E. (2004). Analysis of weighted networks. *Physical Review E*, *70*(5), 056131.

Nguyen-Hoang, P., Yeung, R. (2010). What is paratransit worth? *Transportation research part A: policy and practice 44*, 841–853. doi: 10.1016/j.tra.2010.08.006.

Park, J.-S., & Kang, S.-C. (2010). A model for evaluating the connectivity of multimodal transit networks. *Journal of Korean Society of Transportation*, *28*(3), 85–98.

Piatkowski, D.P., Marshall, W.E. (2015). Not all prospective bicyclists are created equal: The role of attitudes, socio-demographics, and the built environment in bicycle commuting. *Travel behaviour and society 2*, 166–173.
doi: 10.1016/j.tbs.2015.02.001.

RSG. (2015). *User Guide: STOPS: Simplified Trips-on-Project Software, Version 1.50, Prepared for FTA*.

Ruhnau, B. (2000). Eigenvector-centrality—a node-centrality? *Social Networks*, *22*(4), 357–365.

Sarker, A. A., Mishra, S., Welch, T. F., Golias, M. M., & Torrens, P. M. (2015). A Model Framework for Analyzing Public Transit Connectivity and Its Application in a Large-scale Multi-modal Transit Network. Presented at the Transportation Research Board 94th Annual Meeting.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics 6*, 461–464. doi: 10.1214/aos/1176344136.

TCRP (1995). Workbook for Estimating Demand for Rural Passenger Transportation. Synthesis Report 3, Transit Cooperative Research Program (TCRP). Federal
Transit Administration, Washington, DC.

TCRP (2004). Demand Responsive/ADA Traveler Response to Transportation System Changes, Synthesis Report 95, Transit Cooperative Research Program (TCRP). Federal Transit Administration, Washington, DC.

T.D.O.T. (2015). Public Transit, Transportation Demand Management, and Non-Motorized Modes,. In *Mobility Policy Paper*.

Vanasse Hangen Bruslin, Inc., LSC Transportation Consultants, Inc., and Erickson Consulting, LLC. (2013). Methods for Forecasting Demand and Quantifying Need for Rural Passenger Transportation: Final Workbook. TCRP Report 161, Transit Cooperative Research Program, Washington, DC: Transportation Research Board

Wang, C., Quddus, M., Enoch, M., Ryley, T., Davison, L. (2014). Multilevel modelling of Demand Responsive Transport (DRT) trips in Greater Manchester based on area-wide socio-economic data. *Transportation 41*(3), 589–610.
doi: 10.1007/s11116-013-9506-1.

Wang, H., Winter, S. (2010). Utilizing taxi empty cruise time to solve the short distance trip problem. Presented at 17th ITS World Congress, Busan.

Washington, S.P., Karlaftis, M.G., Mannering, F. (2010). Statistical and econometric methods for transportation data analysis. CRC press, NewYork.

Welch, T. F., & Mishra, S. (2013). A measure of equity for public transit connectivity. *Journal of Transport Geography*, *33*, 29–41.

Welch, T. F., & Mishra, S. (2013). A measure of equity for public transit connectivity. *Journal of Transport Geography*, *33*, 29–41.

Yang, H., Cherry, C. R., Zaretzki, R., & Ryerson, M. (2016). Rural to urban Intercity Deviated Fixed Rout Design - an Application in Tennessee,

Yang, H. (2013). Rural to Urban Intercity Transit User Characteristics Analysis, Demand Estimation and Network Design.

Yang, H., & Cherry, C. R. (2012). Statewide rural-urban bus travel demand and network evaluation: An application in Tennessee. *Journal of Public Transportation*, *15*(3), 6.

Yang, H., Cherry, C.R. (2017). Use characteristics and demographics of rural transit riders: a case study in Tennessee. *Transportation Planning and Technology 40*, 213–227. doi: 10.1080/03081060.2016.1266168.

Yang, H., Cherry, C.R., Zaretzki, R., Ryerson, M.S., Liu, X., Fu, Z. (2016). A GIS-based method to identify cost-effective routes for rural deviated fixed route transit. *Journal of Advanced Transportation 50*, 1770-1784. doi: 10.1002/atr.1428.

Yau, K.K., Lee, A.H. (2001). Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. *Statistics in medicine 20*, 2907–2920. doi: 10.1002/sim.860.

Yau, K.K., Wang, K., Lee, A.H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal 45*, 437–452. doi: 10.1002/bimj.200390024.

APPENDIX-A: GTFS DATA STRUCTURE

A typical GTFS data set will include the following files.

1. Agency.txt
2. Calendar.txt
3. Feed_info.txt
4. Routes.txt
5. Shapes.txt
6. Stop_times.txt
7. Stops.txt
8. Trips.txt

For this project, we have utilized GTFS data for three cities and the data sets can be downloaded using following links.

1. GTFS Data for Knoxville: Download
2. GTFS Data for Memphis: Download
3. GTFS Data for Nashville: Download

The most updated datasets are available on the following website.

http://www.gtfs-data-exchange.com/

APPENDIX-B: KNOXVILLE GTFS DATA STRUCTURE and MAPS

A. Agency.txt

agency_id, agency_name, agency_url,
agency_timezone,agency_lang,agency_phone,agency_fare_url

1, Knoxville Area Transit
http://www.katbus.com/,America/New_York,en,865.637.3000,http://www.katbus.com/

B. Calendar.txt

service_id	monday	tuesday	wednesday	thursday	friday	saturday	sunday	start_date	end_date
1	1	1	1	1	1	0	0	20160111	20160531
2	0	0	0	0	0	1	0	20160111	20160531
3	0	0	0	0	0	0	1	20160111	20160531

C. Feed_info.txt

feed_publisher_name, feed_publisher_url, feed_lang,feed_start_date,feed_end_date,feed_version

Knoxville Area Transit, http://www.katbus.com/,en,20160111,20160531,2016 1-11 NO ARRV St_20160405

D. Routes.txt

route_id	agency_id	route_short_name	route_long_name	route_desc	route_type	route_url	route_color	route_text_color
2803	1	10	Sequoyah Hills		3		2673	ffffff
2804	1	11	Kingston Pike		3		FF0000	ffffff
2805	1	12	Western Avenue		3		339900	ffffff
2806	1	13	Beaumont		3		FFBB33	0
2807	1	16	Cedar Bluff Connector		3		2673	ffffff
2808	1	17	Sutherland/Bearden		3		73B2FF	0
2809	1	19	Lakeshore/Lonas Connector		3		339900	ffffff
2810	1	20	Central Avenue		3		FF0000	ffffff
2812	1	21	Lincoln Park		3		2673	ffffff
2813	1	22	Broadway		3		FF0000	ffffff
2814	1	23	Millertown/Fairmont		3		339900	ffffff
2815	1	24	Inskip/Breda		3		2673	ffffff
2816	1	30	Parkridge		3		73B2FF	0
2817	1	31	Magnolia Avenue		3		FF0000	ffffff
2818	1	32	Dandridge Avenue		3		339900	ffffff
2819	1	33	Martin Luther King Jr. Boulevard		3		FFBB33	0
2820	1	34	Burlington		3		2673	ffffff
2821	1	40	South Knoxville		3		2673	ffffff
2823	1	41	Chapman Highway		3		FF0000	ffffff
2824	1	42	UT Hospital		3		339900	ffffff
2826	1	44(F)	Gateway at Knoxville Apartments		3		339900	ffffff
2827	1	45	Vestal		3		FFBB33	0
2833	1	90	Crosstown Connector		3		A900E6	ffffff

E. Shapes.txt

shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence	shape_dist_traveled
23211	35.965431	-83.913153	1	0
23211	35.965466	-83.913176	2	0.0045
23211	35.965582	-83.912985	3	0.0259
23211	35.965582	-83.912974	4	0.0269
23211	35.96552	-83.912929	5	0.0349
23211	35.965299	-83.91331	6	0.0766
23211	35.964909	-83.91399	7	0.1518
23211	35.964839	-83.91412	8	0.1662
23211	35.96479	-83.91422	9	0.1765
23211	35.964659	-83.91449	10	0.2048
23211	35.964579	-83.914679	11	0.224
23211	35.96453	-83.91479	12	0.2357
23211	35.964469	-83.914909	13	0.2473
23211	35.964309	-83.915229	14	0.2815
23211	35.96409	-83.91569	15	0.3295
23211	35.96388	-83.9162	16	0.3809
23211	35.96343	-83.917329	17	0.4941
23211	35.962599	-83.91681	18	0.5983
23211	35.962209	-83.9178	19	0.6976
23211	35.961789	-83.918859	20	0.8027
23211	35.96137	-83.91992	21	0.9086
23211	35.960969	-83.920979	22	1.0138
23211	35.960889	-83.921169	23	1.033
23211	35.960699	-83.92155	24	1.073
23211	35.96057	-83.92172	25	1.0935
23211	35.960419	-83.92187	26	1.1149
23211	35.96019	-83.92205	27	1.1454
23211	35.96001	-83.92222	28	1.171
23211	35.959889	-83.922369	29	1.1901

F. Stop_times.txt

trip_id	arrival_time	departure_time	stop_id	stop_sequence	stop_headsign	pickup_type	drop_off_type	shape_dist_traveled
252030	8:00:00	8:00:00	210	1		0	0	
252030	8:03:56	8:03:56	960	2		0	0	0.5522
252030	8:05:04	8:05:04	32	3		0	0	0.753
252030	8:07:48	8:07:48	1278	4		0	0	1.236
252030	8:10:00	8:10:00	2	5		0	0	1.6169
252030	8:10:35	8:10:35	34	6		0	0	1.7951
252030	8:11:17	8:11:17	154	7		0	0	2.0076
252030	8:12:40	8:12:40	1280	8		0	0	2.4276
252030	8:13:45	8:13:45	1643	9		0	0	2.7547
252030	8:15:09	8:15:09	471	10		0	0	3.177
252030	8:16:25	8:16:25	1383	11		0	0	3.5583
252030	8:17:23	8:17:23	472	12		0	0	3.8494
252030	8:19:20	8:19:20	473	13		0	0	4.4399
252030	8:21:11	8:21:11	474	14		0	0	4.9975
252030	8:24:33	8:24:33	475	15		0	0	6.0116
252030	8:25:50	8:25:50	476	16		0	0	6.399
252030	8:27:15	8:27:15	457	17		0	0	6.8257
252030	8:30:00	8:30:00	458	18		0	0	7.637
252031	9:00:00	9:00:00	210	1		0	0	
252031	9:03:56	9:03:56	960	2		0	0	0.5522
252031	9:05:04	9:05:04	32	3		0	0	0.753
252031	9:07:48	9:07:48	1278	4		0	0	1.236
252031	9:10:00	9:10:00	2	5		0	0	1.6169
252031	9:10:35	9:10:35	34	6		0	0	1.7951
252031	9:11:17	9:11:17	154	7		0	0	2.0076
252031	9:12:40	9:12:40	1280	8		0	0	2.4276
252031	9:13:45	9:13:45	1643	9		0	0	2.7547
252031	9:15:09	9:15:09	471	10		0	0	3.177
252031	9:16:25	9:16:25	1383	11		0	0	3.5583

G. Stops.txt

stop_id	stop_code	stop_name	stop_desc	stop_lat	stop_lon
2	CumbJAge	Cumberland at James Agee		35.958184	-83.928308
7	CumbPFul	Cumberland at Phillip Fulmer		35.958019	-83.928144
8	KingNth2	Kingston Pk. at Northshore	S NORTHSHORE DR & ACCESS	35.932732	-84.002774
9	GuyBTwr1	Guy B. Love Towers	E ANDERSON AVE & FOLSOM AVE	35.985158	-83.92509
11	SMryHsp1	St. Mary's Hospital	HURON ST & E OAK HILL AVE	35.992074	-83.928204
13	ChicBway	Chickamauga at Broadway		36.001396	-83.926411
14	SMryHsp2	St Mary's Hospital		35.992328	-83.928387
15	GuyBTwr2	Guy B. Love Towers		35.985332	-83.9252
16	KingLynW	Kingston Pk. at Lyons View	KINGSTON PIKE	35.941285	-83.977855
20	RayMWinW	Ray Mears Blvd WB @ Winston Rd	RAY MEARS BLVD & ACCESS	35.924109	-84.044946
23	WindSqre	Windsor Square	N SEVEN OAKS DR	35.913198	-84.098491
27	RayMWinE	Ray Mears Blvd EB @ Winston Rd		35.923771	-84.045071
30	WT@Belk	West Town between Belk & Pk Gar		35.923596	-84.037896
32	CumbLocW	Cumberland Ave WB @ Locust St		35.961533	-83.919847
33	CumbVol2	Cumberland Ave EB @ Volunteer Blvd		35.95718	-83.93042
34	Cumb16th	Cumberland Ave WB @ 16th St		35.957504	-83.930115
35	BoydUnv1	Boyd St NB @ University Ave		35.970919	-83.934563
36	McSpVirg	Virginia at McSpadden	MCSPADDEN ST	35.980403	-83.945369
38	640PLZ	I-640 Plaza	SHOPPERS LN	35.971666	-83.988916
39	UnivColl	College at University	UNIVERSITY AVE	35.968405	-83.937948
44	VirgMcSp	Virginia at McSpadden	VIRGINIA AVE & MCSPADDEN ST	35.980356	-83.94516
46	CollUniv	College at University	COLLEGE ST & UNIVERSITY AVE	35.968267	-83.938281
47	UnivBoyd	University Ave EB @ Boyd St	BOYD ST & UNIVERSITY AVE	35.970734	-83.934867
48	MainCCBl	Main Street EB @ City & County Bldg		35.961248	-83.917355
49	TexShm2	Texas Ave EB @ Sherman St		35.982124	-83.961684
50	CentBax1	Central St NB @ Baxter Ave		35.980642	-83.928218
53	KnoxJuvC	Knox Co. Juvenile Court		35.956007	-83.960633
54	StatOffc	State Office Bldg.		35.963941	-83.956732
55	ChroHltS	Cherokee Health		35.966511	-83.944551

H. Trips.txt

route_id	service_id	trip_id	trip_headsign	trip_short_name	direction_id	block_id	shape_id	wheelchair_accessible	bikes_allowed
2803	1	252030			0	12259	23211	0	0
2803	1	252031			0	12259	23211	0	0
2803	1	252032			0	12260	23212	0	0
2803	1	252033			0	12260	23211	0	0
2803	1	252034			0	12260	23213	0	0
2803	1	252035			1	12259	23215	0	0
2803	1	252036			1	12259	23214	0	0
2803	1	252037			1	12259	23216	0	0
2803	1	252038			1	12260	23214	0	0
2803	1	252039			1	12260	23214	0	0
2804	1	252040			0	12261	23219	0	0
2804	1	252075			0	12268	23219	0	0
2804	1	252041			0	12262	23217	0	0
2804	1	252071			0	12265	23217	0	0
2804	1	252042			0	12263	23217	0	0
2804	1	252074			0	12267	23217	0	0
2804	1	252043			0	12264	23217	0	0
2804	1	252072			0	12266	23217	0	0
2804	1	252044			0	12261	23217	0	0
2804	1	252085			0	12268	23217	0	0
2804	1	252045			0	12262	23217	0	0
2804	1	252073			0	12265	23217	0	0
2804	1	252046			0	12263	23217	0	0
2804	1	252047			0	12264	23217	0	0
2804	1	252048			0	12261	23217	0	0
2804	1	252049			0	12262	23217	0	0
2804	1	252050			0	12263	23217	0	0
2804	1	252051			0	12264	23217	0	0
2804	1	252052			0	12261	23217	0	0

APPENDIX-C: MEMPHIS GTFS DATA STRUCTURE

A. Agency.txt

agency_id, agency_name,
agency_url,agency_timezone,agency_lang,agency_phone,agency_fare_url

MATA, Memphis Area Transit Authority, http://www.matatransit.com,America/Chicago,en,,

B. Calendar.txt

service_id	monday	tuesday	wednesday	thursday	friday	saturday	sunday	start_date	end_date
2	0	0	0	0	0	1	0	20151213	20160423
3	0	0	0	0	0	0	1	20151213	20160423
5	1	1	1	1	1	0	0	20151213	20160423
3405	0	0	0	0	0	0	0	20151213	20160423
3505	0	0	0	0	0	0	0	20151213	20160423

C. Feed_info.txt

feed_publisher_name, feed_publisher_url,
feed_lang,feed_start_date,feed_end_date,feed_version

Memphis Area Transit Authoritiy,
http://www.matatransit.com,en,20151213,20160423,DEC2015_20151204

D. Routes.txt

route_id	agency_id	route_short_name	route_long_name	route_desc	route_type	route_url	route_color	route_text_color
2742	MATA	2	Madison		3		800000	0
2743	MATA	4	Walker		3		800000	0
2744	MATA	5	Central		3		8000	0
2745	MATA	6	Northaven		3		800080	0
2746	MATA	7	Air Park		3		FF00FF	0
2747	MATA	8	Chelsea		3		8080FF	0
2748	MATA	9	Highland		3		8080	0
2749	MATA	100	Trolley Main Line		3		FF80C0	0
2750	MATA	101	Trolley Riverfront		3		FF80C0	0
2751	MATA	102	Trolley Madison Line		3		FF80C0	0
2752	MATA	11	Thomas		3		800080	0
2753	MATA	12	Florida		3		8080	0
2754	MATA	13	Lauderdale		3		800000	0
2755	MATA	15	President's Island		3		C0C0C0	0
2756	MATA	17	McLemore		3		C0DCC0	0
2757	MATA	19	Vollintine		3		A6CAF0	0
2758	MATA	20	Bellevue Winchester		3		808080	0
2759	MATA	30	Brooks		3		FF0000	0
2760	MATA	32	East Parkway / Hollywood		3		80	0
2761	MATA	34	Walnut Grove		3		C0C0C0	0
2762	MATA	35	South Parkway		3		800080	0
2763	MATA	36	Hacks Cross		3		80	0
2764	MATA	37	Perkins		3		FF0000	0
2765	MATA	38	Boxtown-Westwood		3		0	0
2766	MATA	39	South Third		3		8080	0
2767	MATA	40	Wolfchase		3		C0C0C0	0
2768	MATA	42	Crosstown		3		0	0
2769	MATA	46	Whithaven		3		0	0
2770	MATA	50	Poplar		3		800080	0

…. continue

E. Shapes.txt

shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence	shape_dist_traveled
28567	35.044702	-89.981017	1	0
28567	35.044729	-89.980169	2	0.0771
28567	35.044749	-89.980098	3	0.0838
28567	35.044781	-89.98005	4	0.0888
28567	35.044835	-89.980022	5	0.0955
28567	35.044886	-89.980024	6	0.1015
28567	35.044909	-89.979992	7	0.1057
28567	35.044943	-89.979968	8	0.1093
28567	35.044988	-89.979953	9	0.1144
28567	35.045069	-89.979921	10	0.1239
28567	35.045232	-89.979922	11	0.1429
28567	35.045454	-89.979927	12	0.1679
28567	35.046169	-89.979984	13	0.2481
28567	35.046407	-89.980038	14	0.2746
28567	35.046484	-89.980056	15	0.2836
28567	35.046658	-89.980134	16	0.3039
28567	35.046703	-89.980159	17	0.3097
28567	35.046755	-89.980184	18	0.316
28567	35.046835	-89.980253	19	0.3268
28567	35.046906	-89.980342	20	0.3382
28567	35.046951	-89.980417	21	0.3468
28567	35.046973	-89.980462	22	0.3512
28567	35.047014	-89.980562	23	0.3615
28567	35.047028	-89.980667	24	0.3717
28567	35.047038	-89.981032	25	0.4047
28567	35.047059	-89.981162	26	0.4169
28567	35.047107	-89.981296	27	0.4299
28567	35.047182	-89.981401	28	0.4426
28567	35.047323	-89.98148	29	0.4601

…. continue

F. Stop_times.txt

trip_id	arrival_time	departure_time	stop_id	stop_sequence	stop_headsign	pickup_type	drop_off_type
355855	21:43:00	21:43:00	30	1	0	0	
355855	21:43:36	21:43:36	3342	2	0	0	0.2621
355855	21:43:52	21:43:52	3343	3	0	0	0.3821
355855	21:44:13	21:44:13	3344	4	0	0	0.5362
355855	21:44:27	21:44:27	3345	5	0	0	0.6382
355855	21:44:48	21:44:48	3526	6	0	0	0.7944
355855	21:45:01	21:45:01	3527	7	0	0	0.8944
355855	21:45:16	21:45:16	3528	8	0	0	1.0035
355855	21:45:40	21:45:40	3529	9	0	0	1.1836
355855	21:45:54	21:45:54	3530	10	0	0	1.2876
355855	21:46:22	21:46:22	3386	11	0	0	1.4916
355855	21:46:36	21:46:36	3387	12	0	0	1.5997
355855	21:46:48	21:46:48	3388	13	0	0	1.6928
355855	21:47:24	21:47:24	3389	14	0	0	1.96
355855	21:47:48	21:47:48	3842	15	0	0	2.1351
355855	21:47:48	21:47:48	3843	16	0	0	2.1351
355855	21:48:00	21:48:00	3390	17	0	0	2.2272
355855	21:48:21	21:48:21	3391	18	0	0	2.3813
355855	21:49:00	21:49:00	747	19	0	0	2.6032
355855	21:49:40	21:49:40	748	20	0	0	2.8606
355855	21:50:02	21:50:02	749	21	0	0	2.9998
355855	21:50:22	21:50:22	750	22	0	0	3.1299
355855	21:51:01	21:51:01	751	23	0	0	3.3775
355855	21:51:23	21:51:23	752	24	0	0	3.5169
355855	21:51:41	21:51:41	753	25	0	0	3.6329
355855	21:51:57	21:51:57	754	26	0	0	3.738
355855	21:52:16	21:52:16	755	27	0	0	3.8607
355855	21:52:37	21:52:37	756	28	0	0	3.9963
355855	21:52:49	21:52:49	757	29	0	0	4.076

…. continue

G. Stops.txt

stop_id, stop_code, stop_name, stop_desc, stop_lat,stop_lon,zone_id,stop_url,location_type,parent_station,stop_timezone,wheelchair_boarding

7, AIRKETSN, AIRWAYS BLVD@KETCHUM, AIRWAYS BLVD & KETCHUM RD,35.079543, -89.984917,,,,,,0

H. Trips.txt

route_id	service_id	trip_id	trip_headsign	direction_id	block_id	shape_id	wheelchair_accessible	bikes_allowed
2742	2	355865	2 Downtown	0	26872		0	0
2742	2	355864	2 Downtown	0	26872		0	0
2742	2	355863	2 Downtown	0	26872		0	0
2742	2	355862	2 Downtown	0	26872		0	0
2742	2	355861	2 Downtown	0	26872		0	0
2742	2	355860	2 Downtown	0	26872		0	0
2742	2	355859	2 Medical Center - Fairgrounds	0	26872		0	0
2742	2	355858	2 Medical Center - Fairgrounds	0	26872		0	0
2742	2	355857	2 Medical Center - Fairgrounds	0	26872		0	0
2742	2	355856	2 Medical Center - Fairgrounds	0	26872		0	0
2742	2	355855	2 Medical Center - Fairgrounds	0	26872		0	0
2742	2	355875	2 Medical Center Airport	1	26872		0	0
2742	2	355874	2 Medical Center Airport	1	26872		0	0
2742	2	355873	2 Medical Center Airport	1	26872		0	0
2742	2	355872	2 Medical Center Airport	1	26872		0	0
2742	2	355871	2 Medical Center Airport	1	26872		0	0
2742	2	355870	2 Medical Center - Fairgrounds	1	26872		0	0
2742	2	355869	2 Medical Center - Fairgrounds	1	26872		0	0
2742	2	355868	2 Medical Center - Fairgrounds	1	26872		0	0
2742	2	355867	2 Medical Center - Fairgrounds	1	26872		0	0
2742	2	355866	2 Medical Center - Fairgrounds	1	26872		0	0
2743	2	355977	4 Walker Alcy	0	26881		0	0
2743	2	355975	4 Walker Castalia	0	26880		0	0
2743	2	358881	4 Walker Alcy	0	26881		0	0
2743	2	358882	4 Walker Castalia	0	26880		0	0
2743	2	355973	4 Walker Alcy	0	26881		0	0
2743	2	355972	4 Walker Castalia	0	26880		0	0
2743	2	355971	4 Walker Alcy	0	26881		0	0
2743	2	355970	4 Walker Castalia	0	26880		0	0

…. continue

APPENDIX-D: NASHVILLE GTFS DATA STRUCTURE

A. Agency.txt

agency_id, agency_name, agency_url, agency_timezone,agency_lang,agency_phone,agency_fare_url

Nashville MTA, Nashville Metropolitan Transit Authority, http://www.nashvillemta.org/,America/Chicago,en,615-862-5950,

Nashville RTA,Regional Transportation Authority of Middle Tennessee,http://rtarelaxandride.com/,America/Chicago,en,615-862-8833,

B. Calendar.txt

service_id	monday	tuesday	wednesday	thursday	friday	saturday	sunday	start_date	end_date
1	1	1	1	1	1	0	0	20150927	20160326
2	0	0	0	0	0	1	0	20150927	20160326
3	0	0	0	0	0	0	1	20150927	20160326
301	0	0	0	0	0	0	0	20150927	20160326
401	0	0	0	0	0	0	0	20150927	20160326
201	0	0	0	0	0	0	0	20150927	20160326
101	0	0	0	0	0	0	0	20150927	20160326
901	0	0	0	0	0	0	0	20150927	20160326
1001	0	0	0	0	0	0	0	20150927	20160326
4	1	1	1	1	1	0	0	20150927	20160326
3204	0	0	0	0	0	0	0	20150927	20160326

C. Feed_info.txt

D. Routes.txt

route_id	agency_id	route_short_name	route_long_name	route_desc	route_type	route_url	route_color
1	Nashville	MTA	1	100 OAKS	3	00A651	FFFFFF
2	Nashville	MTA	2	BELMONT	3	00A651	FFFFFF
3	Nashville	MTA	3	WEST END WHITE BRIFGE	3	ED1C24	FFFFFF
4	Nashville	MTA	4	SHELBY	3	ED1C24	FFFFFF
5	Nashville	MTA	5	WEST END BELLEVUE	3	FF0000	FFFFFF
6	Nashville	MTA	6	LEBANON PIKE	3	00A651	FFFFFF
7	Nashville	MTA	7	HILLSBORO	3	ED1C24	FFFFFF
8	Nashville	MTA	8	8TH AVENUE SOUTH	3	00A651	FFFFFF
9	Nashville	MTA	9	METROCENTER	3	00A651	FFFFFF
10	Nashville	MTA	10	CHARLOTTE	3	ED1C24	FFFFFF
12	Nashville	MTA	12	NOLENSVILLE PIKE	3	ED1C24	FFFFFF
14	Nashville	MTA	14	WHITES CREEK	3	00A651	FFFFFF
15	Nashville	MTA	15	MURFREESBORO PIKE	3	ED1C24	FFFFFF
17	Nashville	MTA	17	12TH AVENUE SOUTH	3	ED1C24	FFFFFF
18	Nashville	MTA	18	AIRPORT DOWNTOWN HOTELS	3	00A651	FFFFFF
19	Nashville	MTA	19	HERMAN	3	ED1C24	FFFFFF
20	Nashville	MTA	20	SCOTT	3	00A651	FFFFFF
21	Nashville	MTA	21	UNIVERSITY CONNECTOR	3	FF0000	FFFFFF
22	Nashville	MTA	22	BORDEAUX	3	ED1C24	FFFFFF
23	Nashville	MTA	23	DICKERSON PIKE	3	ED1C24	FFFFFF
24	Nashville	MTA	24	BELLEVUE EXPRESS	3	F18C03	FFFFFF
25	Nashville	MTA	25	MIDTOWN CONNECTOR	3	00A651	FFFFFF
26	Nashville	MTA	26	GALLATIN PIKE	3	ED1C24	FFFFFF
27	Nashville	MTA	27	OLD HICKORY	3	FF8000	FFFFFF
28	Nashville	MTA	28	MERIDIAN	3	ED1C24	FFFFFF
29	Nashville	MTA	29	JEFFERSON	3	ED1C24	FFFFFF
30	Nashville	MTA	30	McFERRIN	3	00A651	FFFFFF

….continue

E. Shapes.txt

shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence	shape_dist_traveled
9649	36.166454	-86.782269	1	0
9649	36.166567	-86.781977	2	0.0291
9649	36.166454	-86.782268	3	0.0291
9649	36.166409	-86.78224	4	0.0349
9649	36.166126	-86.782064	5	0.0702
9649	36.165851	-86.781893	6	0.1042
9649	36.165682	-86.781792	7	0.1253
9649	36.165278	-86.781551	8	0.1758
9649	36.164543	-86.781098	9	0.2675
9649	36.164493	-86.78107	10	0.2729
9649	36.164019	-86.780756	11	0.3337
9649	36.16316	-86.780181	12	0.4424
9649	36.163038	-86.780099	13	0.4577
9649	36.162594	-86.779797	14	0.5145
9649	36.16188	-86.779326	15	0.6049
9649	36.161622	-86.779158	16	0.6375
9649	36.160863	-86.77868	17	0.7323
9649	36.160642	-86.778541	18	0.7605
9649	36.160493	-86.778442	19	0.7784
9649	36.15967	-86.777898	20	0.8835
9649	36.159401	-86.777718	21	0.9175
9649	36.158716	-86.777259	22	1.0039
9649	36.158437	-86.777097	23	1.0388
9649	36.158226	-86.777614	24	1.0902
9649	36.158031	-86.778087	25	1.1376
9649	36.157959	-86.778267	26	1.1555
9649	36.157847	-86.778563	27	1.1846
9649	36.157294	-86.778204	28	1.2544
9649	36.156927	-86.777958	29	1.3009

…continue

F. Stop_times.txt

trip_id	arrival_time	departure_time	stop_id	stop_sequence	stop_headsign	pickup_type	drop_off_type
116305	6:18:00	6:18:00	MCC5_11	1	0	1	
116305	6:19:23	6:19:23	CHA7AWN	2	0	0	0.3178
116305	6:20:00	6:20:00	CHA8AWN	3	0	0	0.4601
116305	6:22:13	6:22:13	CXONGULC	4	0	0	0.9683
116305	6:36:00	6:36:00	100OAKS	5	1	0	10.4874
116306	7:15:00	7:15:00	MCC5_11	1	0	1	
116306	7:16:32	7:16:32	CHA7AWN	2	0	0	0.3178
116306	7:17:13	7:17:13	CHA8AWN	3	0	0	0.4601
116306	7:19:41	7:19:41	CXONGULC	4	0	0	0.9683
116306	7:35:00	7:35:00	100OAKS	5	1	0	10.4874
116307	8:15:00	8:15:00	MCC5_11	1	0	1	
116307	8:16:23	8:16:23	CHA7AWN	2	0	0	0.3178
116307	8:17:00	8:17:00	CHA8AWN	3	0	0	0.4601
116307	8:19:13	8:19:13	CXONGULC	4	0	0	0.9683
116307	8:33:00	8:33:00	100OAKS	5	1	0	10.4874
116308	15:15:00	15:15:00	MCC5_11	1	0	1	
116308	15:16:49	15:16:49	5AUNISM	2	0	0	0.3315
116308	15:18:16	15:18:16	5AVCOMSN	3	0	0	0.5946
116308	15:19:13	15:19:13	5AVBROSN	4	0	0	0.7695
116308	15:22:00	15:22:00	6AVDEMSF	5	0	0	1.2678
116308	15:22:42	15:22:42	6APEASN	6	0	0	1.5456
116308	15:23:12	15:23:12	LAFEWIEN	7	0	0	1.7464
116308	15:24:53	15:24:53	6AOAKSN	8	0	0	2.4127
116308	15:26:05	15:26:05	FORVINSM	9	0	0	2.8898
116308	15:27:22	15:27:22	FORCHESN	10	0	0	3.3999
116308	15:28:18	15:28:18	HAMFOREF	11	0	0	3.7719
116308	15:29:30	15:29:30	HAMMAREN	12	0	0	4.2497
116308	15:29:49	15:29:49	MARMOOSF	13	0	0	4.3805
116308	15:31:00	15:31:00	MARWEDSN	14	0	0	4.8138

…continue

G. Stops.txt

stop_id	stop_code	stop_name	stop_desc	stop_lat	stop_lon
10ABENNN	10ABENNN	10TH AVE S & BENTON AVE NB		36.132025	-86.785641
10AGILSN	10AGILSN	10TH AVE S & GILMORE AVE SB	10TH AV S & GILMORE AV	36.124667	-86.786403
10AHALNN	10AHALNN	10TH AVE S & HALCYON AVE NB		36.122265	-86.78647
10AHALSN	10AHALSN	10TH AVE S & HALCYON AVE SB	10TH AV S & HALCYON AV	36.122488	-86.786537
10AHERNN	10AHERNN	10TH AVE N & HERMAN ST NB	10TH AV N & HERMAN ST	36.168831	-86.792339
10ALAWSN	10ALAWSN	10TH AVE S & LAWRENCE AVE SB	10TH AV S & LAWRENCE AV	36.129275	-86.786044
10ALAWNN	10ALAWNN	10TH AVE S & LAWRENCE AVE NB	10TH AV S & LAWRENCE AV	36.128739	-86.785943
10ASDONN	10ASDONN	10TH AVE S & S DOUGLAS AVE NB	10TH AV S & S DOUGLAS AV	36.13031	-86.785797
10AWALNN	10AWALNN	10TH AVE S & WALDKIRCH AVE NB	10TH AV S & WALDKIRCH AV	36.126855	-86.786145
11AWHESN	11AWHESN	11TH AVE N & WHEELESS ST SB		36.179003	-86.803879
12AARCSN	12AARCSN	12TH AVE S & ARCHER ST SB	12TH AV S & ARCHER ST	36.145597	-86.786258
12AARCSM	12AARCSM	12TH AVE S & ARCHER ST SB	12TH AV S & ARCHER ST	36.144535	-86.786549
12AARGNN	12AARGNN	12TH AVE S & ARGYLE AVE NB	12TH AV S & ARGYLE AV	36.137677	-86.788074
12ACALSN	12ACALSN	12TH AVE S & CALDWELL AVE SB	12TH AV S & CALDWELL AV	36.132811	-86.788803
12ACALNM	12ACALNM	12TH AVE S & CALDWELL AVE NB	12TH AV S & CALDWELL AV	36.132436	-86.788657
12ADEMNN	12ADEMNN	12TH AVE S & DEMONBREUN ST NB	12TH AV S & DEMONBREUN ST	36.154509	-86.786056
12ADIVNN	12ADIVNN	12TH AVE S & DIVISION ST NB		36.150705	-86.784306
12AEDGSN	12AEDGSN	12TH AVE S & EDGEHILL AVE SB	12TH AV S & EDGEHILL AV	36.142972	-86.787031
12AEDGNN	12AEDGNN	12TH AVE S & EDGEHILL AVE NB	12TH AV S & EDGEHILL AV	36.142624	-86.786998
GRAFERSN	GRAFERSN	GRANNY WHITE PIKE & FERGUSON AVE SB	GRANNY WHITE PK & FERGUSON AV	36.118934	-86.791122
GRAFERNN	GRAFERNN	GRANNY WHITE PIKE & FERGUSON AVE NB	GRANNY WHITE PK & FERGUSON AV	36.118756	-86.791055
12AHAWSN	12AHAWSN	12TH AVE S & HAWKINS ST SB	12TH AV S & HAWKINS ST	36.148669	-86.785136
12AHAWNN	12AHAWNN	12TH AVE S & HAWKINS ST NB	12TH AV S & HAWKINS ST	36.148383	-86.785282
12AHORSM	12AHORSM	12TH AVE S & EDGEHILL AVE SB	12TH AV S & HORTON AV	36.141267	-86.78739
12AHORSN	12AHORSN	12TH AVE S & HORTON AVE SB	12TH AV S & HORTON AV	36.140043	-86.787626
12AHORNN	12AHORNN	12TH AVE S & HORTON AVE NB	12TH AV S & HORTON AV	36.139606	-86.787525
12ALAUSN	12ALAUSN	12TH AVE S & LAUREL ST SB	12TH AV S & LAUREL ST	36.153723	-86.785686
12ALAWNN	12ALAWNN	12TH AVE S & LAWRENCE AVE NB	12TH AV S & UNKNOWN AL	36.129319	-86.788814
12ALINNN	12ALINNN	12TH AVE S & LINDEN AVE NB	12TH AV S & LINDEN AV	36.127614	-86.789082

…continue

H. Trips.txt

route_id	service_id	trip_id	trip_headsign	trip_short_name	direction_id	block_id	shape_id
1	1	116305	100 OAKS EXPRESS		0	12879	9650
1	1	116306	100 OAKS EXPRESS		0	12879	9650
1	1	116307	100 OAKS EXPRESS		0	12879	9650
1	1	116311	100 OAKS MALL		0	12881	9649
1	1	116308	100 OAKS MALL		0	12880	9649
1	1	116309	100 OAKS MALL		0	12880	9649
1	1	116310	100 OAKS MALL		0	12880	9649
1	1	116312	DOWNTOWN		1	12879	9651
1	1	116313	DOWNTOWN		1	12879	9651
1	1	116314	DOWNTOWN		1	12879	9651
1	1	116315	DOWNTOWN		1	12879	9651
1	1	116319	DOWNTOWN EXPRESS		1	12881	9654
1	1	116316	DOWNTOWN EXPRESS		1	12880	9653
1	1	116317	DOWNTOWN EXPRESS		1	12880	9653
1	1	116318	DOWNTOWN EXPRESS		1	12880	9653
2	1	116323	BELMONT		0	12883	9655
2	1	116324	BELMONT		0	12885	9655
2	1	116321	BELMONT		0	12883	9655
2	1	116325	BELMONT		0	12885	9655
2	1	116330	BELMONT		0	12882	9655
2	1	116322	BELMONT		0	12884	9655
2	1	116329	BELMONT		0	12882	9655
2	1	116326	BELMONT		0	12884	9655
2	1	116327	BELMONT		0	12882	9655
2	1	116328	BELMONT		0	12884	9655
2	1	116320	BELMONT		0	12882	9655
2	1	116332	DOWNTOWN		1	12883	9656
2	1	116335	DOWNTOWN		1	12885	9656
2	1	116336	DOWNTOWN		1	12883	9656

…continue

APPENDIX-E: SCRIPT TO PROCESS GTFS DATA

Example code-Nashville

```
#####################################

### GTFS CONNECTIVITY TOOL v1.0###

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

#Version Date: 12.07.2015

#Coded by: Tim Welch, Georgia Tech and Sabya Mishra, University of Memphis

#You are free to use, modify and distribute this code

library("maptools")

library("foreign")

library("plyr")

library("dplyr")

library("Hmisc")

library("ggplot2")

library("ggthemes")

library("parallel")

library("doParallel")

library("snow")

library("data.table")

library("sp")

library("leaflet")


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

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

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

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

# THESE ARE THE ONLY TWO LINES YOU CHANGE AND IT SHOULD RUN
```

```
LOC<-"Nashville"

PDFFILENAME<-"Nashville.pdf"

PDFFILENAME2<-"Nashville.png"

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

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

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

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

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



polygon<-readShapePoly("ActivityShape.shp")

#Convert Inputs to CSV

stop_timeTXT = read.delim("stop_times.txt",sep=",") #Stop-Time file

write.table(stop_timeTXT, file="stop_times.csv",sep=",",col.names=TRUE,row.names=FALSE)


tripTXT = read.delim("trips.txt",sep=",") #Stop-Time file

write.table(tripTXT, file="trips.csv",sep=",",col.names=TRUE,row.names=FALSE)


routesTXT = read.delim("routes.txt",sep=",") #Stop-Time file

write.table(routesTXT, file="routes.csv",sep=",",col.names=TRUE,row.names=FALSE)


stopsTXT = read.delim("stops.txt",sep=",") #Stop-Time file

write.table(stopsTXT, file="stops.csv",sep=",",col.names=TRUE,row.names=FALSE)


ShapesTXT = read.delim("shapes.txt",sep=",") #Stop-Time file
```

```
write.table(ShapesTXT, file="Shapes.csv",sep=",",col.names=TRUE,row.names=FALSE)


#Read Back CSVs

stop_times <- read.csv("stop_times.csv",sep=",")

trips <- read.csv("trips.csv",sep=",")

routes <- read.csv("routes.csv",sep=",")

stops <- read.csv("stops.csv",sep=",")

shapes <- read.csv("Shapes.csv",sep=",")


# merge Routes with Trips

trips_routes = merge(trips, routes, by="route_id")

Stops_times_trips_routes = merge(stop_times, trips_routes, by="trip_id")


Stops_times_trips_routes$arrival_time <- as.character(Stops_times_trips_routes$arrival_time)

Stops_times_trips_routes$arrival_time <-
sapply(strsplit(Stops_times_trips_routes$arrival_time,":"),

                                   function(x) {

                                     x <- as.numeric(x)

                                     x[1]+x[2]/60+x[3]/3600

                                   }

)


#Select 1 direction

Stops_times_trips_routes_Dir1 <- subset(Stops_times_trips_routes,direction_id==1)


###CALC TRIP TIME

# use aggregate to create new data frame with the maxima

Stops_times_trips_routes_Dir1_aggMAX <- aggregate(arrival_time ~ trip_id,
Stops_times_trips_routes_Dir1, max)
```

```
names(Stops_times_trips_routes_Dir1_aggMAX)[1] <- "trip_id"

names(Stops_times_trips_routes_Dir1_aggMAX)[2] <- "timeMAX"

Stops_times_trips_routes_Dir1_aggMIN <- aggregate(arrival_time ~ trip_id,
Stops_times_trips_routes_Dir1, min)

names(Stops_times_trips_routes_Dir1_aggMIN)[1] <- "trip_id"

names(Stops_times_trips_routes_Dir1_aggMIN)[2] <- "timeMIN"


Stops_times_trips_routes_Dir1_aggMINMAX = merge(Stops_times_trips_routes_Dir1_aggMIN,
Stops_times_trips_routes_Dir1_aggMAX, by="trip_id")


Stops_times_trips_routes_Dir1_aggMINMAX$RouteTime <-
(Stops_times_trips_routes_Dir1_aggMINMAX$timeMAX-
Stops_times_trips_routes_Dir1_aggMINMAX$timeMIN)*60


###CALC TRIP DISTANCE

Stops_times_trips_routes_Dir1_TripDISTMAX <- aggregate(shape_dist_traveled ~ trip_id,
Stops_times_trips_routes_Dir1, max)

names(Stops_times_trips_routes_Dir1_TripDISTMAX)[1] <- "trip_id"

names(Stops_times_trips_routes_Dir1_TripDISTMAX)[2] <- "distMAX"

Stops_times_trips_routes_Dir1_TripDISTMIN <- aggregate(shape_dist_traveled ~ trip_id,
Stops_times_trips_routes_Dir1, min)

names(Stops_times_trips_routes_Dir1_TripDISTMIN)[1] <- "trip_id"

names(Stops_times_trips_routes_Dir1_TripDISTMIN)[2] <- "distMIN"


Stops_times_trips_routes_Dir1_TripDISTMINMAX =
merge(Stops_times_trips_routes_Dir1_TripDISTMIN,
Stops_times_trips_routes_Dir1_TripDISTMAX, by="trip_id")


Stops_times_trips_routes_Dir1_TripDISTMINMAX$RouteDIST <-
(Stops_times_trips_routes_Dir1_TripDISTMINMAX$distMAX-
Stops_times_trips_routes_Dir1_TripDISTMINMAX$distMIN)
```

###Merge Time and Distance, CALC SPPED

```
Stops_times_trips_routes_Dir1_TripSPEED =
merge(Stops_times_trips_routes_Dir1_aggMINMAX,
Stops_times_trips_routes_Dir1_TripDISTMINMAX, by="trip_id")

Stops_times_trips_routes_Dir1_TripSPEED$TripSpeed <-
((Stops_times_trips_routes_Dir1_TripSPEED$RouteDIST/5280)/Stops_times_trips_routes_Dir1
_TripSPEED$RouteTime)*60
```

###CALC HEADWAY

```
Stops_times_trips_routes_Dir1_aggMAX["Count"] <-1
```

```
# use aggregate to create new data frame with the maxima

Stops_times_trips_routes_Dir1_RouteTimeMAX <- aggregate(arrival_time ~ route_id,
Stops_times_trips_routes_Dir1, max)

names(Stops_times_trips_routes_Dir1_RouteTimeMAX)[1] <- "route_id"

names(Stops_times_trips_routes_Dir1_RouteTimeMAX)[2] <- "timeMAX"

Stops_times_trips_routes_Dir1_RouteTimeMIN <- aggregate(arrival_time ~ route_id,
Stops_times_trips_routes_Dir1, min)

names(Stops_times_trips_routes_Dir1_RouteTimeMIN)[1] <- "route_id"

names(Stops_times_trips_routes_Dir1_RouteTimeMIN)[2] <- "timeMIN"


Stops_times_trips_routes_Dir1_RouteTimeMINMAX =
merge(Stops_times_trips_routes_Dir1_RouteTimeMAX,
Stops_times_trips_routes_Dir1_RouteTimeMIN, by="route_id")

Stops_times_trips_routes_Dir1_RouteTimeMINMAX$trips_RouteTime <-
Stops_times_trips_routes_Dir1_RouteTimeMINMAX$timeMAX-
Stops_times_trips_routes_Dir1_RouteTimeMINMAX$timeMIN


#Select 1 direction

trips_Dir1 <- subset(trips,direction_id==1)


trips_Dir1_TripCount <- aggregate(direction_id ~ route_id, trips_Dir1, sum)

names(trips_Dir1_TripCount)[1] <- "route_id"
```

```
names(trips_Dir1_TripCount)[2] <- "TripCount"


##Merge Route Info acn CALC final headways

Trips_Routes_Headway = merge(Stops_times_trips_routes_Dir1_RouteTimeMINMAX,
trips_Dir1_TripCount, by="route_id")

Trips_Routes_Headway$RouteHeadway <-
(Trips_Routes_Headway$trips_RouteTime/Trips_Routes_Headway$TripCount)*60

Trips_Routes_Headway$RouteFrequency <- 60/Trips_Routes_Headway$RouteHeadway


###CALC CAPACITY and CAP

ifelse(routes$route_type==3, routes$CAPACITY <- 50, 200)

routes_ALLDATA = merge(routes, Trips_Routes_Headway, by="route_id")

routes_ALLDATA$CAP <-  routes_ALLDATA$CAPACITY*routes_ALLDATA$RouteFrequency


###CALC DIST IN-OUT

stop_times_trips_routes_TRIPDIST = merge(Stops_times_trips_routes,
Stops_times_trips_routes_Dir1_TripSPEED, by="trip_id")

stop_times_trips_routes_TRIPDIST$OriginDistance =
stop_times_trips_routes_TRIPDIST$shape_dist_traveled -
stop_times_trips_routes_TRIPDIST$distMIN

stop_times_trips_routes_TRIPDIST$DestinationDistance =
stop_times_trips_routes_TRIPDIST$RouteDIST-
stop_times_trips_routes_TRIPDIST$OriginDistance


###CALC ACTIVITY

coordinates(stops)=~stop_lon+stop_lat

stops$Activity = over(stops,polygon)


stop_times_trips_routes_TRIPDIST_Activity = merge(stop_times_trips_routes_TRIPDIST,
stops, by="stop_id")


###Construct Connectivity INPUT FILE
```

```
stop_times_trips_routes_TRIPDIST_ROUTEDATA =
merge(stop_times_trips_routes_TRIPDIST_Activity, routes_ALLDATA, by="route_id")

INPUT <-
stop_times_trips_routes_TRIPDIST_ROUTEDATA[,c("route_id","RouteDIST","stop_id","OriginD
istance","DestinationDistance","TripSpeed","RouteFrequency","CAPACITY","CAP","Activity")]

names(INPUT)[1] <- "Line"

names(INPUT)[2] <- "Distance"

names(INPUT)[3] <- "Node"

names(INPUT)[4] <- "Origin Distance"

names(INPUT)[5] <- "Destination Distance"

names(INPUT)[6] <- "Speed"

names(INPUT)[7] <- "Frequency"

names(INPUT)[8] <- "Capacity"

names(INPUT)[9] <- "CAP"

names(INPUT)[10] <- "Activity"


write.table(INPUT, file="input.csv",sep=",",col.names=TRUE,row.names=FALSE)


##########CALC CONNECTIVITY##########

n0<-read.csv("input.csv", header=TRUE)

n0[["Activity"]][is.na(n0[["Activity"]])] <- 1

n<-na.omit(n0)


#Mean and Stdev.

# add .rm=TRUE if missing values

mean_Odist=mean(n$Origin.Distance);sd_Odist=sd(n$Origin.Distance);

mean_Ddist=mean(n$Destination.Distance);sd_Ddist=sd(n$Destination.Distance);

mean_speed=mean(n$Speed);sd_speed=sd(n$Speed);

mean_cap=mean(n$CAP); sd_cap=sd(n$CAP);

mean_actv=mean(n$Activity); sd_actv=sd(n$Activity)
```

```
#Scaling Coefficients

alpha<-pnorm(n$CAP, mean = mean_cap, sd = sd_cap, log = FALSE)

beta<-pnorm(n$Speed, mean = mean_speed, sd = sd_speed, log = FALSE)

gamma<-pnorm(n$Origin.Distance, mean = mean_Odist, sd = sd_Odist, log = FALSE)

phi<-pnorm(n$Activity, mean = mean_actv, sd = sd_actv, log = FALSE)


n<-cbind(n,alpha,beta,gamma,phi)


#Connecting power

length=nrow(n)

inbound=((n$alpha*0.01*n$CAP)*(n$beta*0.01*n$Speed)*(n$gamma*n$Destination.Distance)*(
n$phi*n$Activity))

outbound=((n$alpha*0.01*n$CAP)*(n$beta*0.01*n$Speed)*(n$gamma*n$Origin.Distance)*(n$p
hi*n$Activity))

Avg_CP=((inbound+outbound)/2)

n<-cbind(n,inbound,outbound,Avg_CP)


## Node Connecting power(node_CP)

node_step1<-aggregate(.~Node,data=n,sum);

freq<-as.data.frame(table(n$Node))

node_CP= (node_step1[,ncol(node_step1)])/freq[2];

node_CP=cbind(node_step1[1],node_CP); names(node_CP)<- c("Node", "Connectivity Index");


## Line Connecting Power(LCP)

Node= n$Node

A<-cbind(n[3],n[1],n[17])

sum_CP=rep(NA,length(Node))

A<-data.frame(A,sum_CP)
```

```
A$sum_CP<-node_step1$Avg_CP[match(A$Node,node_step1$Node)]


require(dplyr)

A<-group_by(A,Line)

A.summary <- summarise(A,Line_CP=sum(sum_CP))

freq_L<-as.data.frame(table(n$Line))

LCP= (A.summary[2])/((freq_L[2])-1);

LCP=cbind(A.summary[1],LCP); names(LCP)<- c("Line", "Line Connecting Power");


#Export the Output

write.csv(n,"Output_pt.1.csv",row.names=FALSE)

write.csv(node_CP,"Output_Node.csv",row.names=FALSE)

write.csv(LCP,"Output_Line.csv",row.names=FALSE)


#Aggregate and Export Line (Route) Data to Shape ID for mapping

names(LCP)[names(LCP)=="Line"] <- "route_id"

tripsRED<-trips[,c("route_id","shape_id")]

LCP_Routes = merge(tripsRED, LCP, by="route_id")

LCP_RoutesRED<-LCP_Routes[,c("shape_id","Line Connecting Power")]

LCP_AGG<-aggregate(.~shape_id,data=LCP_RoutesRED,sum)

LCP_AGG_Out = merge(shapes, LCP_AGG, by="shape_id",all = TRUE)

LCP_AGG_Out[is.na(LCP_AGG_Out)] <- 0

names(LCP_AGG_Out)[names(LCP_AGG_Out)=="Line Connecting Power"] <- "connectivity"

#Normalize scores

LCP_AGG_Out$CONN_NORM<-((LCP_AGG_Out$connectivity-
min(LCP_AGG_Out$connectivity))/(max(LCP_AGG_Out$connectivity)-
min(LCP_AGG_Out$connectivity)))*100

LCP_AGG_Out$CONN_GROUP <- as.numeric(cut2(LCP_AGG_Out$CONN_NORM, g=5))

colors <- read.csv('colors.csv')
```

```
LCP_AGG_Out$CONNCOLOR = colors[match(LCP_AGG_Out$CONN_GROUP,
colors$CONN_GROUP),"CONCOLOR"]

write.csv(LCP_AGG_Out,"Shape_Output.csv",row.names=FALSE)


###Plot Lines and save map

Shape_Output <- read.csv("Shape_Output.csv")

Shape_Output=Shape_Output[ order(-Shape_Output[,4], -Shape_Output[,1]), ]

Shape_Output$size<-Shape_Output$CONN_GROUP/5

p <- ggplot(Shape_Output) + geom_path(aes(shape_pt_lon, shape_pt_lat, group = shape_id),
size = Shape_Output$size, alpha = .5, colour=Shape_Output$CONNCOLOR) + coord_equal()
+ theme_map()

p<- p + theme(plot.title = element_text(size=15, face="bold"))


p

ggsave(file=PDFFILENAME)

ggsave(file=PDFFILENAME2)



###Move image to main directory



#Export Node (Stop) Data to Stop ID for mapping

names(node_CP)[names(node_CP)=="Node"] <- "stop_id"

node_CP_Out = merge(stops, node_CP, by="stop_id")

write.csv(node_CP_Out,"Stop_Output.csv",row.names=FALSE)
```

APPENDIX-F: SCRIPT TO DEVELOP COUNT MODEL FOR FREQUENCY OF DEMAND RESPONSE TRANSIT TRIPS

```
rm(list=ls())
require(IDPmisc);
library(MASS);
require(ggplot2);
require(pscl);
require(boot);
require(Metrics)
require(repmis)
require(foreign)
library(readr)
library(glmmTMB);
TrainingDataDRT5=read.table(file="E:/DRT5/TrainingDataDRT5.txt",sep ="\t",header = TRUE)
testDataDRT5=read.table(file="E:/DRT5/testDataDRT5.txt",sep = "\t",header =TRUE)
#factor origin and destination zipcode
TrainingDataDRT5$OrgZip1<-as.factor(TrainingDataDRT5$OrgZip)
TrainingDataDRT5$DestZip1<-as.factor(TrainingDataDRT5$DestZip)
testDataDRT5$OrgZip1<-as.factor(testDataDRT5$OrgZip)
testDataDRT5$DestZip1<-as.factor(testDataDRT5$DestZip)
#Poisson
summary(combTrainfinal.pos <- glm(Trips ~
org_DomSEx+org_DomRace+des_RetTrad+distanc+org_AHHSize+logoforg_Disablep08t012+logoforg
_Popden+logoforg_64over+org_HOVR+logoforg_HHMein
,data = TrainingDataDRT5, trace = F,family = poisson))
test1.actual=testDataDRT5$Trips
test.predict.pos=predict(combTrainfinal.pos, testDataDRT5, type="response")
#Negative Binomial
summary(combTrainfinal.nb <- glm.nb(Trips ~
prwhite+logoforg_Disablep08t012+logoforg_HHMein+logoforg_64over+logofdes_RetTrad+logofdes_W
hSTrad+logoforg_Ageupto14,data = TrainingDataDRT5,control=glm.control(maxit=100))) ;
test1.actual=testDataDRT5$Trips
test.predict.nb=predict(combTrainfinal.nb, testDataDRT5, type="response")
#Hurdle poisson
summary(combTrainfinal.hp <- hurdle(Trips
~logoforg_Ageupto14+logoforg_Disablep08t012+logoforg_Popden+logofdes_RetTrad+logofdes_WhSTr
ad+distanc+logoforg_RentOCC|org_AHHSize+logoforg_Disablep08t012+logoforg_Popden+distanc+log
oforg_RentOCC+logoforg_Ageupto14+logoforg_HHMein ,data = TrainingDataDRT5, link = "logit", dist
= "poisson"))
test1.actual=testDataDRT5$Trips
test.predict.hp=predict(combTrainfinal.hp, testDataDRT5, type="response")
#Zero-inflted Negative Binomial
summary(combTrainfinal.zinb <- zeroinfl(Trips
~logoforg_Ageupto14+logoforg_Disablep08t012+logoforg_Popden+logofdes_RetTrad+logofdes_WhSTr
ad+distanc+logoforg_RentOCC|org_AHHSize+logoforg_64over+distanc+logoforg_RentOCC+logofdes_
RetTrad+logoforg_Ageupto14 ,data = TrainingDataDRT5, link = "logit", dist = "negbin", trace = F, EM =
F))
test1.actual=testDataDRT5$Trips
test.predict.zinb=predict(combTrainfinal.zinb, testDataDRT5, type="response")
#Hurdle NB
```

```
summary(combTrainfinal.hnb <- hurdle(Trips
~logoforg_Ageupto14+logoforg_Disablep08t012+logoforg_Popden+logofdes_RetTrad+logofdes_WhSTr
a
|org_AHHSize+logoforg_64over+distanc+logoforg_RentOCC+logoforg_Ageupto14+logoforg_Disablep
08t012+logoforg_HHMein  ,data = TrainingDataDRT5, link = "logit", dist = "negbin"))
summary(combTrainfinal.hnb)
test1.actual=testDataDRT5$Trips
test.predict.hnb=predict(combTrainfinal.hnb, testDataDRT5, type="response")
#Zero inflated poisson
summary(combTrainfinal.zip <-
zeroinfl(Trips~logofdes_RetTrad+distanc+org_AHHSize+logoforg_Disablep08t012+logoforg_Popden+l
ogoforg_HHMein+org_HOVR
|org_HHInG200+org_AHHSize+logofdes_RetTrad+logoforg_HHMein+logoforg_Ageupto14+logoforg_
RentOCC+logoforg_Popden ,data = TrainingDataDRT5, link = "logit", dist = "poisson", trace = F, EM =
F))
test1.actual=testDataDRT5$Trips
test.predict.zip=predict(combTrainfinal.zip, testDataDRT5, type="response")
#ZINB Mixed effect
m102 <- glmmTMB(Trips~logoforg_Ageupto14+logoforg_Disablep08t012
+logoforg_Popden+logofdes_RetTrad+logofdes_WhSTrad+distanc+logoforg_RentOCC+(1|org_RUCA)
+(1|des_RUCA),zi=~org_AHHSize+logoforg_64over+distanc+logoforg_RentOCC+logofdes_RetTrad+l
ogoforg_Ageupto14+(1|OrgZip1)+(1|DestZip1),family=nbinom2, TrainingDataDRT5)
summary(m102)
test1.actual=testDataDRT5$Trips
test.predict.m102=predict(m102, testDataDRT5, zitype="response")
# Calculate Elasticity of Zero-inflted Negative Binomial Mixed effect
as.numeric(fixef(m102)$cond["logoforg_Ageupto14"]
*mean(TrainingDataDRT5$logoforg_Ageupto14)/mean(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$cond["logoforg_Disablep08t012"]*mean(TrainingDataDRT5$logoforg_Disable
p08t012)/mean(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$cond["logoforg_Popden"] *
mean(TrainingDataDRT5$logoforg_Popden)/mean(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$cond["logofdes_RetTrad"]
*mean(TrainingDataDRT5$logofdes_RetTrad)/mean(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$cond["logofdes_WhSTrad"]*mean(TrainingDataDRT5$logofdes_WhSTrad)/me
an(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$cond["distanc"] *
mean(TrainingDataDRT5$distanc)/mean(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$cond["logoforg_RentOCC"] *
mean(TrainingDataDRT5$logoforg_RentOCC)/mean(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$zi["org_AHHSize"] *
mean(TrainingDataDRT5$org_AHHSize)/mean(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$zi["logoforg_64over"] *
mean(TrainingDataDRT5$logoforg_64over)/mean(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$zi["logoforg_RentOCC"] *
mean(TrainingDataDRT5$logoforg_RentOCC)/mean(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$zi["logoforg_Ageupto14"] *
mean(TrainingDataDRT5$logoforg_Ageupto14)/mean(TrainingDataDRT5$Trips))
as.numeric(fixef(m102)$zi["distanc"] *
mean(TrainingDataDRT5$distanc)/mean(TrainingDataDRT5$Trips))
```

```
as.numeric(fixef(m102)$zi["logofdes_RetTrad"] *
mean(TrainingDataDRT5$logofdes_RetTrad)/mean(TrainingDataDRT5$Trips))
#AAPD Calculation (Predictive performance of model)
sum(test1.actual[1:36250]<1)
sum(test1.actual[36251:36761]>=1 & test1.actual[36251:36761]<10)
sum(test1.actual[36762:36963]>=10 & test1.actual[36762:36963]<100)
sum(test1.actual[36964:37046]>=100)
#Zero-inflted Negative Binomial
sum(test.predict.zinb[1:36250]<1)
sum(test.predict.zinb[36251:36761]>=1 & test.predict.zinb[36251:36761]<10)
sum(test.predict.zinb[36762:36963]>=10 & test.predict.zinb[36762:36963]<100)
sum(test.predict.zinb[36964:37046]>=100)
#Poisson
sum(test.predict.pos[1:36250]<1)
sum(test.predict.pos[36251:36761]>=1 & test.predict.pos[36251:36761]<10)
sum(test.predict.pos[36762:36963]>=10 & test.predict.pos[36762:36963]<100)
sum(test.predict.pos[36964:37046]>=100)
#Negative Binomial
sum(test.predict.nb[1:36250]<1)
sum(test.predict.nb[36251:36761]>=1 & test.predict.nb[36251:36761]<10)
sum(test.predict.nb[36762:36963]>=10 & test.predict.nb[36762:36963]<100)
sum(test.predict.nb[36964:37046]>=100)
#Zero inflated poisson
sum(test.predict.zip[1:36250]<1)
sum(test.predict.zip[36251:36761]>=1 & test.predict.zip[36251:36761]<10)
sum(test.predict.zip[36762:36963]>=10 & test.predict.zip[36762:36963]<100)
sum(test.predict.zip[36964:37046]>=100)
#Hurdle poisson
sum(test.predict.hp[1:36250]<1)
sum(test.predict.hp[36251:36761]>=1 & test.predict.hp[36251:36761]<10)
sum(test.predict.hp[36762:36963]>=10 & test.predict.hp[36762:36963]<100)
sum(test.predict.hp[36964:37046]>=100)
#Hurdle NB
sum(test.predict.hnb[1:36250]<1)
sum(test.predict.hnb[36251:36761]>=1 & test.predict.hnb[36251:36761]<10)
sum(test.predict.hnb[36762:36963]>=10 & test.predict.hnb[36762:36963]<100)
sum(test.predict.hnb[36964:37046]>=100)
#Zero-inflated Negative Binomial Mixed effect
sum(test.predict.m102[1:36250]<1)
sum(test.predict.m102[36251:36761]>=1 & test.predict.m102[36251:36761]<10)
sum(test.predict.m102[36762:36963]>=10 & test.predict.m102[36762:36963]<100)
sum(test.predict.m102[36964:37046]>=100)
#Goodness of Fit
#Poisson
logLik(combTrainfinal.pos)
AIC(combTrainfinal.pos)
BIC(combTrainfinal.pos)
#Negative Binomial
logLik(combTrainfinal.nb)
AIC(combTrainfinal.nb)
BIC(combTrainfinal.nb)
```

```
#Zero-inflated Poisson
AIC(combTrainfinal.zip)
AIC(combTrainfinal.zip, k = log(nrow(TrainingDataDRT5))) #for BIC
logLik(combTrainfinal.zip)
#Zero-inflated Negative Binomial
logLik(combTrainfinal.zinb)
AIC(combTrainfinal.zinb)
AIC(combTrainfinal.zinb, k = log(nrow(TrainingDataDRT5))) #for BIC
#Hurdle NB
logLik(combTrainfinal.hnb)
AIC(combTrainfinal.hnb)
AIC(combTrainfinal.hnb, k = log(nrow(TrainingDataDRT5)))#for BIC
#Hurdle poisson
logLik(combTrainfinal.hp)
AIC(combTrainfinal.hp)
AIC(combTrainfinal.hp, k = log(nrow(TrainingDataDRT5)))#for BIC
#Zero-inflted Negative Binomial Mixed effect
logLik(m102)
AIC(m102)
AIC(m102, k = log(nrow(TrainingDataDRT5)))#for BIC
```

APPENDIX-G: ADDITIONAL MAPS FOR DEMAND BASED TRIPS

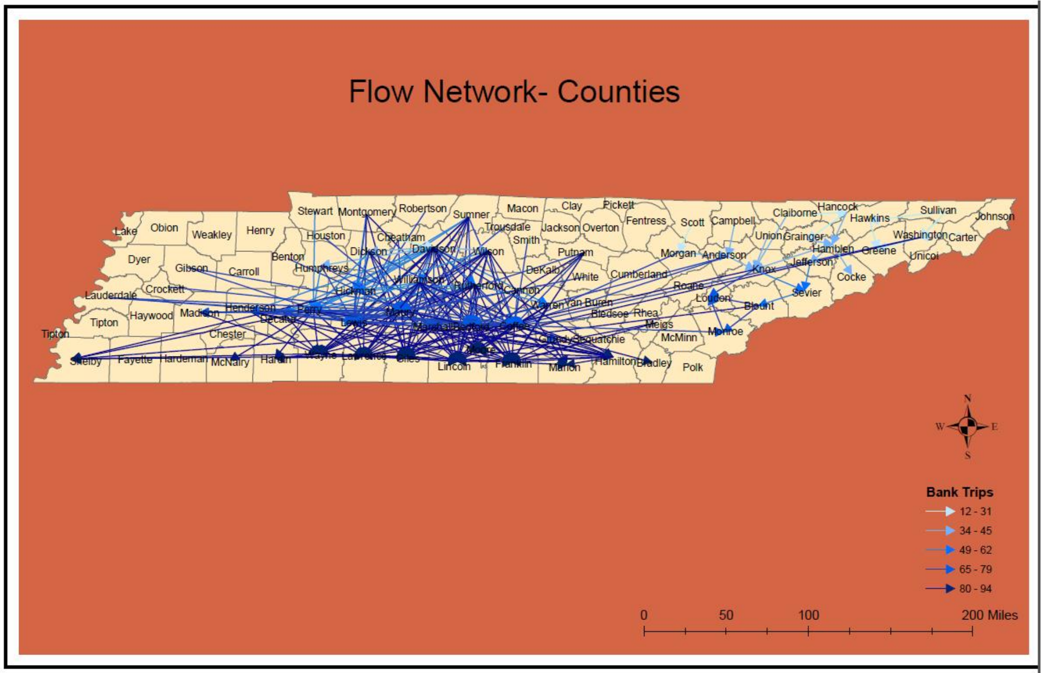

Figure F1: TN Bank trips

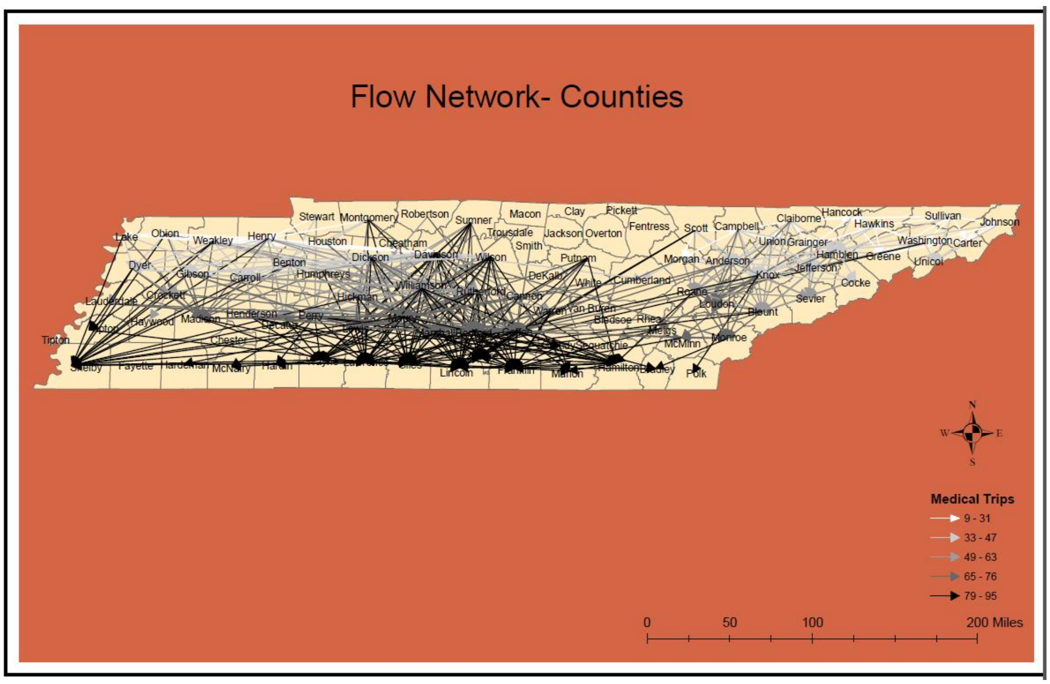

Figure F2: TN Medical trips

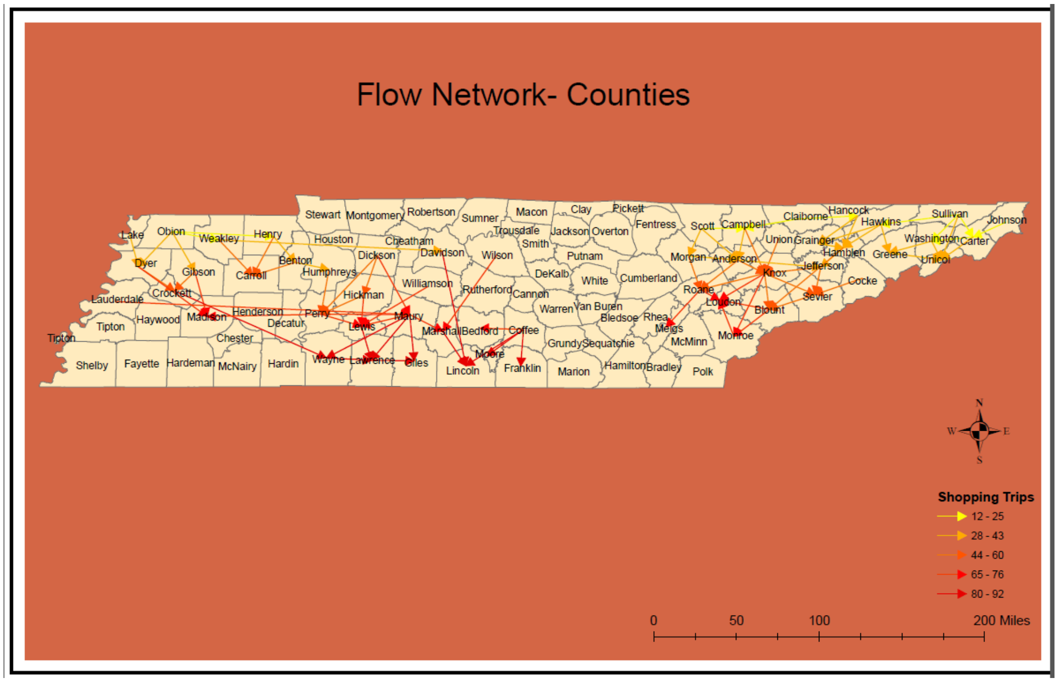

Figure F3: TN shopping trips

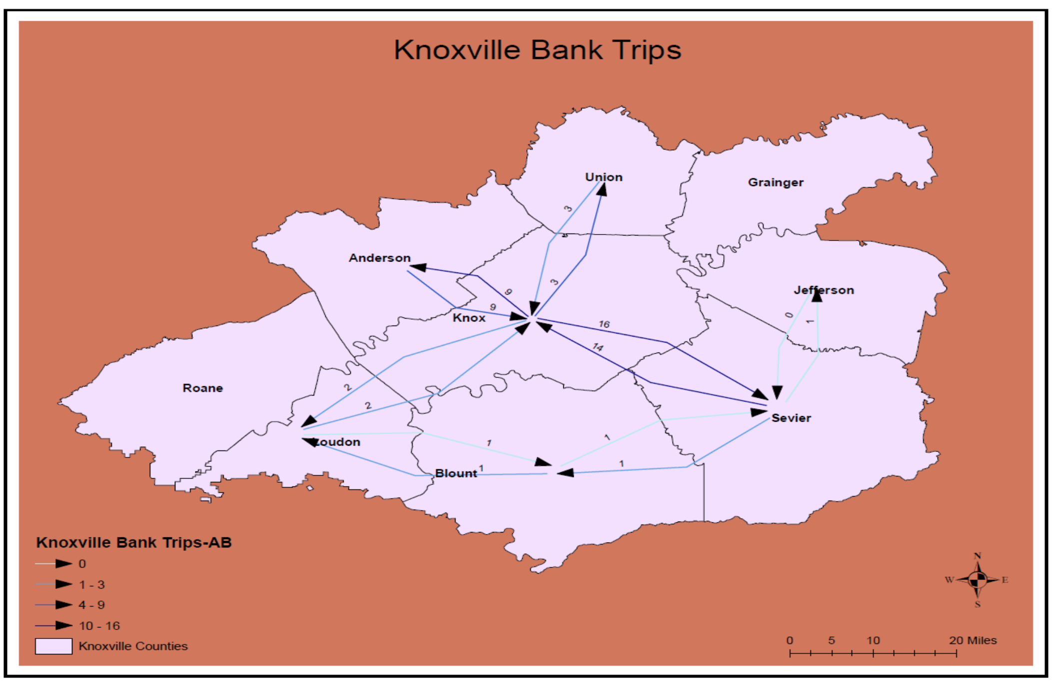

Figure F4: Knoxville bank trips

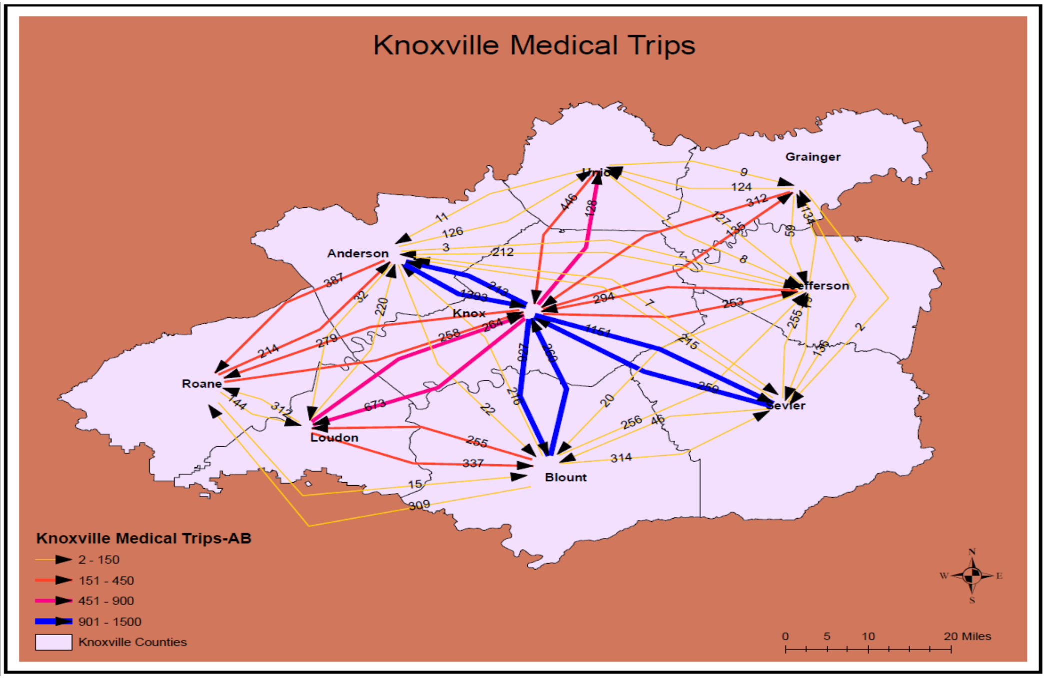

Figure F5: Knoxville medical trips

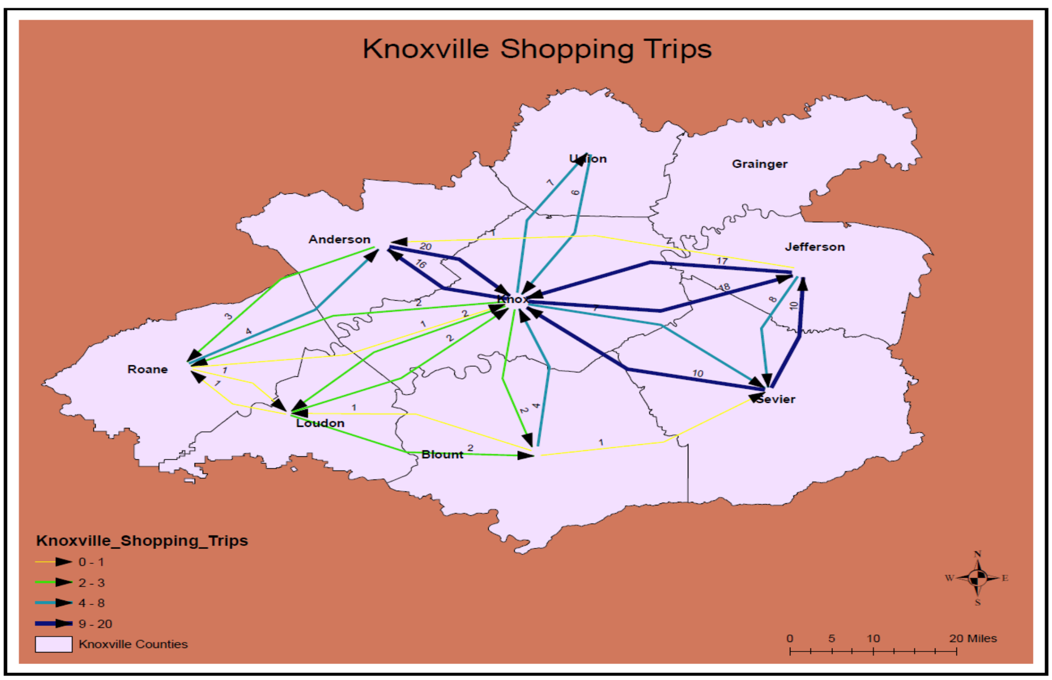

Figure F6: Knoxville shopping trips

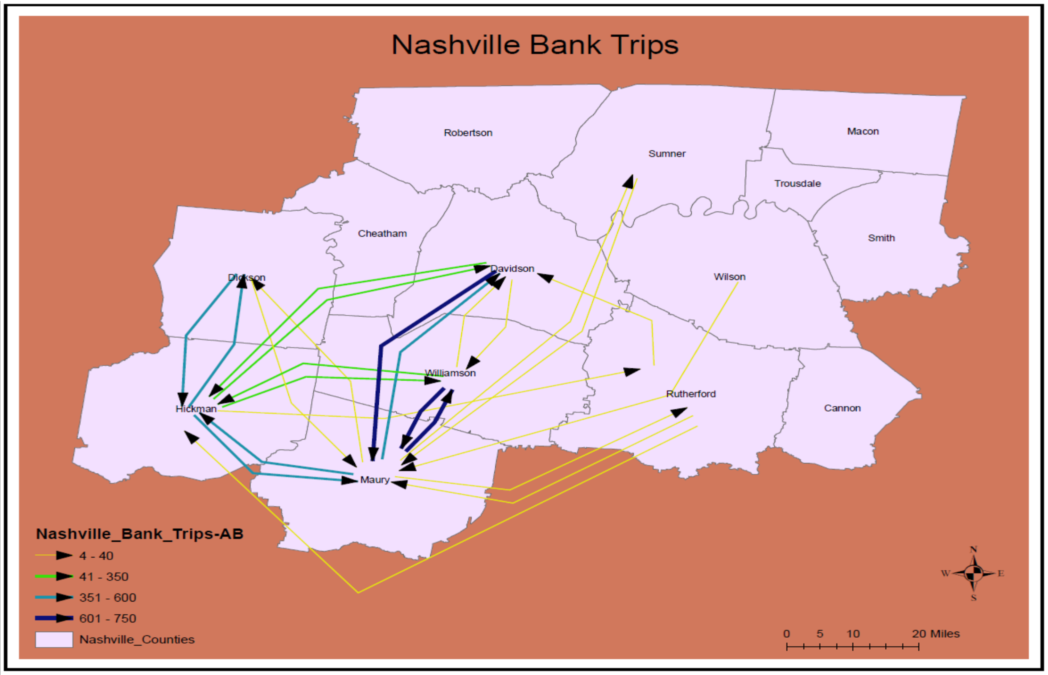

Figure F7: Nashville bank trips

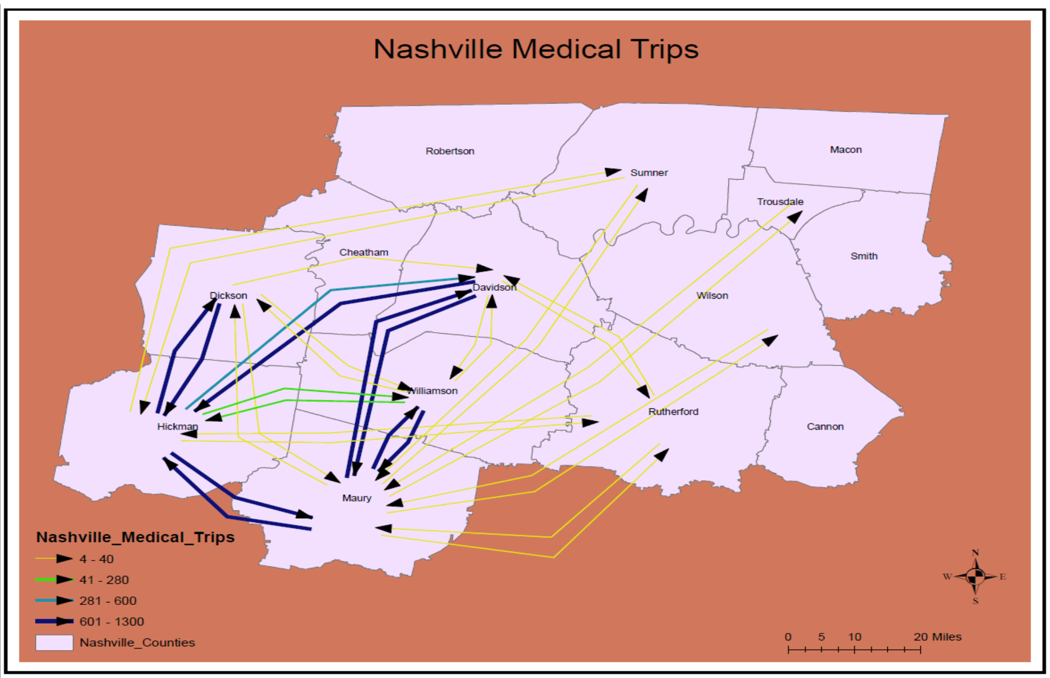

Figure F8: Nashville medical trips