Network Accessibility and Population Change: Historical Analysis of Transportation in Tennessee, 1830-2010

by

Ruth A. Zipfel

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

August 2015

DEDICATION

I dedicate this thesis to my husband, Joshua Smotherman, and my children, for sacrificing so many hours of time while learning patience through the moments of chaos this journey has taken us on. Secondly, to my parents, John and Cheryl, who know more than anyone the perseverance that has brought me to this point in my life and for always being my biggest support system. For you all, I am forever grateful.

ACKNOWLEDGMENTS

I would like to thank my chair, Dr. Jennifer Swift, whose guidance and encouragement made this thesis possible. I also extend full gratitude to my other committee members, Dr. Robert Vos and Dr. David Rigby, for your incredible insight and support in making such a huge concept a reality.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This thesis examines how potential accessibility (A(P)) through transportation over the time span of this study (1830-2010) affected population growth by county for the state of Tennessee. It focuses on shifts in transportation networks from waterways to rail, and rail to roadways, using decennial census data and likewise temporally adjusted county boundaries. The span of this study was broken into four individual time periods to best measure major transitions in transportation: waterways (1830 - 1850), railways (1860 - 1920), historic roads (1930 - 1970), and modern roads (1980 - 2010). Potential accessibility, which was anticipated to have influenced the population change taking place within the state over time, was measured using Esri ArcMap geographical information system (GIS) and a series of network datasets. Calculations of population sums, geographic measurements, and network accessibility were accomplished using both Microsoft (MS) Excel and Esri ArcMap. Linear regression modeling was performed using Statistical Package for the Social Sciences (SPSS). The results suggest that the variable influence was dependent on the study period, and although conclusively correlated at times, other variables in addition to or other than transport accessibility also proved significant in several of the study periods. Specifically, the waterways study period showed a direct correlation with the population growth and transport networks during this time, though additional variables could have contributed to population change as well. The railway network did not significantly contribute to population changes going on during this time, likely directly related to the onset of the civil war which hindered the development and growth of this transport system. While starting population share proved to be significant, with higher growth in counties that started out with larger populations, again additional variables could help explain population growth during the railway study period. Potential accessibility and starting share collectively explained almost 90%

of the variance within the historic road model, proving significant and likewise leaving very little of the change in population unexplained during that time period. Oddly, while the potential accessibility was significant, unlike theorized within this study counties less accessible to transport networks actually grew more quickly than those with higher accessibility. Finally, modern roads were found to be significant in population change as well and highly correlated. Additional steps to improve on this study in the future would include considering connections outside of the state, particularly in non-Tennessee peripheral localities with high populations. Secondly, investigation of additional variables such as economic data over a shorter overall time span, or using dasymetric allocation methods, could also provide further explanation behind population changes taking place over time.

CHAPTER 1: INTRODUCTION

Human migration has occurred in the United States throughout history, from early nomadic tribes to the present day. Understanding the forces that drive migration and population growth, from socioeconomic influences to topography and accessibility is crucial for numerous fields of study, included but not limited to urban development, cultural and natural resource conservation, and economic development and planning. In addition to contributing a key role in shaping the present-day cultural geography of the landscape and political boundaries, accessibility has also limited development in specific regions based on transport networks in the past (Wellman 2014). Discerning the processes involved in migration is vital to government officials charged with making sensible economic and political decisions, community planning, considering cultural perspectives and adaptations, and to academics and others making contributions to future research in numerous fields of study. Furthermore, having a firm grasp on the both natural and anthropogenic factors influencing migration in the past is essential in shaping and understanding historical perspectives (Trotter 1991).

1.1 Area of Study

The state of Tennessee, the chosen project area for this study, is located in the southeastern region of the United States (Figure 1). This state was chosen as the area of study for several reasons. To begin with, Tennessee has a longer available study period than half of the present day states in the U.S. This thesis study begins in 1830, when there were only 24 states existing in the Union. Secondly, it is a diverse and integral part of the history of the American South as well as having a contributing transport network affected by a major war. Furthermore, it lies in a prime location for migration from east to west towards what was for much of the period explored here the expanding Western Territory (what is now the western United States). Finally, inland regions such as the Midwest and Southeast have been relatively scarce in terms of heavy incoming populations (particularly immigrant concentrations), thus have not been the focus of heavy research in regard to human migration (Portes and Rumbaut 2006).

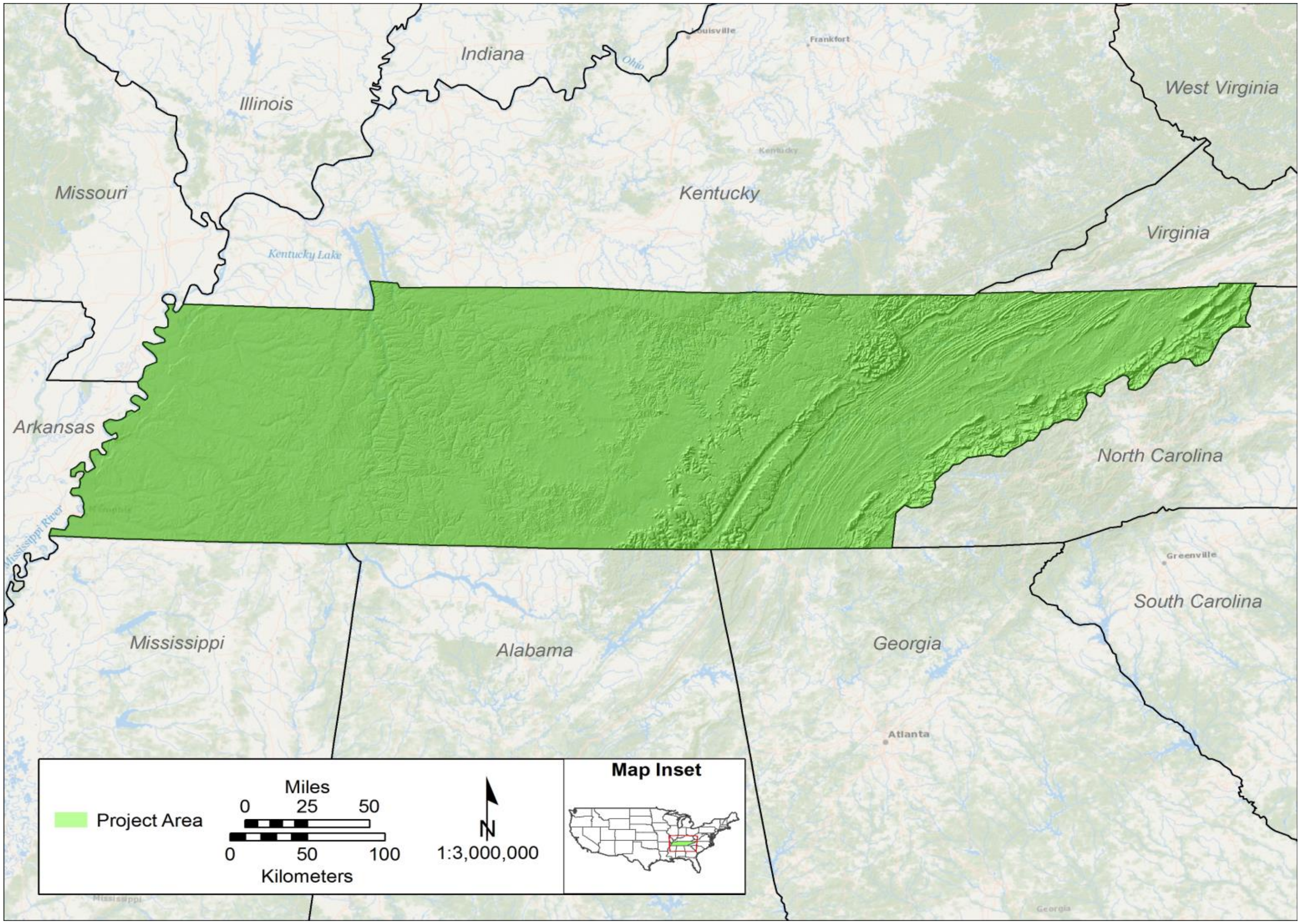

Figure 1. Map of the Project Area: the State of Tennessee

This study uses transport networks to determine potential factors influencing population changes in the state of Tennessee by considering likely variables that would contribute or inhibit these demographic fluxes (i.e. accessibility and starting population). Additionally, this study aims to identify how transport networks through time have heightened or weakened population densities and influenced the spatial geography that exists today throughout the state. The data covers the time span of 180 years, using county boundaries at a statewide scale and following a temporally chronological format. The correlation between independent factors (e.g. accessibility and economics) and population change is analyzed in order to identify these influences. A final goal of this thesis was to develop the historical dataset for the state of Tennessee that was be used to visually assess the changes in population. The process of developing the geodatabase and maps may prove useful for future studies in other regions using similar approaches.

1.2 Research Statement

Transport networks are generally related to economic growth and the validity of this connection has proven significant in multiple studies within economic geography (Krugman 1991; Pavlínek 2003; Zondag et al. 2014; Yuwei and Pengfei 2013; Ding, 2013). Krugman (1991) specifically discussed the idea of spatial turn in economic geography, suggesting a pattern of both the core and periphery of urban areas defined by manufacturing status, transportation costs, and likewise economic scale dependencies. Yuwei and Pengfei (2013) look at regional economic growth in several cities in China by expanding on the traditional neo-classical model in order to assess transport accessibility and externalities, showing that inner cities were influenced by transport networks based on differing agglomeration factors and developed infrastructure. This is particularly true for growth between these inner cities and peripheral outliers, but still indicative of the importance of the transport network overall. Likewise, economic data was considered as a potential variable within this study, but due to inconsistent data, data restrictions, and enumerator category changes, contiguous data throughout the entire study period was not available for the entire region and across all included counties, and therefore population was used as a proxy in order to measure economic growth in the absence of this data.

Specifically, this thesis is an attempt to replicate the case study conducted by Kotavaara et al. (2011) in Finland, albeit in a different location and geography, and over a much longer history. Kotavaara et al. (2011) analyzed accessibility in Finland by generating a network dataset with a GIS of railways and roads, using census districts and rail stations as destination points in a series of generalized additive models (GAMS). This thesis applies a similar framework, aiming to identify which components most heavily impacted migration in the state of Tennessee over what time periods between 1830 and 2010: accessibility, transport potential, or distance to large cities? In the context of this study, potential accessibility, referred to simply as accessibility in this thesis, is defined as the measured potential of traveling from a given county to all other counties within the state (Bugromenko 2010) with the population weights of the counties built in and transport as a type of network that moves populations from one location to another (Park and Allaby 2013).

Early in 19th century, waterways controlled movement of commodity and migration to the integral towns and establishments at the time, particularly after the Cherokee Treaty in 1791 (Cherokee Nation 2015). Topography influenced development and population growth, particularly in early decades when ease of development and roads were essentially non-existent. In the early 1900's when mechanization began to become established, transportation moved from waterways to railroads along with significant roadways and became the primary mode of public commute, opening a whole new world for an influx of migration highways (New Deal Network 2003). It is expected there was growth that correlated with the connectivity of these transportation networks, but that outward population growth also developed within a given distance.

This thesis applies two analytical approaches to address the above question using Esri's ArcMap version 10.3 Geographical Information System (GIS): The first approach was calculating population for every given county within the state using a decennial sequence over ten years and examining it along with transportation (i.e. roads and railway) to determine the contribution of connectivity and the development attributing to accessibility and how it relates to population change. Further, this method will also employ a hydrology layer calculated at a given range to assess migration movement early in the time series when commute was primarily done on the waterways (New Deal Network 2003). A timeline was generated using spatially relevant information in the literature, to determine if socioeconomic events might have been the dominating force behind migration movement and current ethnic enclave concentrations.

This thesis first begins with reviewing ideologies behind migration and transportation theory, and then discusses research conducted in relation to populations and transportation using a GIS, continues with a brief assessment of migration within Tennessee since colonial times, and finally reviews studies of existing GIS-based transportation demographic analyses for evaluating population fluxes and potential accessibility. This research provides a newly generated dataset for population change in the state of Tennessee from 1830 to 2010. The resulting dataset is geared toward the public as well as academic researchers and government entities that wish to undertake demographic mapping within their own state(s) for private or future planning needs. This dataset can potentially be further used to predict future migrations patterns in Tennessee.

1.3 Motivation

Demography and transportation are highly researched topics and heavily debated by politicians and scholars alike. The reasons for the intense research and debate stem from political and economic agendas regarding human and civil rights, cultural diffusion and identity, legalization and classification, trade and labor economics, and homogeneity verses heterogeneity, to name but a few. This emphasizes a need to maintain consistent and historical spatial documentation of where these changes are taking place and analyze reasons behind population changes. Ease in visualizing these changes is one of the main impetuses of this research.

CHAPTER 2: BACKGROUND

This chapter first reviews the historical and socio-economic influences during the span of this study within the state of Tennessee as well as the history of transportation networks on which this study is based to distinguish particularly notable time frames that vary quite extensively by length. The next section comprises a literature review of other work conducted in transport geography, including both potential accessibility measurements and transportation analysis topics. This chapter will then conclude by synthesizing these influences and past research to set the stage for the methodology developed to carry out this this study.

2.1 A Brief History of Tennessee

Early in the 18^{th} century, primarily the native Cherokee and the Chickasaw nations occupied Tennessee (Sturtevant 1966) (Figure 2). The Shawnee had also lived in the Cumberland Valley that is now the city of Nashville (previously Nashboro), and the surrounding areas until the larger two tribes drove the population out. Several French explorers found their way down the river and into the western portion of Tennessee, while Europeans integrated through the Appalachians from the east (Corlew et al. 1990).

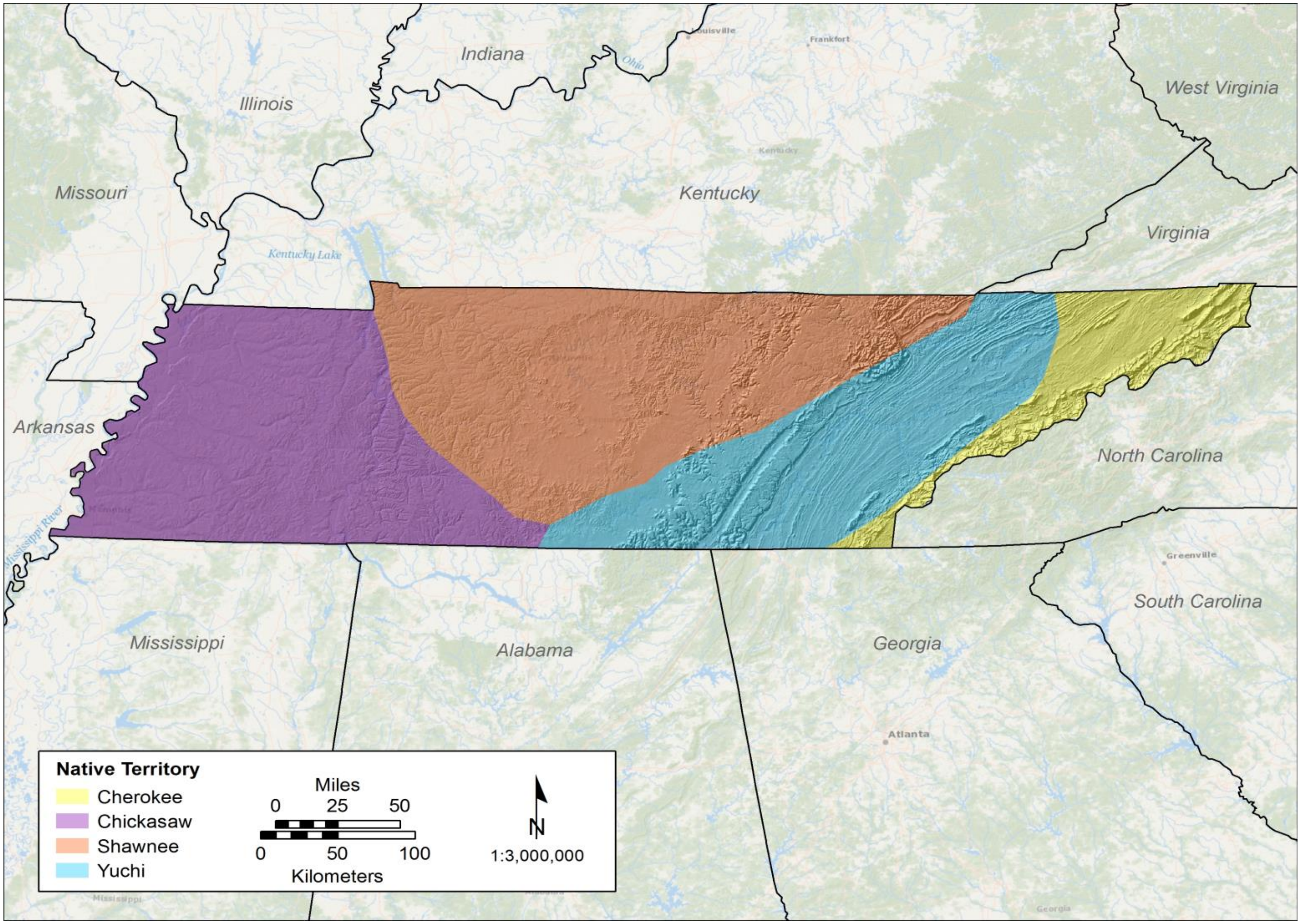

Figure 2. Map of Native American Territories at 1700 A.D. in Modern Day State Boundary (Sturtevant 1966)

As trade and colonization expanded, more immigrants (European-whites) came into the Tennessee region. Military forts were constructed, and animosity soon arose between the newcomers and the natives as land ownership was contested (Rothrock and Smith 1973). In the latter part of the 18th century, pioneers of mixed European decent expanded southwestward into what is today Tennessee, and it joined the Union in 1796 (U.S. Immigration Support 2011) with a population of around 77,000 (Figure 3). During the 19th century, European pioneers expanded into western Tennessee, which at the time was mostly unoccupied (Rothrock and Smith 1973). With the war of 1812 and subsequent conflict with the natives, Tennessee became a battleground for land rights. Most of the natives either succumbed to the white migrants or died both defending the land and during their forced move west (Bowes 1973).

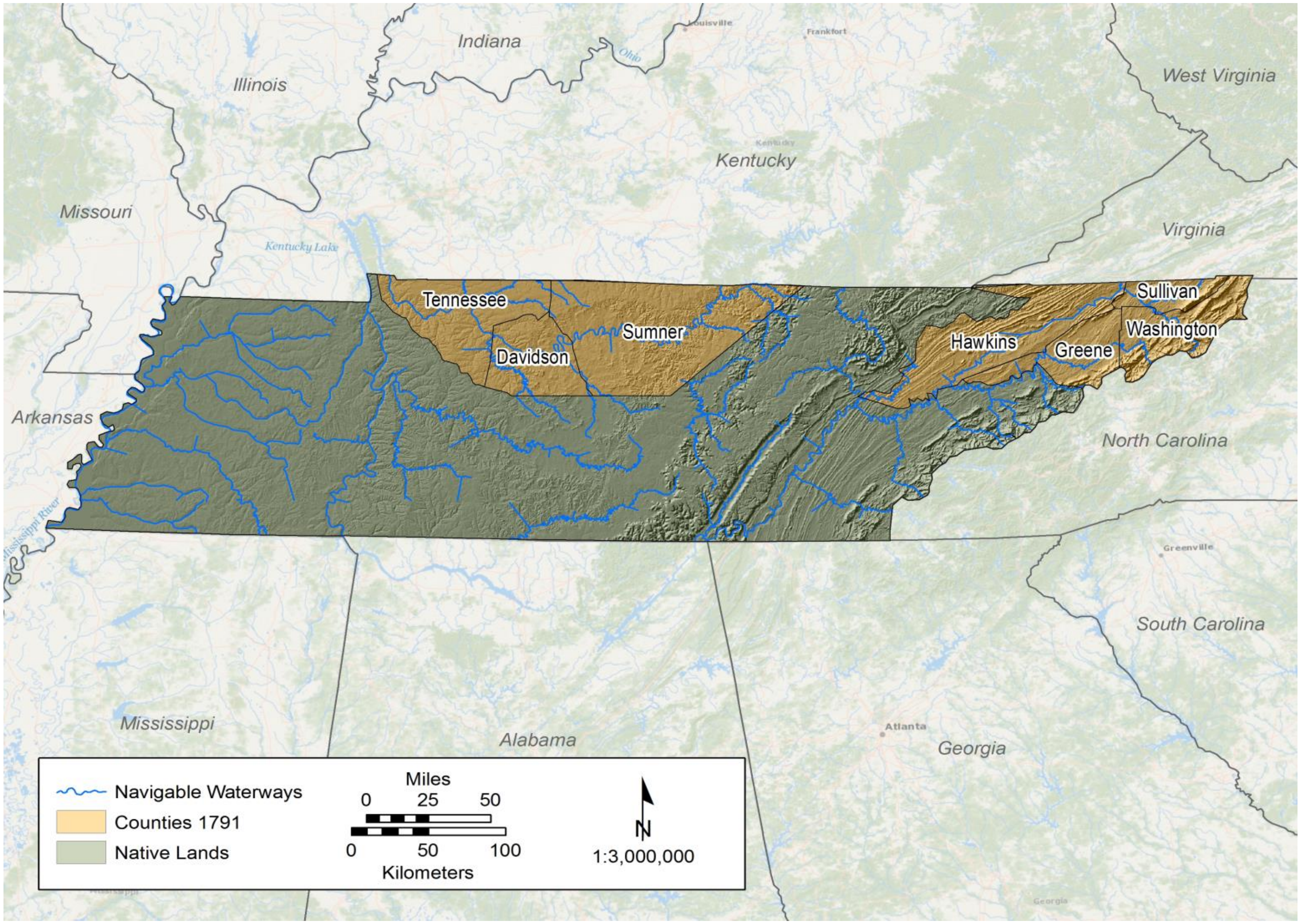

Figure 3. Map of Tennessee Just Prior to Joining the Union in 1796

Early transportation was a problem, with waterways being the common form of travel in the early 1800s. First settlements, and subsequently some of the largest cities in the state today were settled based on navigable waterways (Figure 4)(e.g. Nashville, Memphis, Knoxville, and Chattanooga). The excellent transportation along the river networks contributed to a higher rate of settlement in these years along the Cumberland River than was taking place in eastern Tennessee in the Appalachian region (Corlew et al. 1990), although these two specific regions were first colonized. As shown in Figure 4, these early settlement areas, depicted here as the first Tennessee counties in 1791, fall along the most accessible and strongest areas of these regional waterways, specifically those areas derived from bordering state populations. It was soon around this time that interests' arose regarding railroad development, although construction did not begin (and very briefly at first) until the late 1830s. The first railroad was not completed until the mid 1850s (Johnson 2010). Railways sprang up quickly thereafter, with a track distance of 1,197 miles by 1860. By this time Tennessee had risen to a population of approximately 1,110,000 (Rothrock and Smith 1973). With the onset of the Civil War in 1861, Tennessee found itself very divided politically. The state became very torn during the war due to split opinions about secession from the Union (Majors 1980) (Figure 5). Slave-run plantations were located primarily in western Tennessee, with fewer being in central Tennessee, and hardly any to the east (Rothrock and Smith 1973).

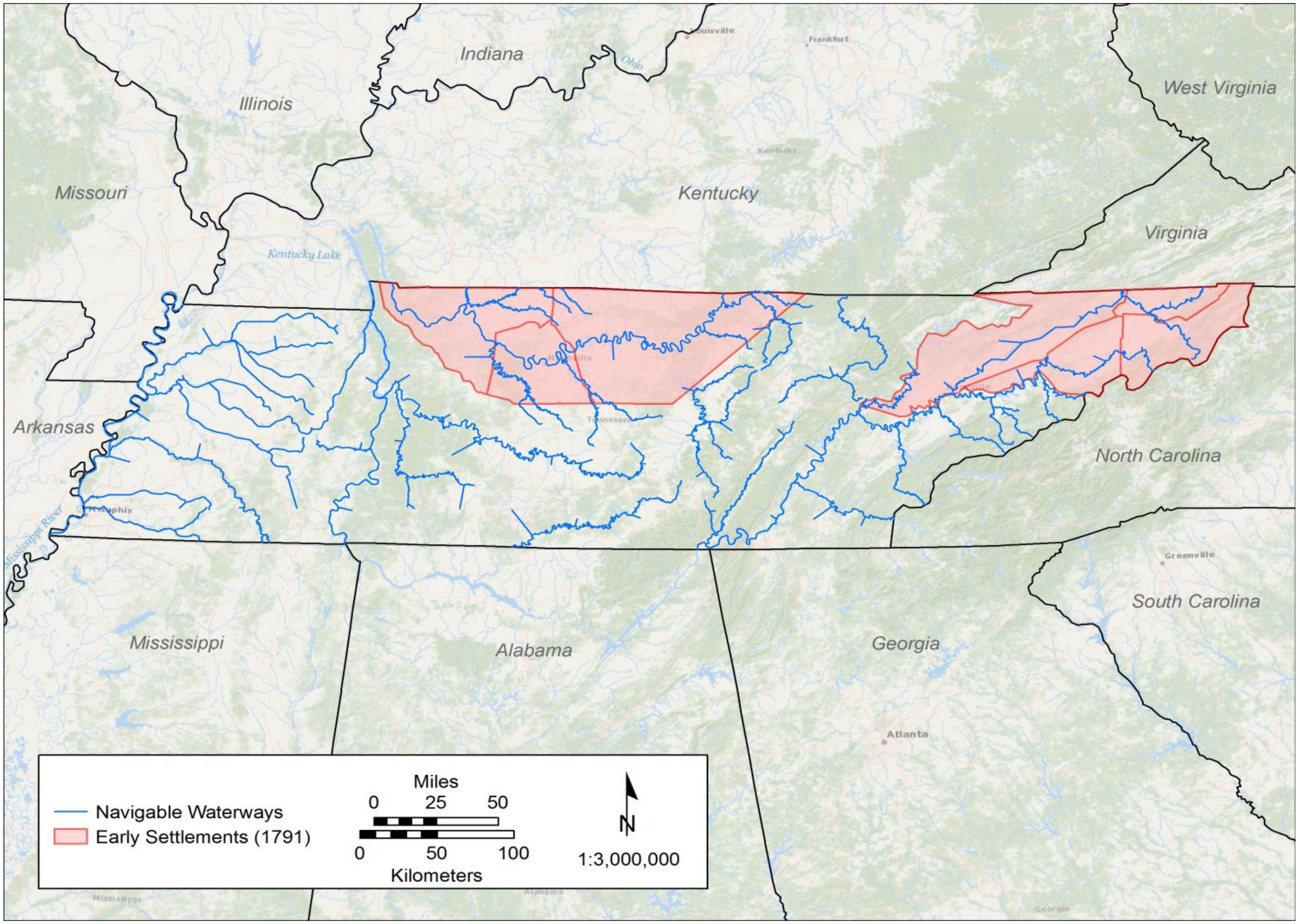

Figure 4. Map of Navigable Waterways in Tennessee and First County Boundaries

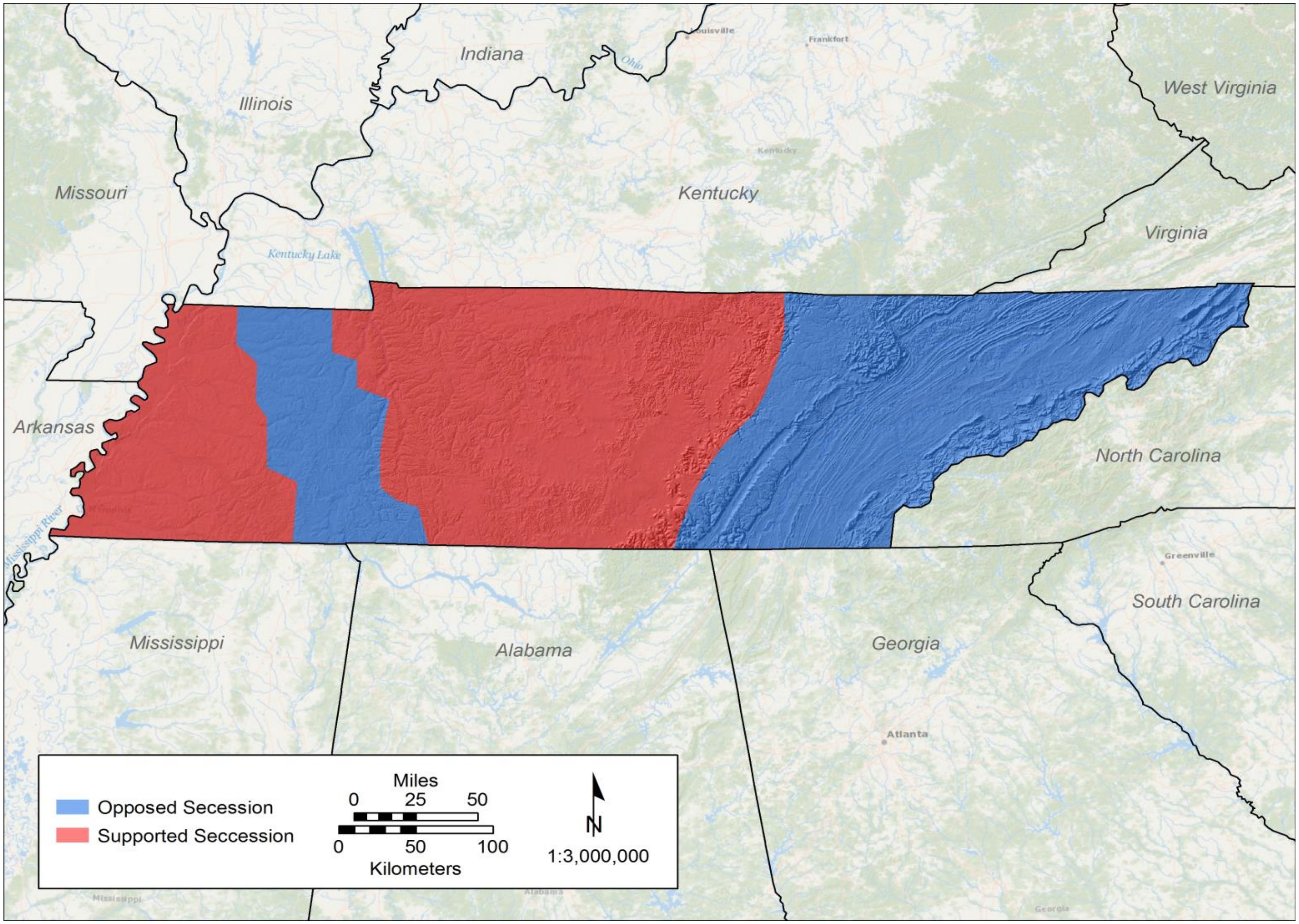

Figure 5. Map Depicting the Split of Tennessee During the Civil War (Encyclopedia Britannica Inc. 2012)

Railroads were used as instruments of communication and spatial movement of supplies and soldiers during the Civil War, leaving excessive track damage as well as diminished resources from the post-war finances into the following decade. The economy eventually improved and re-expansion began during the 1880s. By 1900, Tennessee reached a total of 3,131 miles of in-state track with the most impressive depots springing up over the next two decades, most notably in Memphis and Nashville (Figure 6) (Johnson 2010). The federal government took control of the railroads during the First World War. In 1920 when the control was relinquished, financial hardship set in. In subsequent decades the railways saw a prolonged decline, primarily when interstate road networks began to progress into the state (Johnson 2010) as well as the development of the Dixie Highway Association and the Tennessee Good Roads Association (Sharp 2010). Every county in the state was eventually connected with a complex paved-roads network, going from a meager 244 miles in 1923 to more than 4,000 miles by 1930, transforming Tennessee entirely with trade and tourism (Pierce 2010). Although railways continue to be used into the modern era, it is primarily for industrial transport with little to no passenger use (Figure 7) (Johnson 2010).

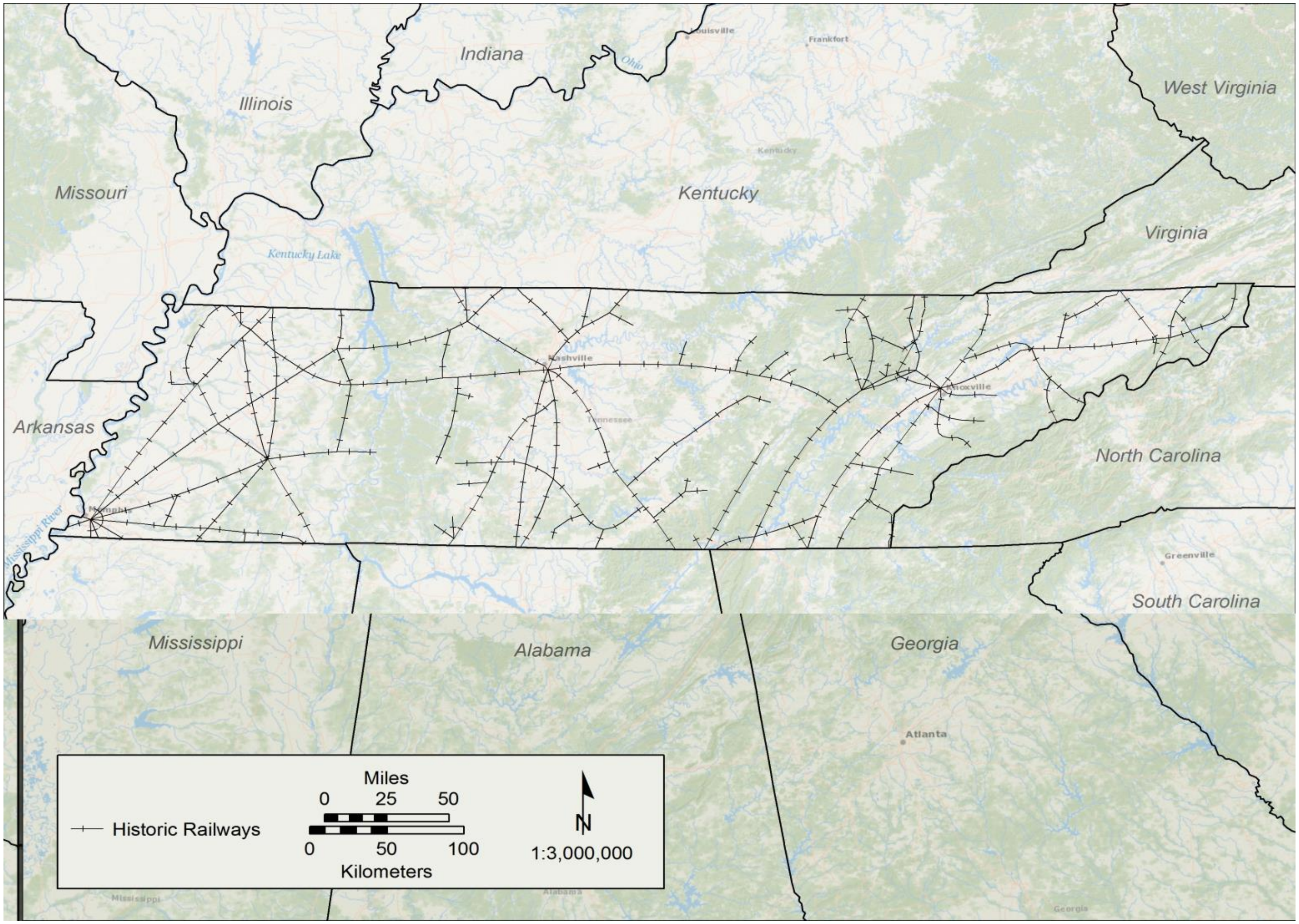

Figure 6. Historic Railway Networks in Tennessee (1856-1930) (Leahy 1934)

Figure 7. Modern Railway Networks in Tennessee

Eisenhower pushed for improved roads during World War II, believing it would strengthen the nation against postwar economic failure. Construction for such an expansion finally began in the mid 1950s once Eisenhower became president and the Federal-Aid Highway Act of 1956 was passed, although there was considerable roadway development already in existence at this time in Tennessee (Figure 8). Major interstate construction in Tennessee began in 1957 with a section of new I-65, and I-40 through the longest stretch of the state soon thereafter. I-24 followed in the 60s, extending over the Appalachians in the east through Monteagle and down into Chattanooga, the rugged terrain presenting quite a challenge for the crew involved. A bridge was eventually built in Memphis during the 1970s spanning the Mississippi River into Arkansas along I-40, a major highway that presently extends from North Carolina to California. Highway development continued steadily into the present day (Figure 9), overseen by the Tennessee Department of Transportation (TDOT) and the Federal Highway Administration (Sellers 2010).

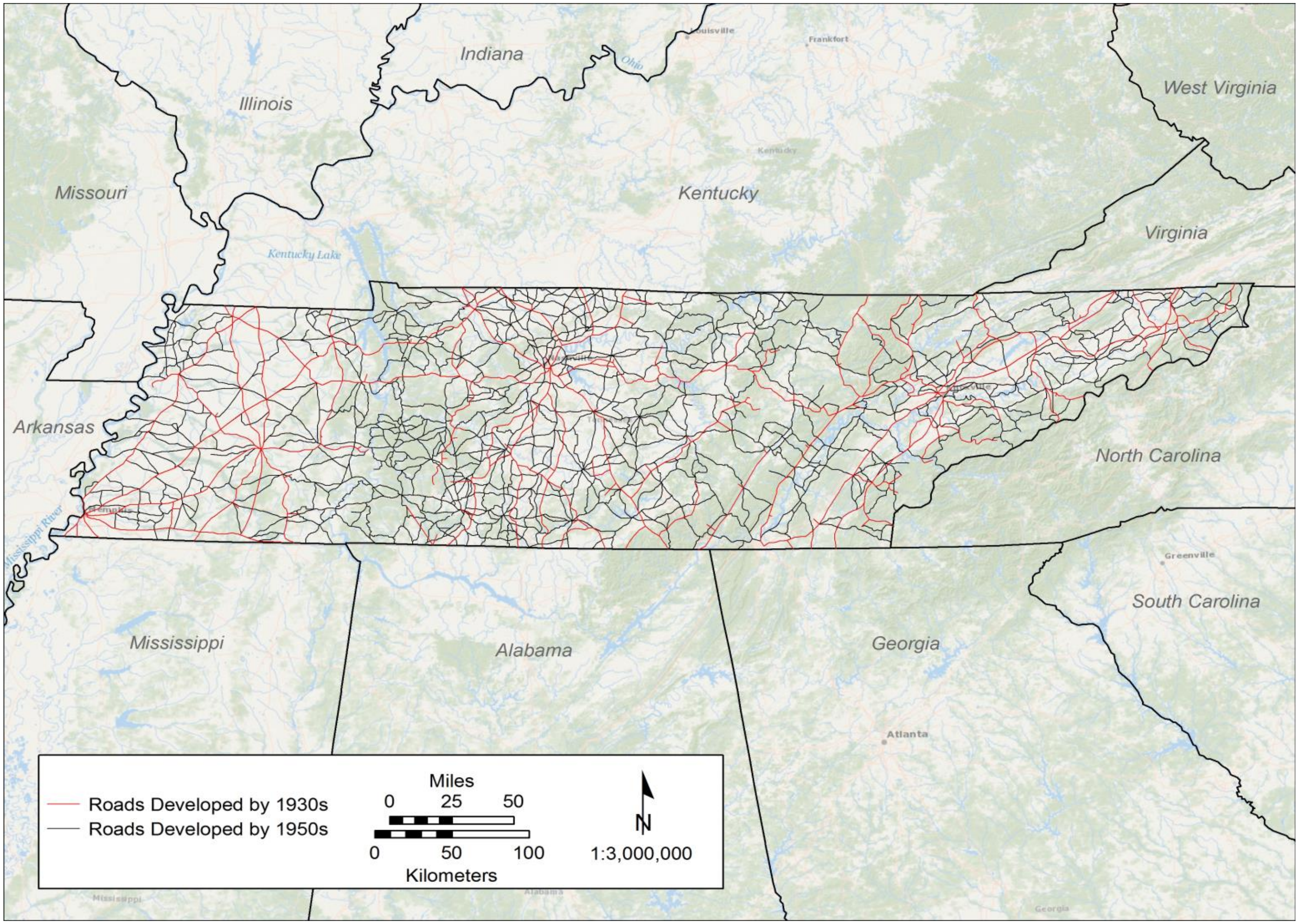

Figure 8. Historic Roadway Networks in Tennessee

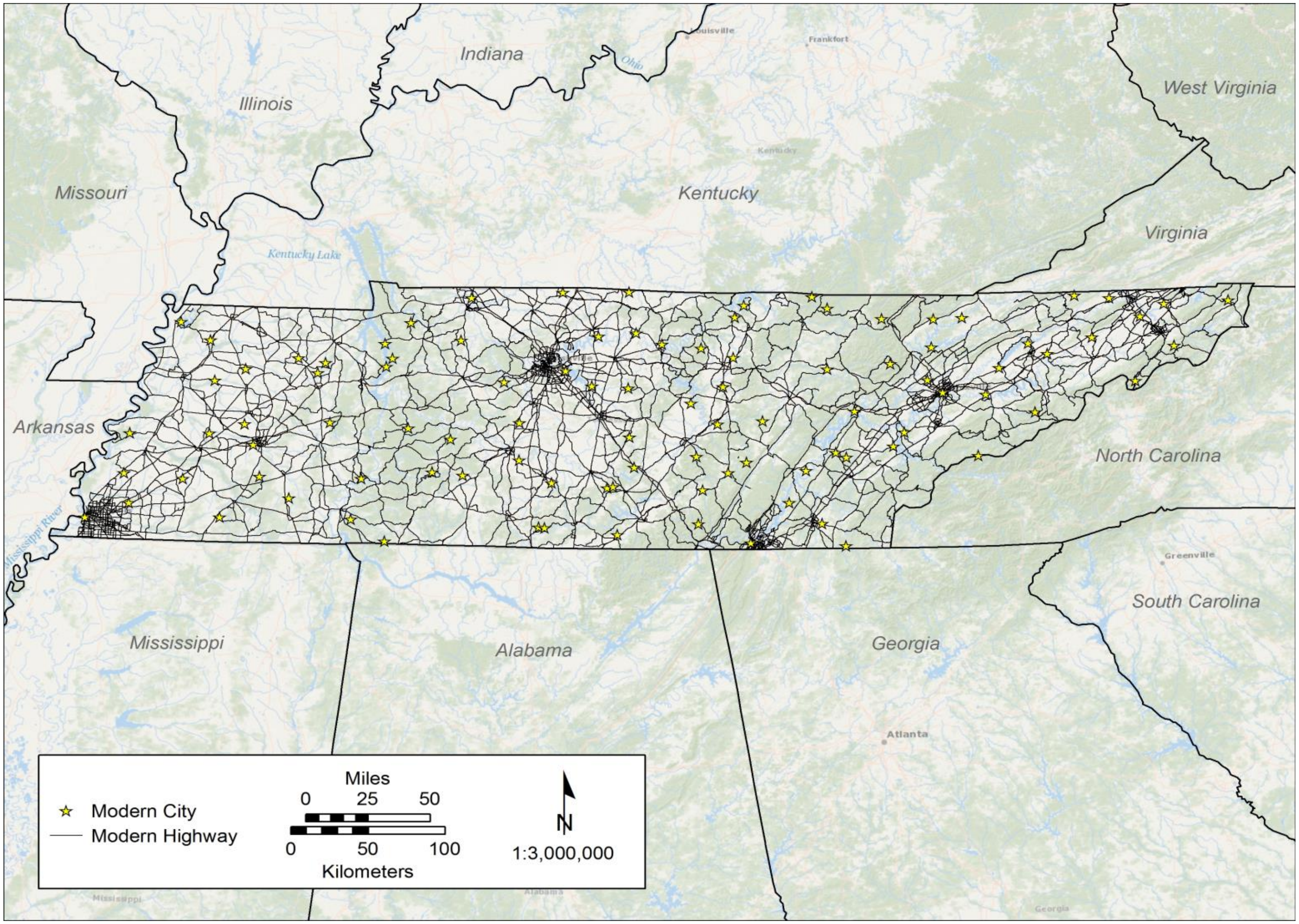

Figure 9. Modern Major Roadway Networks in Tennessee (1980-2010)

2.2 Transportation Geography

Bugrimenko (2010) discusses a common debate among geographers and social scientists pertaining to cause and affect of population trends and transportation potential: do these networks cause the flux of movement, or does the flux instead create the network? The chosen response generally determines approaches to both scholarly study and planned infrastructure if there is to be a general goal determined. Although transportation geography was initially defined as the potential of movement from destination to origin in early studies by W. Hansen (1959), it has since expanded to include additional components of study. Accessibility is generally measured using either land-use or transport models, or a combination of the two (Wang et al.). 2014 Zondag et al. (2014) emphasizes the importance of land-use and the distribution of the accompanied activity that comes with it, indicating many efforts ignore the effects transportation has on them.

Bugrimenko (2010) notes three key components of what he calls the “triad” of transportation planning, all essential to population location theories: flows, mobility, and accessibility. Additionally, land use also shapes transportation patterns and determines route locations, destinations, and often the frequency of trips to given destinations as well (Zondog 2014). Many approaches have been taken to determine the most accurate and effective way to model and measure potential accessibility for planning, predictions, and analytical purposes (Yigitcanlar et al. 2007; Wang et al. 2014; Lopes et al. 2014; Huang and Wei 2002; Zondag et al. 2014; Kotavaara et al. 2011; Kotavaara et al. 2012). In contrast to this study, most of these studies are done on a city wide, urban scale. This comprises a relatively smaller scale analysis in terms of area than this study, which encompasses an entire state.

Yigitcanlar et al. (2007) created the Land Use and Public Transport Accessibility Indexing Model (LUPTAI) to model land use in Queensland, Australia. Their model was designed to measure accessibility at a local scale to destinations such as heath care, shopping, or employment by means of walking and/or public transportation. GIS was used and included a pedestrian and transport network, the latter including railways, bus routes, and ferry destinations. They considered frequencies as well as trips taken in their analysis in order to associate peak travel times with specific accessibility needs, such as places of employment. The pilot study took place on the Gold Coast (Southeast Queensland) in order to test their model in a major urban location. They succeeded in generating a mappable index and demonstrated a feasible accessibility model while noting that the model lacked specifics such as determining proximity and economic affordability (i.e. social class). This is important to consider because economic inequality contributes to accessibility inequality, where the wealthy have advantage to travel affordability over the poor, therefore having a higher sense of control of space and time (Grengs 2014).

Wang et al. (2014) used accessibility indicators to link transport and land-use data in what they termed a land-use and transport interaction (LUTI) model. A Metropolitan Activity Relocation Simulator (MARS) model was chosen to determine variable relation and measurement of cause and effect between these submodels. The case study took place in the Madrid Region in Spain, and by using a larger scale analysis the authors were able to apply walking, vehicle travel, and public transportation in the model along with exogenous variables (i.e. demography, economics, and land use). Based on a regional analysis, they were able to calculate first a Potential Accessibility (PA) and additionally an Adapted Potential Accessibility (AA) using competition results to approximate job opportunities.

Lopes et al. (2014) conducted a study in Porto Alegre, a city in Rio Grande do Sul, Brazil to specifically analyze various methods used for spatial regression modeling in transport planning, noting apparent gaps in this field of study. They hypothesized the importance of using spatial effects in regression modeling in an attempt to determine if the inclusion of these spatial variables indeed improves upon models analyzing transportation demand. Using available data from 1974 and 1986, a handful of models were tested. The best outcome came using the Alternative, Global, and Local-74 (AGL74) model that introduced specific variables. Alternative models were defined as "*spatial regression models or regression models including spatial variables*" (Lopes et al. 2014). These models had better performance and outcomes over traditional models, thereby proving the importance of including spatial dependence when using regression analyses and supporting the usability of linear regression modeling for transport analysis.

Huang and Wei (2002) focused specifically on low-income residents of Milwaukee, Wisconsin who have a higher dependence on public transportation. They created a transportation network in a GIS using bus information from Milwaukee County Transit System (MCTS). Distance was assessed using Esri's Arcview and service capability calculated based on total trips along individual bus routes. They examined the connectivity between low-income locations to bus routes, hence low accessibility for a large portion of the population, hypothesizing that a "spatial mismatch" was to blame, meaning that mass transit only really happens in direction, from urban development to inner cities. Their results supported this theory, but further reinforced the notion that transport is important for both population and economic growth.

As already mentioned, this thesis replicates some of the work done by Kotavaara et al.

(2011), taking on a similar sampling approach and applying potential accessibility calculations, adapting the methods used for the geographic landscape of Finland into four applicable models for the state of Tennessee. A key difference to note between these study cases is Finland's rather secluded locality, particularly for migrant accessibility, in contrast to Tennessee's extreme connectivity to mostly passable terrain. Additionally, Tennessee is landlocked, unlike Finland, which also occupies a considerably larger area. Tennessee, at a current extent of 42,180 square miles, is roughly one-third the size of Finland, which has a landmass of 130,666 square miles.

There have clearly been numerous attempts made to measure accessibility and determine its influence on population trends but none of these studies capture the significant length of time that this study covers. Additionally, such a study has not been conducted in Tennessee, particularly spanning the entire state. Furthermore, most of these previous studies exist at a relatively small spatial scale, such as at city level (i.e. Huang and Wei 2002, Lopes et al. 2014). It should be noted that regarding cartographic scale, the size description (large versus small) describes feature appearance within the map and not the area covered. Thus, a small-scale map could cover a state or a country and a large-scale map could depict a city or neighborhood (Esri 2013).

The study are is extensive, encompassing an entire state with ninety-five currently existing counties, reviewed over a significant period of time. Much of this study is based on the work of Kotavaara et al. (2011) using similar types of transport network datasets and adjusted calculations for the relevant data for the state of Tennessee to measure potential accessibility in Tennessee. The following chapter introduces the methodology used for measuring the potential accessibility of transportation networks using the GIS software Esri ArcGIS and Statistical Package for the Social Sciences (SPSS) to determine population trends over time in Tennessee.

CHAPTER 3: METHODOLOGY

This main goal of this study is to determine likely causes behind population changes in Tennessee over the past 180 years. The overall analysis ultimately measures one consistent dependent variable, population change, and three independent variables, namely accessibility, county shares of the overall state population, and mean geodesic distance to the largest cities in the state throughout the study (Table 1).

Table 1: Study Variables

Variable	**Variable Type**
Population Change	Dependent
Potential Accessibility	Independent
Population Share	Independent
Mean Distance	Independent

A statistical approach was used to model accessibility, similar to the Finnish study (Kotavaara et al. 2011). Significant shifts in transportation networks were allotted to specific time periods to assess how transportation and accessibility may have affected population trends. The starting population share was included as a secondary variable based on the theory of path dependence, an approach to predicting population trends based on places with a head start in economic growth. As mentioned earlier, populations generally grow faster than places that may fall behind at the beginning (Pavlínek 2003). And finally, the average distance to the five highest populated cities (Nashville, Knoxville, Chattanooga, Clarksville, and Memphis) was calculated to assess the possible influence this variable may have had on population changes over time in Tennessee (Figure 10).

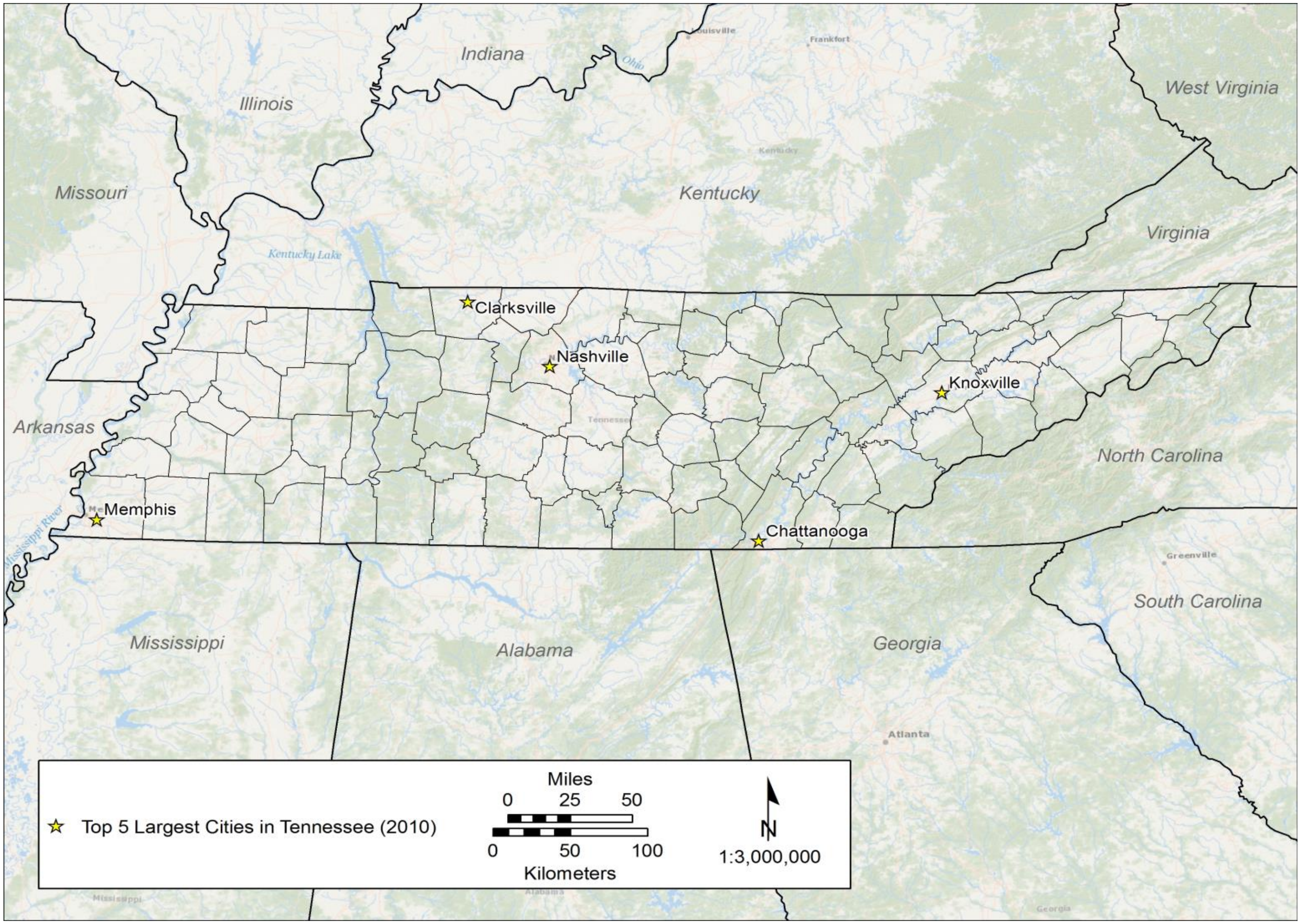

Figure 10. Five Largest Modern Day Tennessee Cities (2010) Used As Third Independent Study Variable

This last variable is included based on the theory of Tobler's first geographic law (1970), stating that although all things have a connection to all other things, those that are nearer have a stronger relationship, in order to examine whether network mode and/or accessibility truly matter or whether proximity has a stronger correlation.

3.1 Study Periods and Hypotheses

In this study, the Esri Network Analyst toolset and extension were used to calculate distance and relative accessibility of waterways, railways and roadway to population centers, centers being heavily populated counties, and their effects on population densities over time. Waterways were derived from a navigable waterway data source (National Waterway Network 2012). Individual vector datasets were digitized from historical maps for linear railways and rail station points (Leahy 1934), and historic roadways (National Map Company 1927, Rand McNally and Company 1927, State Farm Insurance Companies Travel Bureau 1940, Shell Oil Company 1956). Modern roads were selected and clipped for the state of Tennessee from an Esri (2013) shapefile. These datasets were converted into network datasets in ArcMap to assess potential accessibility and population change over time in Tennessee. This calculation included variables of both distance and change in the relative share of population in the state at any time (Kotavaara et al. 2011). Figure 11 highlights the steps taken to achieve the final transport networks.

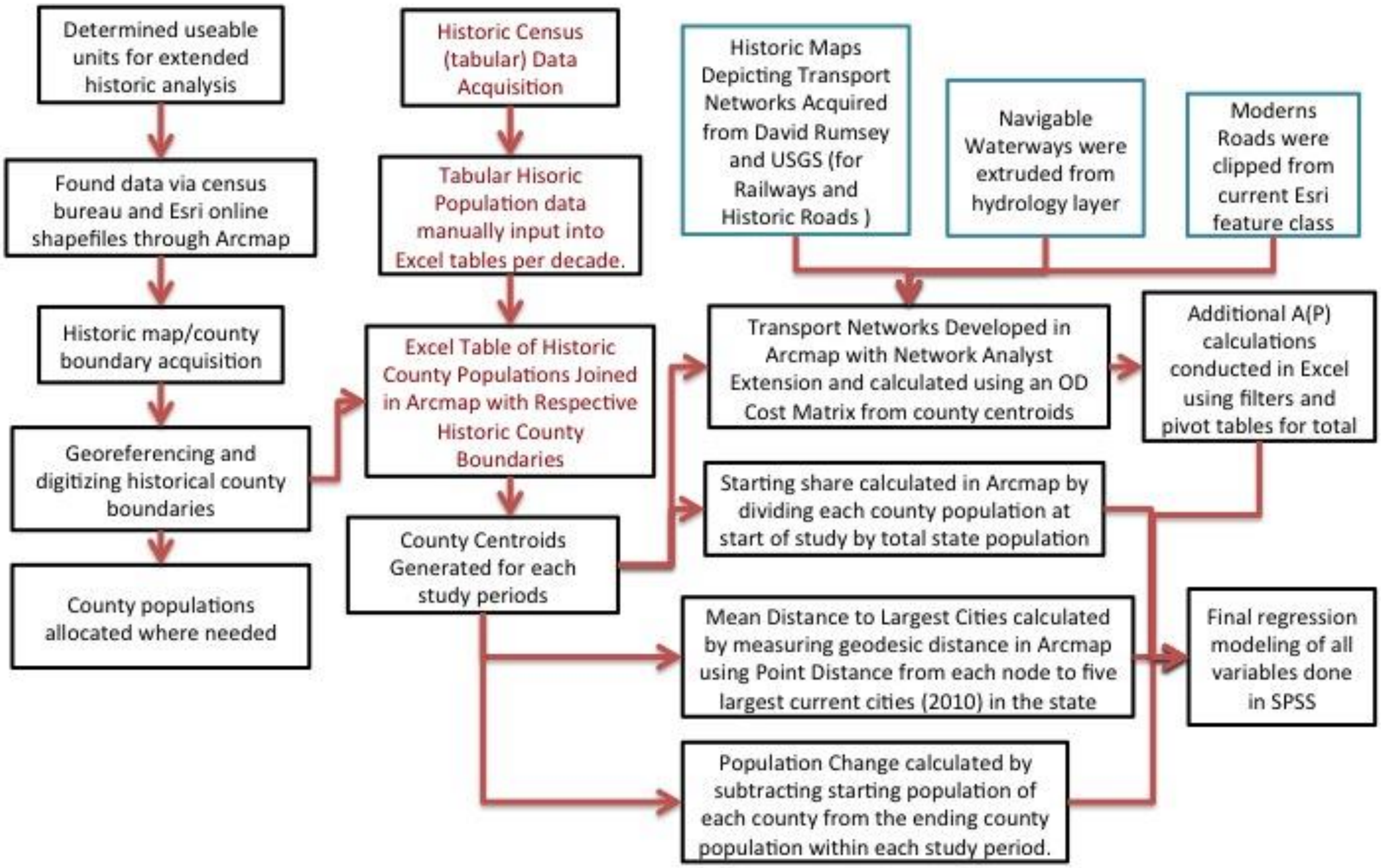

Figure 11. Diagram Highlighting Steps to Achieve Transport Networks

As previously stated, this thesis aims to identify the relationship between population change and transportation methods in Tennessee over a span of 180 years (1830-2010). A study encompassing such a long stretch of time resulted in the need to examine population movement in shorter, segmented study periods focused on the applicable mode of transportation over specific decade ranges (Table 2). The time spans were chosen based on dominant methods of transport gleaned from the literature review of the transportation development history of Tennessee (Johnson 2010, Sharp 2010, Johnson 2010).

Table 2: Study Periods Defined

Study Period	Time Span	Method of Transport
Period A	1830-1850	Waterways
Period B	1860-1920	Railways
Period C	1930-1970	Historic Roads
Period D	1980-2010	Modern Roads

This thesis further aims to identify which changes in transportation methods, if any, most heavily impacted population densities (and to what statistical degree) in Tennessee over those time periods between 1830 and 2010. Changes from transport by waterway, to transport by rail, and then to transport by road were considered. Each change indicates major investments in transport infrastructure and may reflect related aggregation of economic development over a sequence of decades. The following hypotheses were tested in this study:

- Alternative Hypothesis 1: There is a strong correlation between waterway transport network and population densities.
- Null Hypothesis 1: No or weak correlation between the waterway transport network and population densities.

- Alternative Hypothesis 2: With the introduction of the railway network, there is a strong correlation between railway network and population densities.
- Null Hypothesis 2: No or weak correlation between the railway network and population densities.

- Alternative Hypothesis 3: With the introduction of the historic road network, there is a strong correlation between the road network and population densities.
- Null Hypothesis 3: No or weak correlation between the historic road network and population densities.

- Alternative Hypothesis 4: With the introduction of the modern road network, there is a strong correlation between the modern road network and population densities.
- Null Hypothesis 4: No or weak correlation between the modern road network and population densities.

With the introduction of each of these four transport infrastructures, population totals are expected to increase in counties near transportation cores and areas of mass network interchange. It is anticipated that early in the 19th century, waterways controlled movement of commodity and migration to the integral towns and establishments at the time, particularly due to the signing of the Cherokee Treaty in 1791 which allowed free use of the Tennessee River in regards to trade and commerce (Cherokee Nation 2015). Moreover, in recent years there has been an outmigration from mountainous areas due to monetary imbalance (Mather 2004), so continued out-migration from these regions is expected in later decades. Mather (2004) notes that those who stay in these regions have higher poverty due to the scarcity of jobs, less education and/or skills. Due to scarcity of available data for the early time periods, income and education levels will not be explored in this study.

Finally, in the early 1900’s when mechanization began to become established, transportation moved from waterways to railroads and became the primary mode of commerce, opening a whole new world for an influx of migration highways (New Deal Network 2003). It is expected there will be continued growth in the future that correlates with the connectivity of these transportation networks, and that the outward population growth will also develop within a given distance from major railroads and roadways.

3.2 Data Sources

All population data sources were derived from information collected by the United States Census Bureau between 1830 and 2010, providing 180 years of demographic information (www.census.gov). Although the census categories are broad and roughly categorized, it is the only demographic numeration data source that consistently spans the entire study period, and is available in a readily accessible format.

Complications in the study overall included county boundary change over the study time span and census demographic enumerator practices. The latter required calculation of individual county populations in many instances from dozens of historic census sheets, many of these categories further broken down into ethnic, sex, and age classes as well. In all decades, values were tediously, manually input into Excel sheets and summed for grand county totals, using both pivot tables and filtering within the original Excel sheet based on county names. This study capitalizes on input datasets while recognizing important considerations and variables that are not included in this study.

3.2.1 Historic County Boundaries

Continuous county boundary changes by decade were taken into consideration, and since these features did not yet exist in a satisfactory digital form applicable to this study, all county boundary changes were manually digitized from georeferenced historical maps (Long 2000) to generate the historical dataset. Although several attempts have been made to generate these boundaries, the approach taken in this study was important for working with the historical boundaries raw data rather than using second and/or third party sources. Figure 12 highlights the processes taken to achieve these final boundaries.

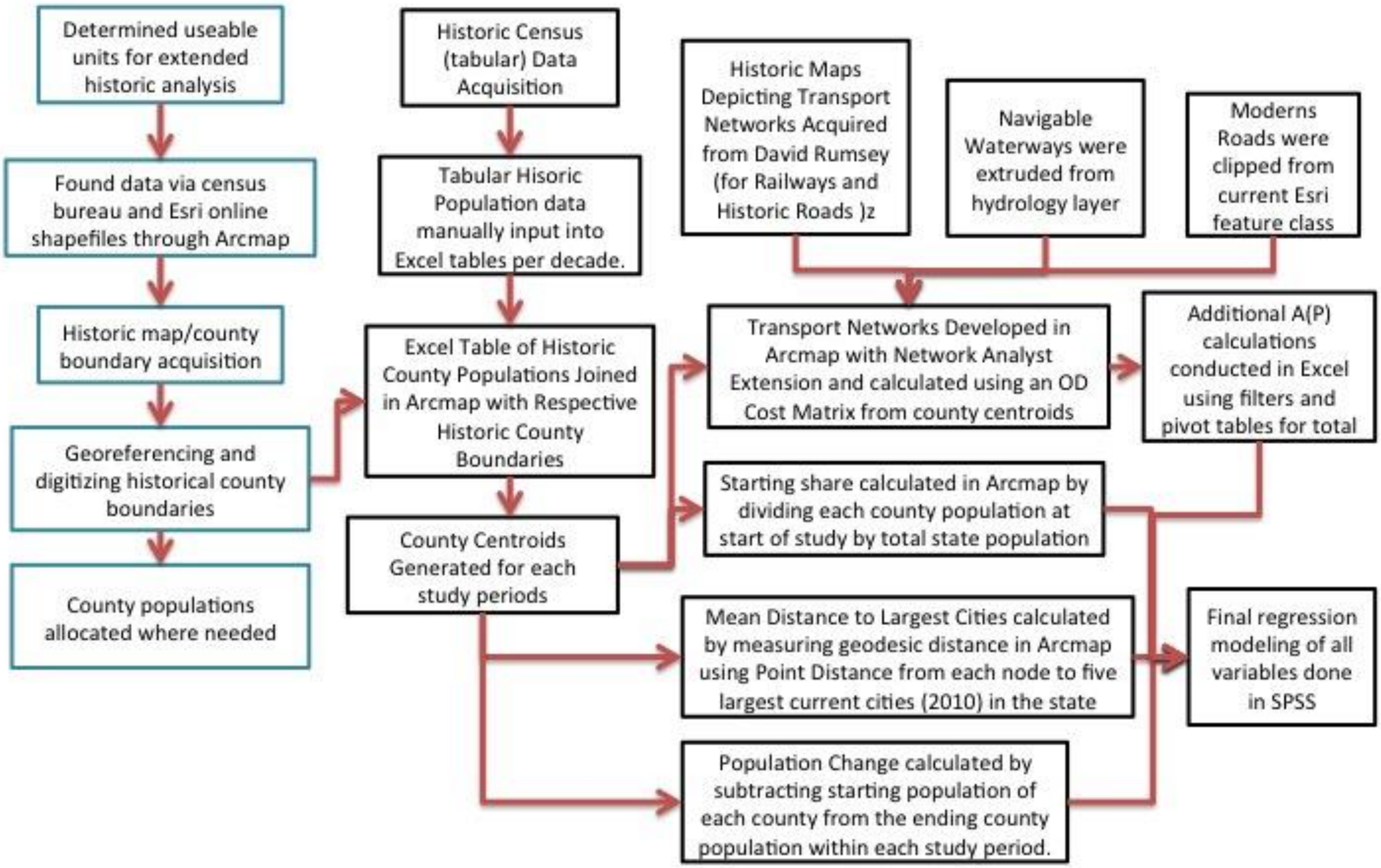

Figure 12. Diagram Highlighting Steps to Achieve Historic County Boundaries

3.2.2 Historic Population Data

The census data from 1830 through 1960 were neither available in spatial datasets nor as tabular data in Microsoft (MS) Excel spreadsheets. This data was available by decade from the Census Bureau as scanned census report volumes that had been collected and entered manually and varying greatly by enumerator practice within each decade. For example, the data from every decade is slightly different in regards to the number of categories labeled and documented. In this study, the census data were manually entered in decadal increments into MS Excel spreadsheet tables, one for each county available at the time of the data collection. The process resulted in sums of population by county for the state for each decade of the study. The population of each county was then calculated based on a total sum. In many instances this included the need to enter individual ethnic categories to achieve totals, which were then calculated and loaded as an attribute table into Esri ArcMap version 10.2. Each record in the table was joined to the corresponding unique polygon on the manually digitized historic county maps (Long 2000). Utilizing Esri ArcMap tools, the area (in square miles) was calculated for each country per decade. These steps are explicitly highlighted in Figure 13.

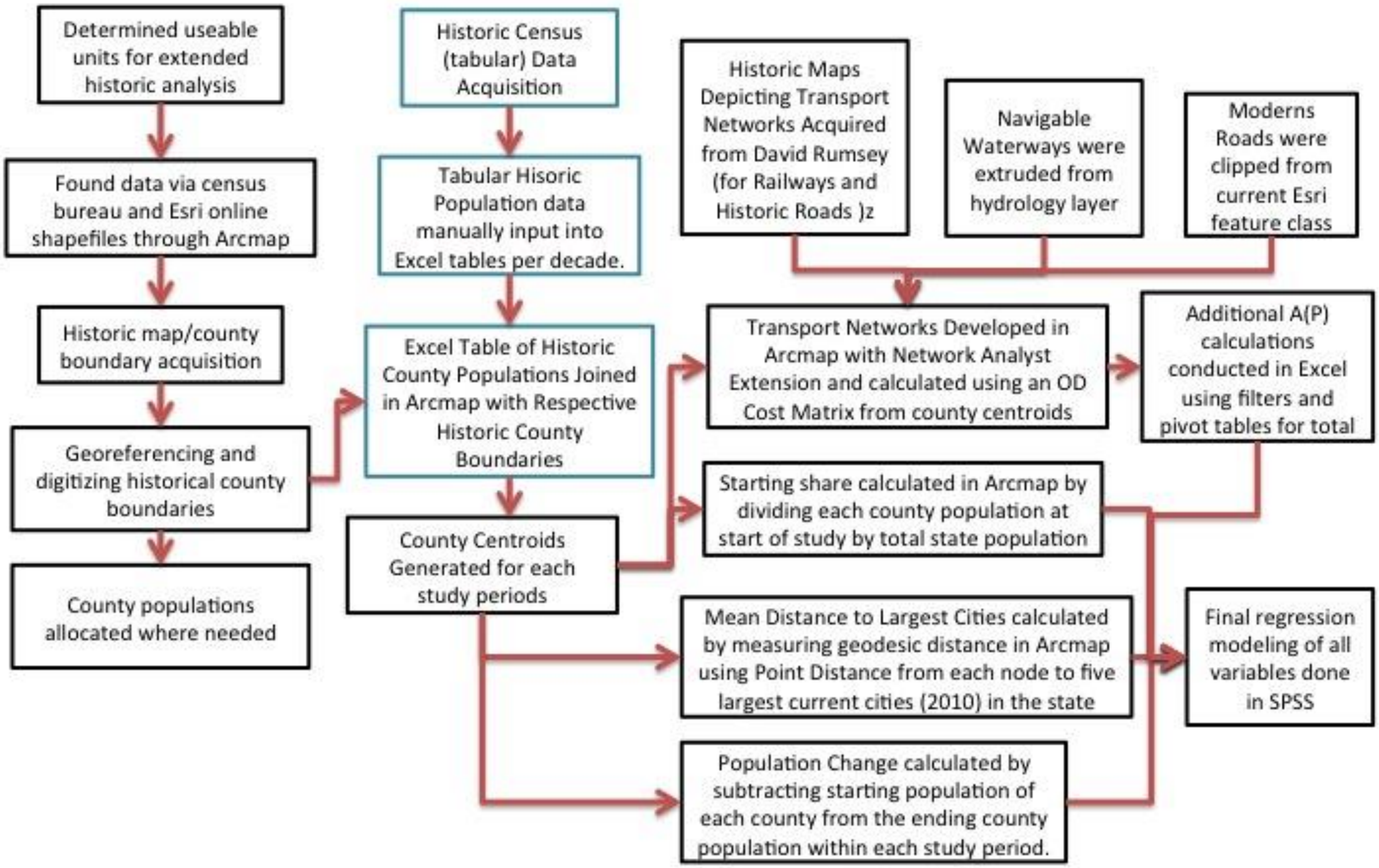

Figure 13. Diagram Highlighting Steps to Achieve Historic Population Totals Per Decade

3.2.3 Current Population Data

For the primary analysis of examining more current population change (i.e. 1970-2010), several specific spatial data extraction functions within the GIS were utilized. First, the outline for the study area (Tennessee) was extracted from a polygon feature class comprising the United States (Esri 2013). Secondly a polygon feature class that comprised all counties from the United States (National Historic Geographic Information System 2010) were extracted and imported into a separate polygon feature class. All MS Excel tables ranging from 1970 through 2010 were edited to only include Tennessee records, and broken down by county. Using GIS, these tables were then joined to the county feature classes by using an identical field within both tables (state name) in order to visualize the changes in population between 1970-2010.

3.3 Variable Definitions

The unit of analysis for the entire study of this thesis is the counties of Tennessee, as they existed in the final study period (Figure 14). In order to historically rectify these unit boundaries, the county boundary changes over time were considered throughout the entire study, but being primarily relevant within the first two study periods (Table 3).

Table 3. County Boundary Changes

Study Period	**Increase in Number of Counties**
Waterways (1830-1850)	64 -> 79
Railways (1860-1920)	84 -> 95

This process began with waterways and covered a thirty-year study period (1830-1850). The population for the 1830 dataset was distributed based on the 1850 county boundaries. For example, if a county lost 30 percent of its square mileage from 1830 into a newly generated county for 1850, 30 percent of the population was calculated and removed from that county for the 1830 population and given to the "eventual" county area. The overall allocated county populations are a rough estimate of the actual population distributions during these years.

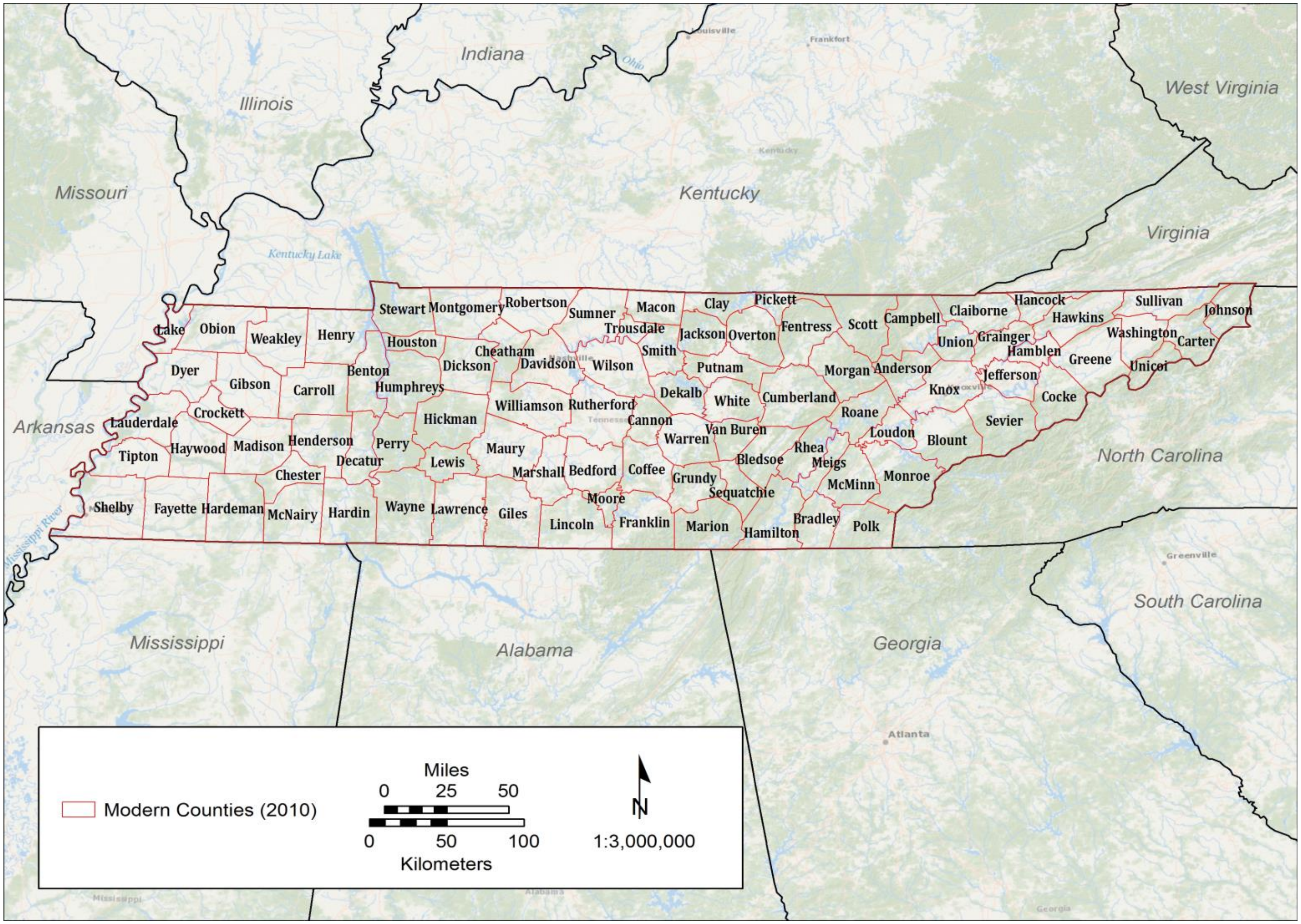

Figure 14. Modern County Boundaries in Tennessee

The final two study periods maintained the same number of counties, with very few counties experiencing boundary changes. The vector boundaries within the feature classes were redefined to reflect these changes throughout each applicable study period, although the waterways was the only study period that had less than the present-day number of counties included in the analysis (Table 4), which were the summation at the end of the study period (1850).

Table 4. Number of Counties (n) Per Study Period

Study Period	**Number of Counties in Analysis**
Waterways (1830-1850)	79
Railways (1860-1920)	95
Historic Roads (1930-1970)	95
Modern Roads (1980-2010)	95

3.4 Dependent and Independent Variables

Conducting statistical regression analysis in order to predict population changes is of long historical convention (Erickson 1974) and is used within this study as the dependent variable within this study, as in Kotavaara et al. (2011). Population change values were determined by subtracting the starting decade population total from the final decade population total per county within each of the four study periods. According to Schnell and Monmonier (1976), changes in population are often caused by one of two instances: an overall increase in reproduction or a change in net-migration, the movement of people across distances. The latter will be explored here in full as the second variable. Further elements to consider include reduced rates of death over time or longevity, especially with improved medical technologies in the more modern era.

Longevity is specifically believed to have biological significance, but is also affected by access to necessities and levels of hygiene (Pearl 1924).

Potential accessibility (A(P)) was used as the first independent variable and determined by first measuring distance from each node, the node being the county centroid, to all connecting centroid node locations within the network for each study period. The A(P) was calculated by dividing the population of each node by the resulting distance to each connecting node within the network, and then summing all of these values per county. This approach is similar to previous studies that defined transport network accessibility measurements (Kotavaara et al. 2011, Geurs and Ritsema van Eck 2001). The centroid, which was used throughout this study, is the geometric center of the polygon mass or gravity in GIS calculations. In GIS, a polygon is an area feature with boundaries at a defined scale, in this case being the county borders within Tennessee (Esri 2013). These county centroids were re-generated for each study period based on the ending county boundaries within the given study period. This allowed for total county use, specifically in earlier study periods where the starting decades had fewer county boundaries. If a centroid fell off of a transport network, a direct geodesic line was measured from the centroid to the closest vertex along the network and this distance was summed with the distance total when measuring the A(P) total of each county node. A node was declared connected to another node if it was along the same continuous transport network within the state. No connections outside of the state were considered within this study. Distinct transport networks (networks that did not connect to all admissible counties) only occurred during the waterways study period, although both study periods listed in Table 2 did have counties with A(P) values of 0 because they contained no network element (i.e. waterway or railway).

Historic roads and railways were manually digitized into a network database and distance fields were calculated using Esri's ArcMap Network Analyst extension, also used to create an

Origin Destination (OD) Cost Matrix for each of the four transport networks. The data was then exported to MS Excel spreadsheets to create conditional formatting and pivot tables for summation of each county's A(P) total. Potential accessibility (A(P)) was calculated by adapting the logarithm applied by Kotavaara et al. (2011), which provides the first independent variable. This variable is the measurement of both distance and population size for any given node within the transport network. This equation can be written as:

$$A(P) = \sum_{z=1}^{n} \frac{P_j}{d_{yz}}$$

In this equation, the A(P) assigns A as the area and P as the population, defining the potential accessibility base. Secondly, the d_{yz} provides the distance between the starting point and the destination, or connected nodes for every county within the study period. Further, P_j becomes the population of the ending destination point for each connected county along the network. Finally, the n is the total starting points, here depicting the county units within each study period. A higher A(P) total would signify a higher accessibility based on a higher population and likely higher connectivity to other other locations. Kotavaara et al. (2011) applied x as the variable for friction to account for the degree of urbanization. Their study applied x as an exponent on d_{yz} in Equation 1 shown above. This study does not apply a friction factor, so A(P) is simply a measure of both the distance and population size for any given node in the network. A higher number indicates greater accessibility based on network proximity to more populated areas (Tobler 1970).

As mentioned above, Kotavaara et al. (2011) further applied x as a variable for friction, noting that 1 is the understood value for the friction at an international-scale analyses. However due to lack of congestion and overall scale of this study, friction was omitted from this study. Friction in transport geography is a variable that can inhibit the effect distance has on growth and connectivity of places, leading to distance decay. It is primarily influenced by cost and time across space, which is often further defined by social status (Johnston 2009).

The second independent variable included in the final regression modeling was the share of the state population by county at the start of each study period. This was calculated by dividing each county population by the total sum of the state (i.e. all counties combined) for the starting decade within each study period. This variable was included for reasons of path dependancy, theorizing that counties with large populations would experience the greatest increase in population (Pavlínek 2003) and was calculated in the final Excel tables along with the A(P) total.

The third and final variable included in the model was mean distance to the top five largest modern-day cities in Tennessee based on theory of distance and effect (Tobler 1970). This measurement was calculated by running a Point Distance tool in ArcMap between the county centroids and the top cities, which in turn output a table containing the distance between every county and each of the five cities. This generated a table with distance measurements, five for each county. This table was exported to Excel and a pivot table was created that summed each county and the mean of all five distance measurements. The same mean measured variable was used within all study periods, which could have influenced some of the output. With this noted, four of the five cities were consistent in maintaining largest populations throughout all man-made study periods (Periods B-C), with the exception of Clarkesville which only saw a high increase in population during World War II with the construction of Fort Campbell just across the border in Kentucky (Muir 2009). All variable calculation methods are depicted in the blue highlighted boxes within the flowchart in Figure 15.

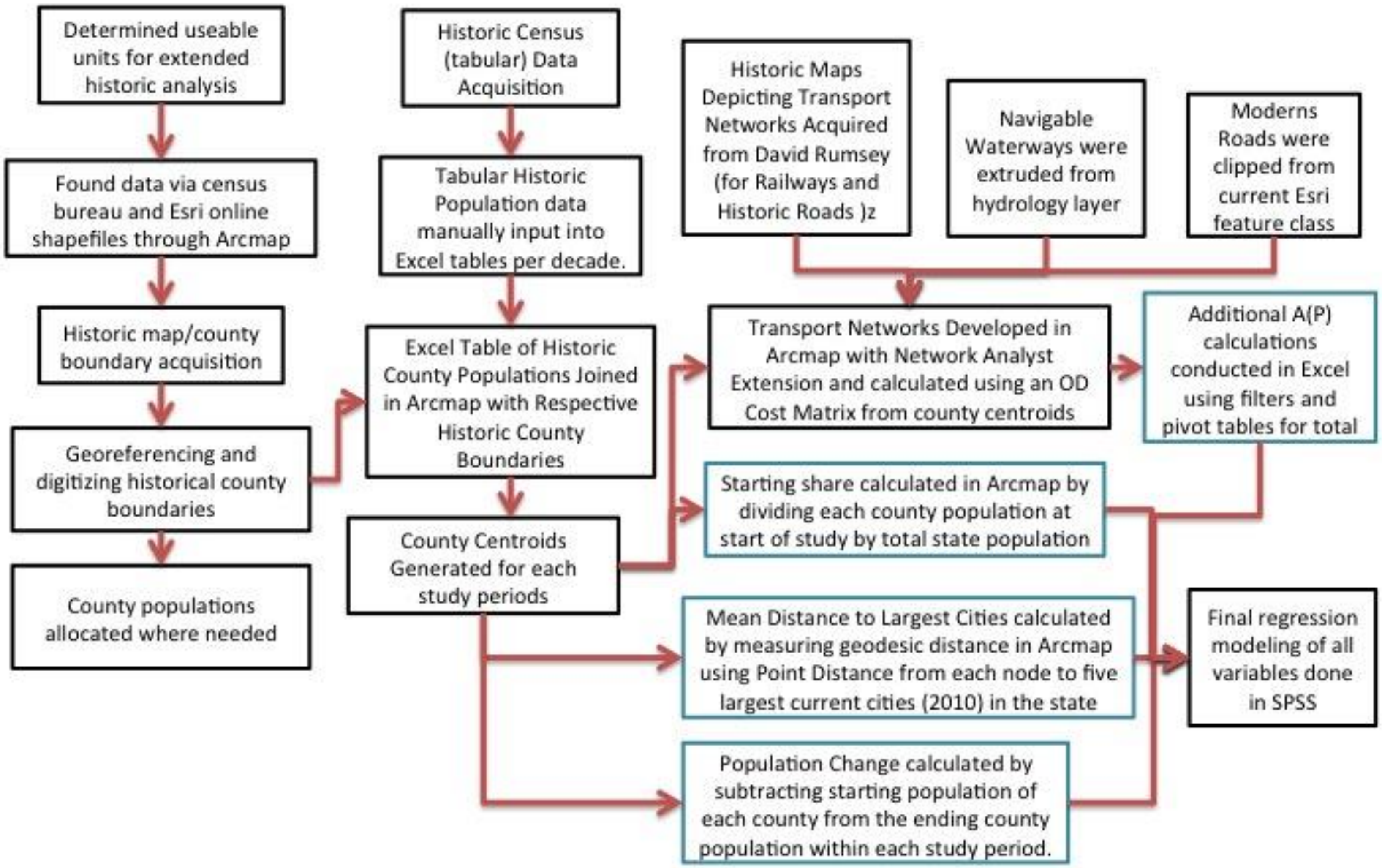

Figure 15. Diagram Highlighting Steps to Achieve Study Variables

3.5 Conclusion

To summarize, population change was used as the dependent variable and A(P) as the primary independent variable with which the hypotheses are framed. Relative share of population per county in the state of Tennessee at any given point in time was calculated and included as the secondary variable within the regression model. Finally, the mean distance to city was applied as the third independent variable per study period using the five largest cities in the study to determine if this may have contributed to the population change as well. The final regression modeling was done using SPSS to process the data, in order to find the most appropriate linear regression models that include all necessary variables (International Business Machines 2015).

The A(P) results from each study period derived from the individual transportation networks measured in ArcMap were first input into SPSS separately, with their respective variables (both independent and dependent). Linear regression was used to statistically model the different study periods and individual outcomes, applying a value of $p \leq 0.05$ as the significance threshold. R-squared is used in regression modeling to determine coefficient variation for the variable included in the model (Dancer and Tremayne 2006). The adjusted R-squared of the model outcome was applied to compare model predictability in order to assess how well the modeled variables support the tested hypotheses (Ohtani 1994). The following chapter evaluates the results from the analyses run over the entire study period chronologically, beginning with navigable waterways and ending with modern roads in Tennessee. Linear regression model outputs are discussed for each of the four study periods listed in Table 1, using the three independent variables: A(P) total, population share and average city distance, and the dependent variable, population change, within each period.

CHAPTER 4: RESULTS

This chapter evaluates the results from the analyses run over the entire study period chronologically, beginning with navigable waterways and ending with modern roads in Tennessee. Linear regression model outputs are discussed for each of the four study periods listed in Table 1, using the three independent variables: A(P) total, population share and average city distance, and the dependent variable, population change, within each period. Next, this chapter comparatively examines all of these variables within the entire study area using a bivariate correlation matrix in order to assess the relationship between all variables within the entire study (Puth 2014).

4.1 Analyses of Study Periods

Initially, the histogram was not normally distributed in modeling the variables, in additional to several other model variables throughout the study (Table 5); therefore a lognormal was run on the independent variables prior to generating the linear regression in these instances.

Table 5. Variable Distribution Throughout Study

Variable	Distribution	Natural Log Transformation Used in Models
Population Change Period A	Normal	No
A(P) Total 1830	Skewed	Yes
Starting Population Share (1830)	Skewed	Yes
Population Change Period B	Normal	No
A(P) Total 1850	Normal	No
Starting Population Share (1850)	Normal	No
Population Change Period C	Normal	No
A(P) Total 1930	Skewed	Yes

Variable	Distribution	Natural Log Transformation Used in Models
Starting Population Share (1930)	Normal	No
Population Change Period D	Normal	No
A(P) Total 1980	Normal	No
Starting Population Share (1980)	Normal	No
Mean Geographic Distance to Largest Cities	Normal	No

4.1.1 Waterway Results

For the waterways analysis, population change between 1860 and 1920 was used as the dependent variable, and accessibility potential (A(P)), starting population share, and mean distance to largest cities as the independent variables. The result following the lognormal run then showed an adjusted R-squared was 0.240, meaning the linear regression explains 24% of the variance in population growth over the study period (Table 6).

Table 6: Linear Regression Model for the Waterways Study Period (1830-1850)

Waterway Study Period		
Adjusted R-Squared	**0.240**	
Model	**Intercept**	**Significance**
Constant	-1420.061	0.955
Accessibility Potential (1830)	2691.979	0.000
Population Share (1830)	-1449.700	0.088
Mean Geographic Distance to Largest Cities	-78.188	0.970
N=79		

The null hypothesis is rejected here due to the resulting correlating relationship in the model between the dependent and independent variable of Accessibility Potential based on a $p \leq 0.05$ level of significance. The additional variable, the population share, also proves to be significant but shows an inverse coefficient relationship, meaning that those counties with a larger population at the start of the study period grew slower than those counties with a smaller starting population. In other words, contrary to the theory of path dependence, counties that started out small and had good access to the navigable waterway network grew faster than counties that started out large with access to waterways. Furthermore, the average geographic distance that turned out to be connected to the largest cities later on is not significant.

Overall, the results for this model point to the importance of the waterway network in the growth in Tennessee counties in the early years of the state, but also indicate that there are other variables that may further explain the change in population that have not yet been uncovered. More specifically, by visually comparing the population of 1830 counties at the start of the study period with the populations of 1840 and 1850 counties (Figure 16) at the end of study period, it is clear that the counties in direct contact with the strongest part of the network have stronger population growth over time than those disconnected or in weaker regions of the network. Specifically, west Tennessee and the central region experienced the highest volume of growth during this time frame, with Shelby and Fayette Counties experiencing the highest rate of change at this time (Figure 17). These findings also correlate with the early 19th century expansion into Western and Central Tennessee along the Mississippi and Cumberland Rivers. Specifically, there is significant growth that is taking place in Memphis in Shelby County during this time period (Rothrock and Smith 1973).

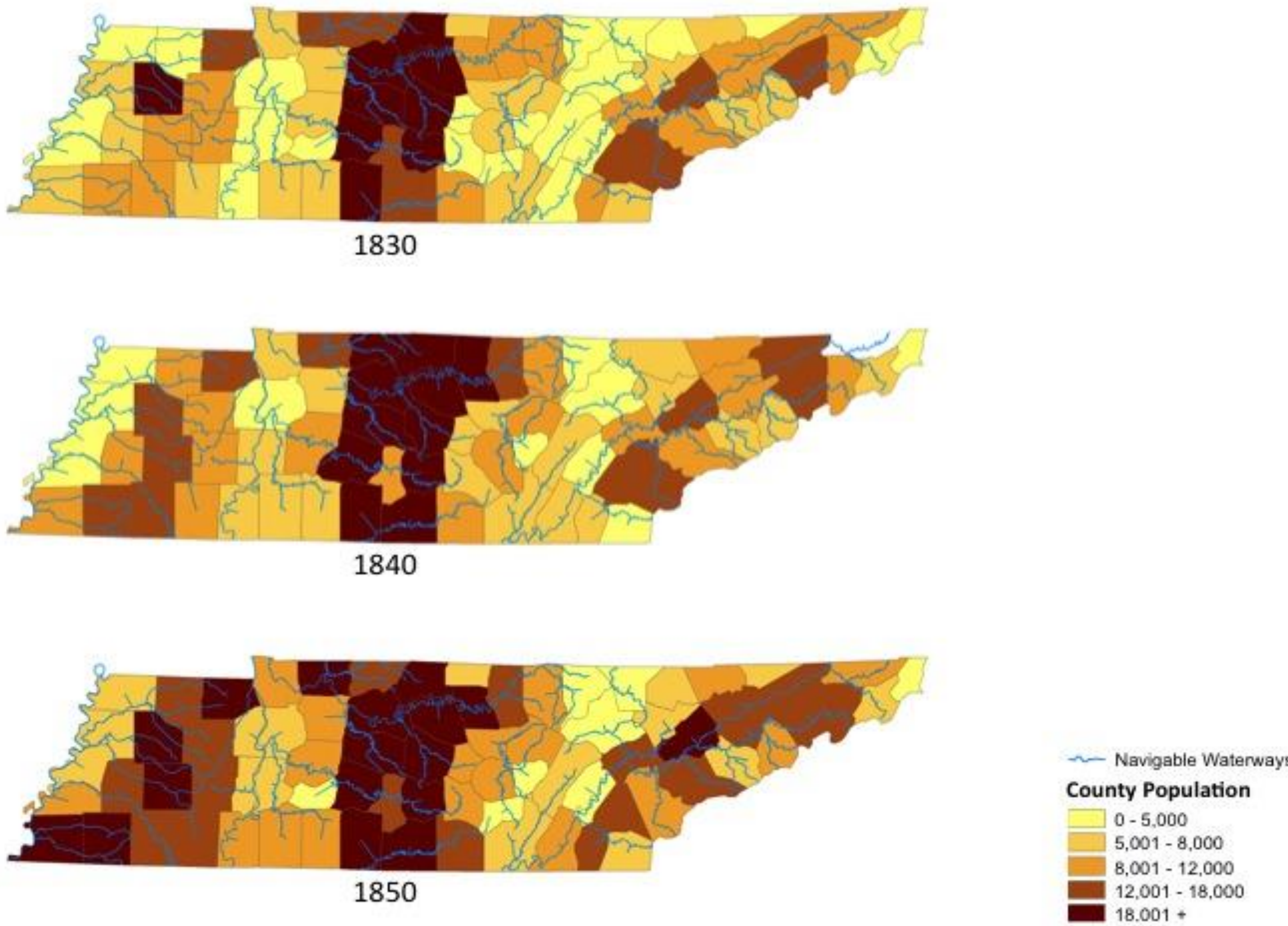

Figure 16. Choropleth Maps of Population Change Maps During the Waterways Study Period

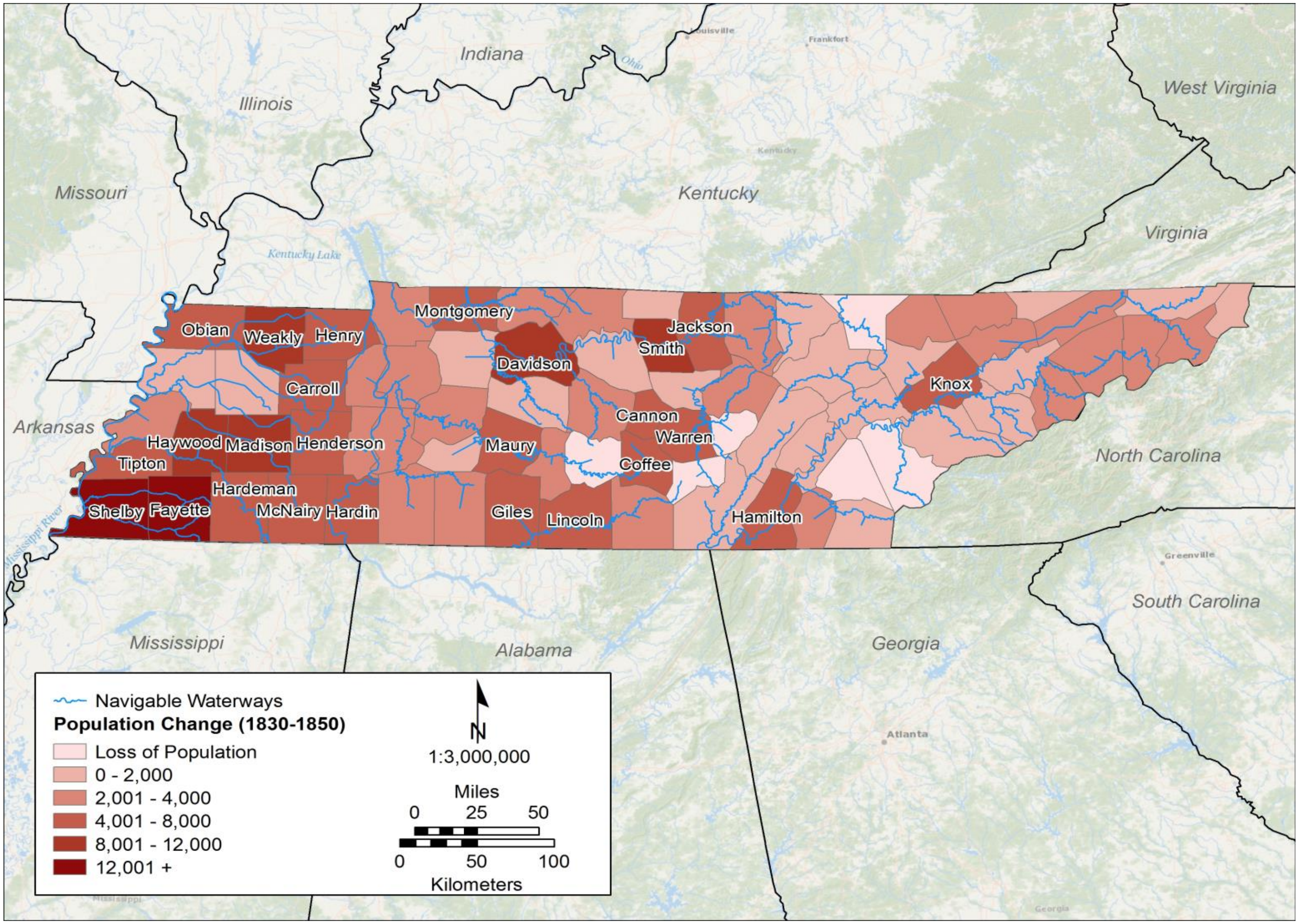

Figure 17. Map of Population Change for Waterways Study (1830-1850)

4.1.2 Railway Results

The railways analysis used population change between 1860 and 1920 as the dependent variable and accessibility potential (A(P)), starting population share, and mean distance to largest cities as independent variables. The adjusted R-squared was 0.301, explaining 30% of the variance in the population growth over the study period (Table 7).

Table 7: Linear Regression Model for the Railway Study Period (1860-1920)

Railway Study Period		
Adjusted R-Squared	**0.301**	
Model	**Intercept**	**Significance**
Constant	-33386.299	.019
Accessibility Potential (1860)	-5160.242	.526
Population Share (1860)	3287504.552	.005
Mean Geographic Distance to Largest Cities	.084	.122
N=95		

The results signify that the null hypothesis is confirmed for railways because the A(P) is not significant. In the railway period, the starting share of population really drives the results as the intercept is, as expected, positive, and the variable is significant at the 0.05 level. The mean distance to largest populated cities proves again to be insignificant with a value of 0.122. Unlike the waterways, the railways followed the theory of path dependence, where those counties that started out large and had limited access to the railway network grew faster than counties that started out small with access to railways. Once again, the average geographic distance proves to be insignificant.

Overall, the results for this model signify a lack of importance of the railway network in the growth in Tennessee counties over this time period in the state, but also suggests additional variables that may further explain the change in population that have not yet been discerned.

Moreover, the importance of path dependency in this study period, which unlike the null outcome in the waterways study, is likely due to an increase in the county populations in this later time period which remained relatively small during the first network analysis. By the time the railway network was developed, counties were already larger in population, resulting in a higher standard deviation than observed related to waterways (Figure 11). Additionally, due to the railway obstruction during the civil war era (Johnson 2010) mentioned previously, the network likely failed to become a strong factor behind population growth overall, supporting the outcome of the null hypothesis in this study period.

Finally, the population changes taking place over this study period (Figure 18) illustrate a common increase in counties sharing borders with counties that were already large at the beginning of the study period, indicating that proximity could have been important during this time period. Specifically, Eastern Tennessee saw a dramatic rise in both Knox and Hamilton counties and a majority of their surrounding counties (Figure 19).

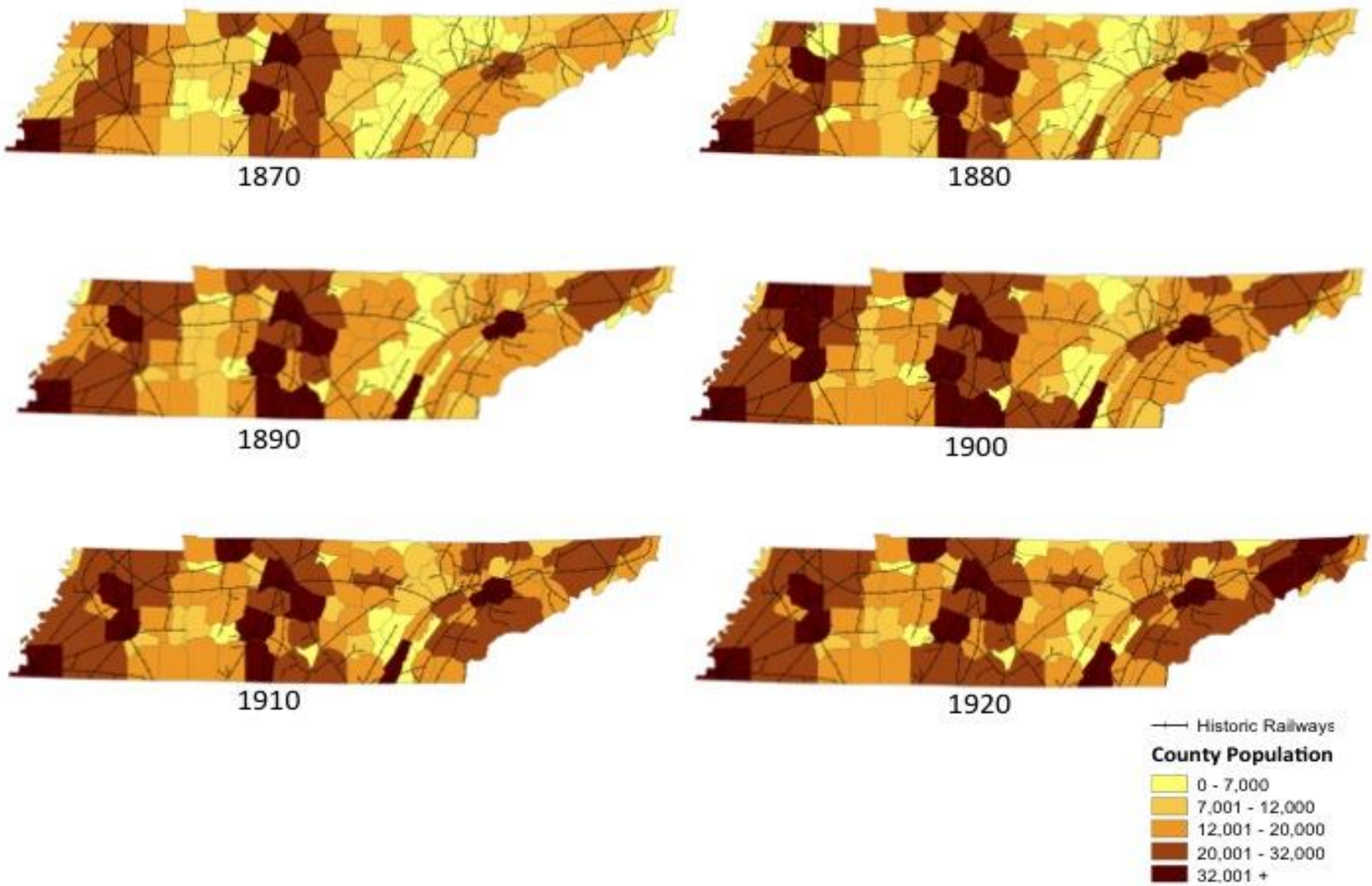

Figure 18. Choropleth Maps of Population Change Maps During the Railways Study Period

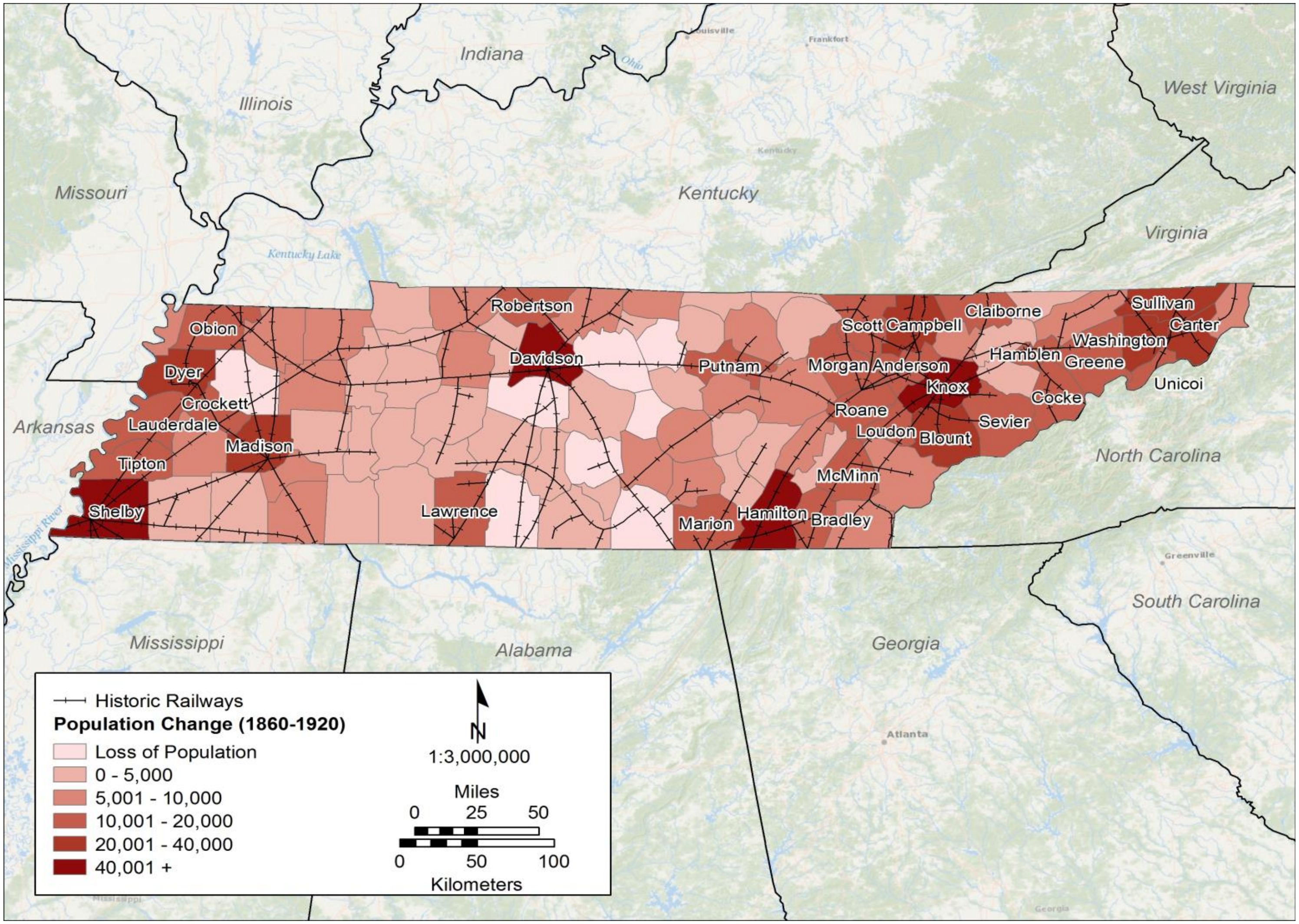

Figure 19. Map of Population Change for Railways Study (1860-1920)

4.1.3 Historic Road Results

For the historic roads analysis, population change between 1930 and 1970 was used as the dependent variable and accessibility potential (A(P)), starting population share, and mean distance to largest cities were the independent variables. The adjusted R-squared was 0.894, providing an almost 90% predictability in population change during this study period (Table 8).

Table 8: Linear Regression Model for the Historic Road Study Period (1930-1970)

Historic Road Study Period		
Adjusted R-Squared	**0.894**	
Model	**Intercept**	**Significance**
Constant	-34244.040	0.000
Accessibility Potential (1930)	-1863.025	0.002
Population Share (1930)	5922295.204	0.000
Mean Geographic Distance to Largest Cities	0.14	0.697
N=95		

In this study period, the null hypothesis is rejected for the A(P) total, with a significance of 0.002, but the relationship is negative, contrary to what had been theorized in this study. In other words, the counties that had less accessibility to transport networks actually grew more quickly than those with higher accessibility. On a bivariate basis, the potential accessibility has a high significance and positive relationship with population, although when including all variables, the sign flips. Based on theory, starting share belongs in the model although this may indicate an instance of multicollinearity. The population share is highly significant, but the mean distance to the largest cities again remains insignificant.

The model results indicate that road network inaccessibility, as well as starting share of the population, were underlying causes behind population changes between 1930 and 1970, which further supports both the theory of path dependency and the importance of the road network. The results of this study period analysis demonstrate the importance of the early road network for Tennessee counties and population growth. It also shows to be more important than proximity of largest cities. Although there may be additional variables contributing to population change, the outcome for this model is statistically significant and leaves relatively little of the change in the dependent variable unexplained.

A visual assessment of these county populations from the start through the end of this study period (Figure 20) for historic roads further supports this outcome. Growth related to the historic roads transport network is very apparent, particularly in central and eastern Tennessee where road development becomes more concentrated. Secondly, the largest counties become significantly more populated as is expected with the model supporting path dependency, where the population has increased the most in counties containing largest cities, specifically with Shelby, Davidson, Hamilton, and Knox counties seeing the highest rate increase (containing the cities of Memphis, Nashville, Chattanooga and Knox, respectively) (Figure 21).

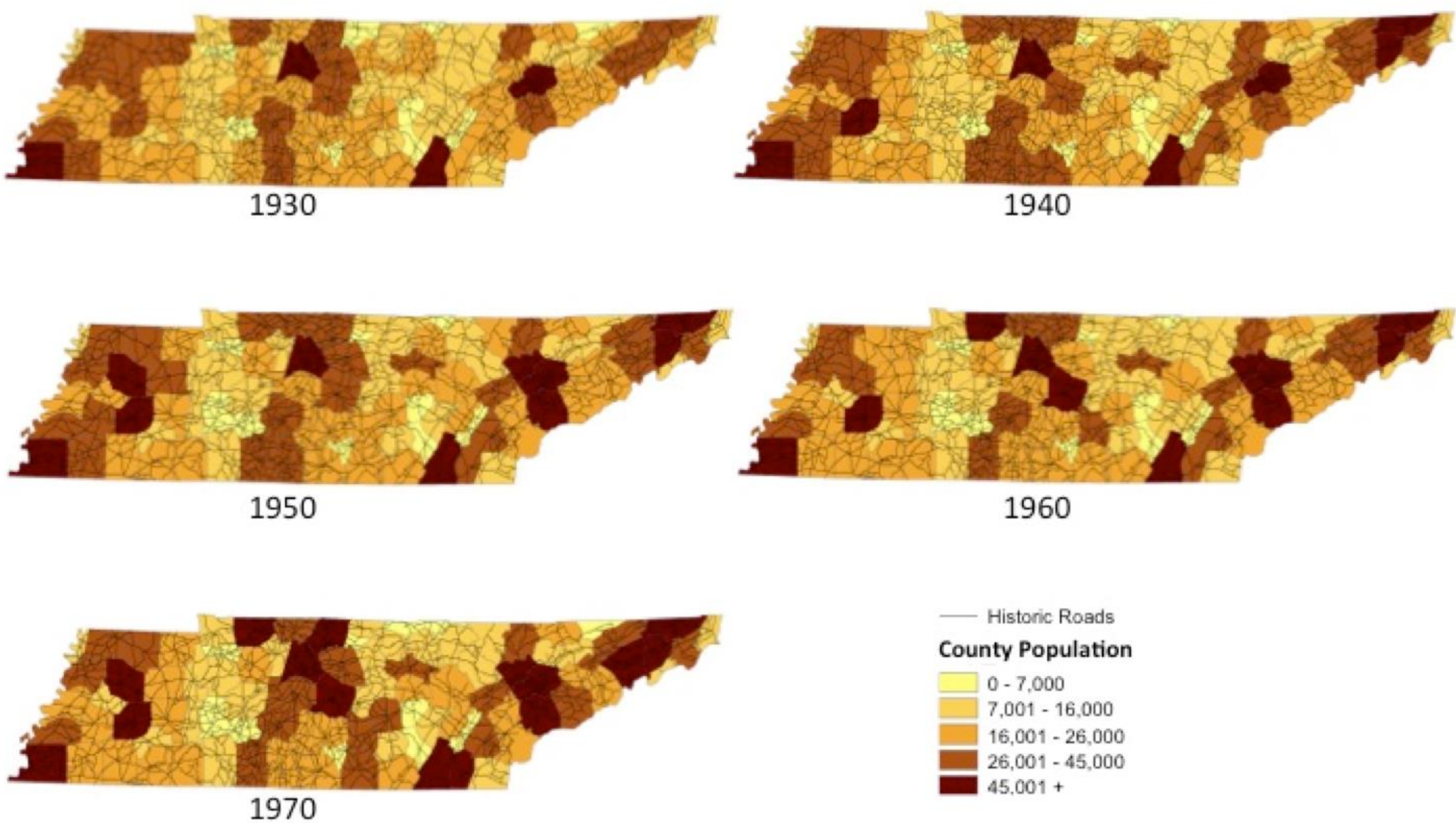

Figure 20. Choropleth Maps of Population Change Maps During the Historic Roads Study Period

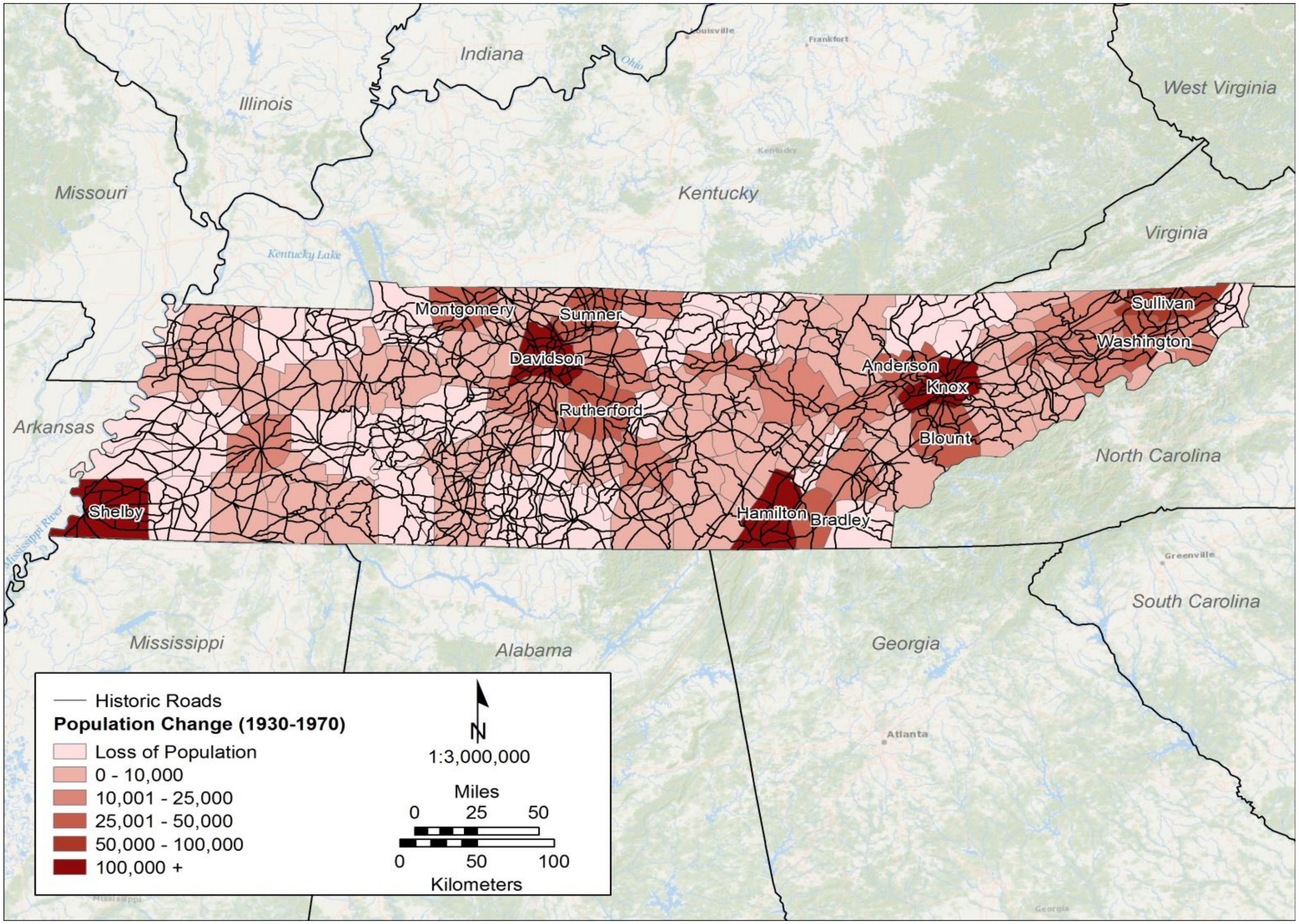

Figure 21. Map of Population Change for Historic Roads Study (1930-1970)

4.1.4 Modern Roads Results

To analyze modern roads, the population change between 1980 and 2010 was used as the dependent variable and accessibility potential (A(P)), starting population share, and mean distance to largest cities as the independent variables. The adjusted R-square was 0.577, accounting for a nearly 58% of the variance for population growth over this time period (Table 9).

Table 9: Linear Regression Model for the Modern Road Study Period (1980-2010)

Modern Road Study Period		
Adjusted R-Squared	**0.577**	
Model	**Intercept**	**Significance**
Constant	20057.547	0.045
Accessibility Potential (1980)	482.964	0.005
Population Share (1980)	165257.969	0.640
Mean Geographic Distance to Largest Cities	-0.063	0.116
N=95		

The null hypothesis is again rejected for the final study period based on an A(P) significance of 0.005. Although the starting share of population is not significant in the overall regression model, it is worth noting that it is significant on a bivariate basis when modeled separately. This signifies that the potential accessibility total is such a strong predictor of population change that it overwhelms the starting share of population within this particular model. Lastly, the mean distance to largest cities remains insignificant in this final study period, indicating that network distance has greater importance in the model than geodesic distance. This could be due to a decrease in the influx of rural migrants, or a possible increase in out-migration from large cities to more rural peripheries or the movement of businesses and infrastructure to the outer urban areas, a concept termed “edge city”, which is common in the United States by the closing decades of this study period (Garreau 1991). The road network also becomes so extensive that major cities become less important due to high levels of connectivity throughout the state at this time. Overall, the results support the theory of potential accessibility in this model but leave room to suggest that additional variables may be the key to further understanding the most recent population trends.

Upon doing a visual analysis of populations per decade over the study period (Figure 22), it appears that the network is contributing to population growth particularly in counties of mass highway interchange (i.e. Davidson, Shelby, Hamilton, Knox) and the growth of the counties surrounding these major-city-containing counties that are likely directly correlated with the expansion of the roads networks (Figure 23).

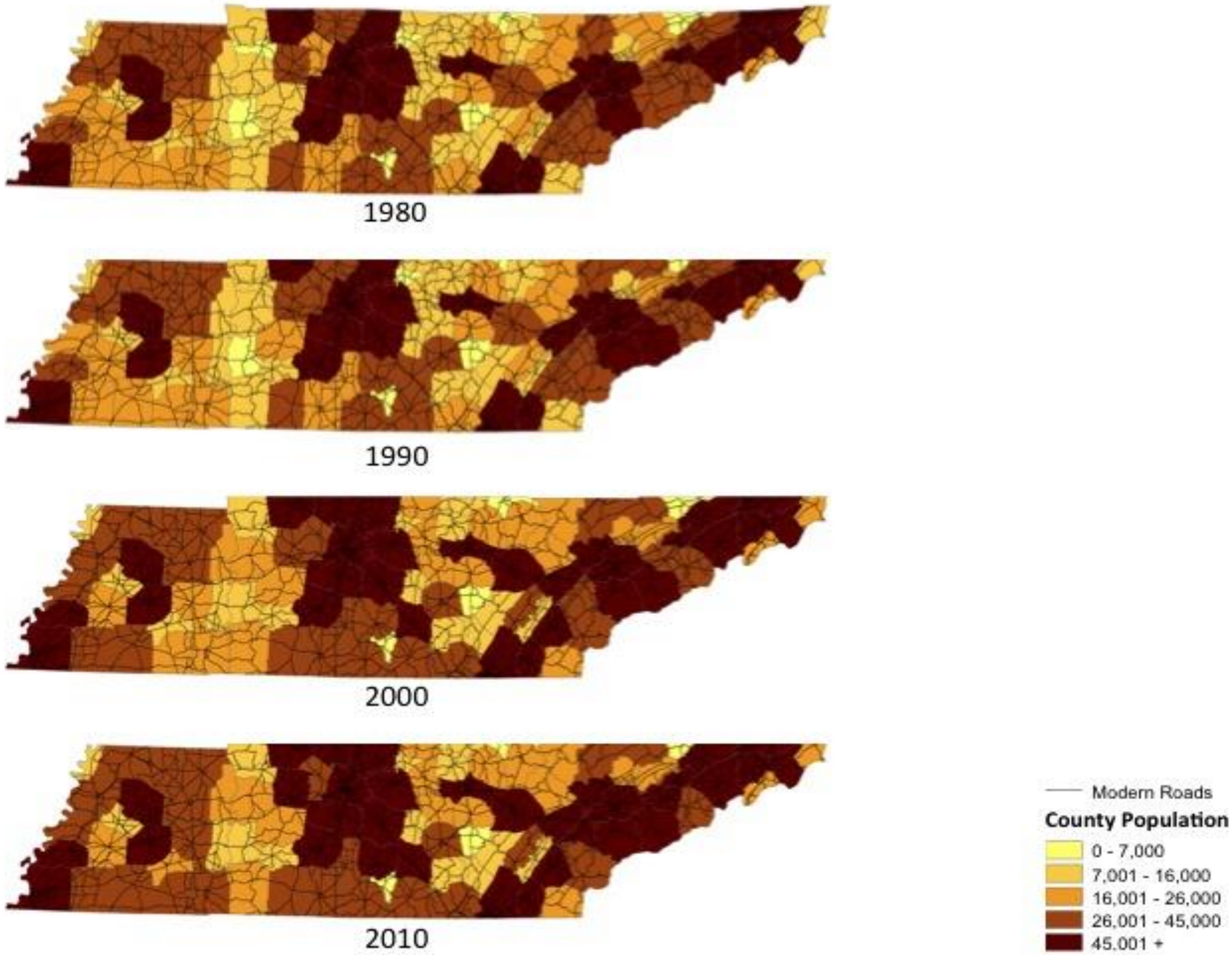

Figure 22. Choropleth Maps of Population Change Maps During the Modern Roads Study Period

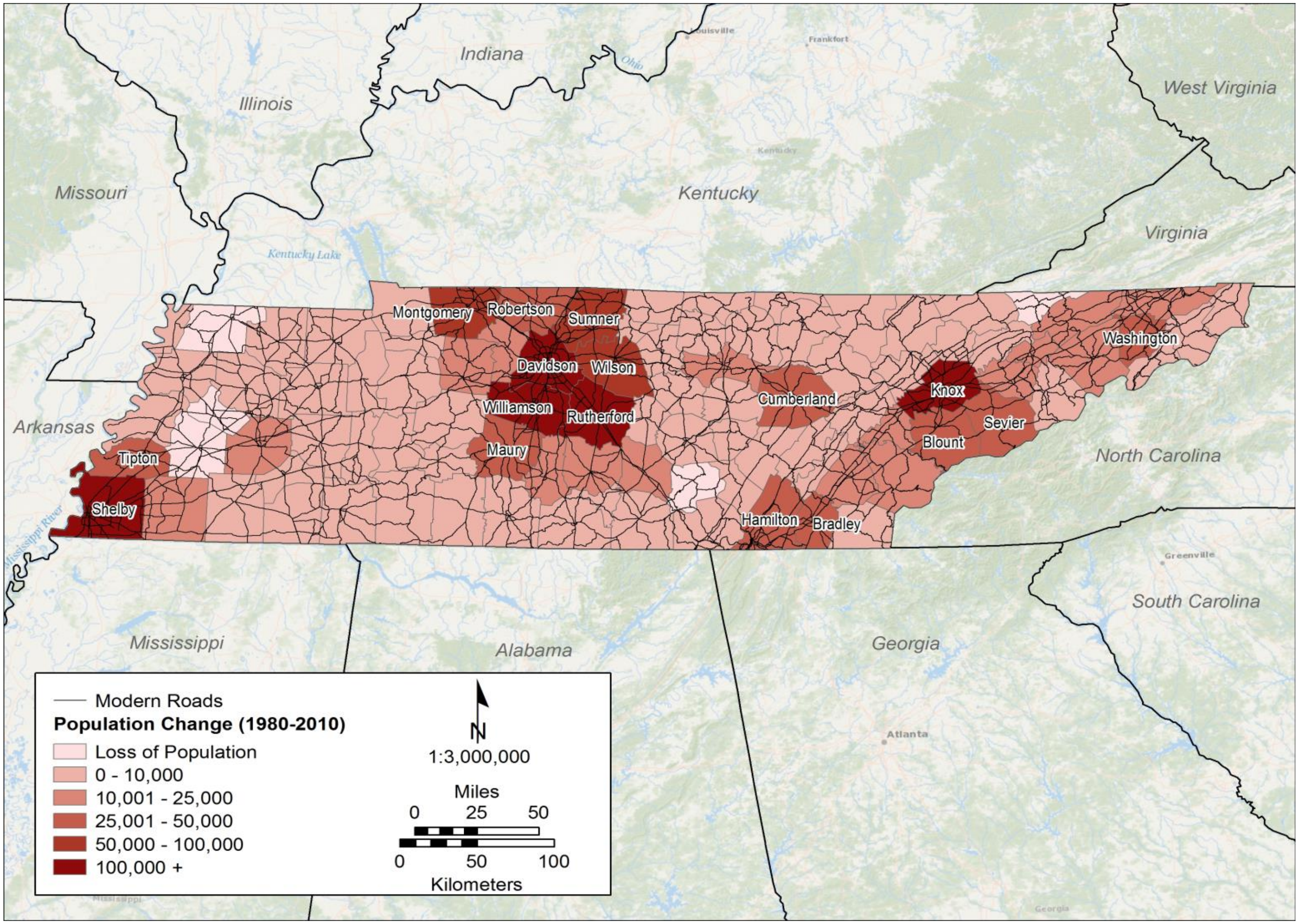

Figure 23. Map of Population Change for Modern Roads Study (1980-2010)

4.2 Bivariate Correlations

An additional step was taken to examine all of the variables within the entire study area using a bivariate correlation matrix in order to assess whether multicollinearity existed between any of the variables throughout the study. Specifically, a bivariate Pearson correlation was performed on all variables throughout all four study periods (Puth 2014). The following table (Table 10) reports on the variables that had significant correlations between them, and disregarded all others for comparison purposes. A full bivariate matrix was generated in SPSS including each variable in the entire study. The matrix was scrutinized for correlations pertaining to the dependent variable, but also included all independent variables to assess correlations within each unique study period. In regression modeling, particularly in considering the overall results, a high correlation between multiple variables is critical to consider because it becomes strenuous to compare the explanatory power of these variables (Allison 1999). The bivariate Pearson correlation specifically depicts associations between variables, but does not impart speculation about causation.

Table 10: Table of Significant Pearson Correlations

Bivariate Variable	**Significant Correlations**	**Significance (2-tailed)**	**R^2**
A(P) Total 1830 (I)	A(P) Total 1860 (I)	0.000*	0.683
	A(P) Total 1930 (I)	0.000*	0.426
	A(P) Total 1980 (I)	0.001*	0.323
	Population Share 1830 (I)	0.000*	0.698
	Population Share 1860 (I)	0.000*	0.638
	Population Share 1930 (I)	0.001*	0.341
	Population Share 1980 (I)	0.012*	0.256
	Population Change Period A (D)	0.025*	0.229
	Population Change Period D (D)	0.001*	0.335
A(P) Total 1860 (I)	A(P) Total 1830 (I)	0.000	0.683
	A(P) Total 1930 (I)	0.000	0.743
	A(P) Total 1980 (I)	0.000*	0.637
	Population Share 1830 (I)	0.000*	0.797
	Population Share 1830 (I)	0.000*	0.924
	Population Share 1930 (I)	0.000*	0.646
	Population Share 1980 (I)	0.000*	0.561
	Population Change Period A (D)	0.000*	0.527
	Population Change Period B (D)	0.000*	0.448
	Population Change Period C (D)	0.000*	0.489
	Population Change Period D (D)	0.000*	0.692
A(P) Total 1930 (I)	A(P) Total 1830 (I)	0.000*	.426
	A(P) Total 1860 (I)	0.000*	.743
	A(P) Total 1980 (I)	0.000*	.965
	Population Share 1830 (I)	0.000*	.422
	Population Share 1860 (I)	0.000*	.731
	Population Share 1930 (I)	0.000*	.944
	Population Share 1980 (I)	0.000*	.896
	Population Change Period A (D)	0.000*	.545
	Population Change Period B (D)	0.000*	.887
	Population Change Period C (D)	0.000*	.850
	Population Change Period D (D)	0.000*	.716

Bivariate Variable	Significant Correlations	Significance (2-tailed)	R^2
A(P) Total 1980 (I)	A(P) Total 1830 (I)	0.001*	.323
	A(P) Total 1860 (I)	0.000*	.637
	A(P) Total 1930 (I)	0.000*	.965
	Population Share 1830 (I)	0.001*	.343
	Population Share 1860 (I)	0.000*	.652
	Population Share 1930 (I)	0.000*	.949
	Population Share 1980 (I)	0.000*	.947
	Population Change Period A (D)	0.000*	.513
	Population Change Period B (D)	0.000*	.936
	Population Change Period C (D)	0.000*	.919
	Population Change Period D (D)	0.000*	.752
Population Share 1830 (I)	A(P) Total 1830 (I)	0.000*	.698
	A(P) Total 1860 (I)	0.000*	.797
	A(P) Total 1930 (I)	0.000*	.422
	A(P) Total 1980 (I)	0.001*	.343
	Population Share 1860 (I)	0.000*	.773
	Population Share 1930 (I)	0.002*	.320
	Population Share 1980 (I)	0.015	.248
	Mean Distance to Cities (I)	0.049	-0.203
	Population Change Period D (D)	0.000*	0.563
Population Share 1860 (I)	A(P) Total 1830 (I)	0.000*	.638
	A(P) Total 1860 (I)	0.000*	.924
	A(P) Total 1930 (I)	0.000*	.731
	A(P) Total 1980 (I)	0.000*	.652
	Population Share 1830 (I)	0.000*	.773
	Population Share 1930 (I)	0.000*	.727
	Population Share 1980 (I)	0.000*	.652
	Population Change Period A (D)	0.000*	.691
	Population Change Period B (D)	0.000*	.523
	Population Change Period C (D)	0.000*	.591
	Population Change Period D (D)	0.000*	.668

Bivariate Variable	Significant Correlations	Significance (2-tailed)	R^2
Population Share 1930 (I)	A(P) Total 1830 (I)	0.001*	0.341
	A(P) Total 1860 (I)	0.000*	0.646
	A(P) Total 1930 (I)	0.000*	0.944
	A(P) Total 1980 (I)	0.000*	0.949
	Population Share 1830 (I)	0.002*	0.320
	Population Share 1860 (I)	0.000*	0.727
	Population Share 1980 (I)	0.000*	0.972
	Population Change Period A (D)	0.000*	0.658
	Population Change Period B (D)	0.000*	0.958
	Population Change Period C (D)	0.000*	0.939
	Population Change Period D (D)	0.000*	0.685
Population Share 1980 (I)	A(P) Total 1830 (I)	0.012*	256
	A(P) Total 1860 (I)	0.000*	561
	A(P) Total 1930 (I)	0.000*	896
	A(P) Total 1980 (I)	0.000*	947
	Population Share 1830 (I)	0.015*	248
	Population Share 1860 (I)	0.000*	652
	Population Share 1930 (I)	0.000*	972
	Population Change Period A (D)	0.000*	602
	Population Change Period B (D)	0.000*	968
	Population Change Period C (D)	0.000*	991
	Population Change Period D (D)	0.000*	709
Mean Distance to Cities (I)	Population Share 1830 (I)	0.049*	-0.203

*Correlation is significant at the 0.05 level (2-tailed)
D = Dependent Variable
I = Independent Variable

To begin with, the waterways show a correlation between population share with both A(P) and average distance to largest cities, indicating that both network potential and distance are already associated with population share at the start of the study period, although a higher significance is shown between accessibility and starting share. Measuring geodesic distance, the top largest cities in Tennessee that were used in this study are located along the navigable waterways, which may help explain some of the connection between average distance and potential accessibility, especially since this is the only instance where mean distance shows any significance in the entire study. Since population share and A(P) are both significant and they are related to one another, there is multicollinearity in this case which creates uncertainty as to which variable really explains the results.

Railways also show a strong correlation between A(P) total and population share. Surprisingly, the correlation between the dependent variable, population change, and A(P) now shows a very high significance when modeled on a bivariate basis. This is the second instance where one independent variable simply overpowers the other possible explanatory variables within the model. Looking at the resulting figures for population changes for Period B (Figures 18 and 19), it could easily be suggested that the pattern of population growth is correlated with the path of the railways as well as with the obvious growth in starting larger counties.

Historic Roads exhibit a strong correlation between county share and A(P) total, suggesting that the starting population most likely influenced accessibility, which makes sense since more populated areas tend to have better transport access based on the theory applied Krugman 1991). Nevertheless there may be additional variables not included in this study that could further explain population change or the inverse relationship resulting from accessibility potential. More specifically, both variables are significant in the regression model and are both correlated so it is not possible to determine certain causality in this case. It is also is possible that populations are in fact driving the development of the transport network contrary to the opposite hypothesis.

As previously discussed, starting population share within Period D, modern roads, shows high correlation with population change on a bivariate level. Also, as seen in all other study periods in the bivariate matrix, county share and A(P) total had a high correlation. With the level of growth occurring in this final study period, and the density of the transport networks, this is not surprising. Overall, study periods where more than one independent variable became significant with population change on a bivariate basis could indicate that it may be either variable explaining the change, or a combination of the significant variables for reasons mentioned above.

As a final measure, the potential accessibility for Period C (historic roads) and Period D (modern roads) based on the corresponding relationship change between these two study periods and similar methods of transportation were checked for differences by running a correlation on both A(P) totals in SPSS. Based on the multivariable results, there was a possibility that the networks were constructed to serve varying purposes (accessibility to cities versus simply connecting counties). This was shown not to be the case, as the results showed a high correlation between the two with a significance of 0.000 and an R-squared of 0.965, predicting over 96 percent of the variance in the model. Figure 24 and Figure 25 compares the potential accessibility between 1930 and 1980, displaying higher rates of A(P) in counties with respective increases in road networks, particularly in areas of mass highway interchange.

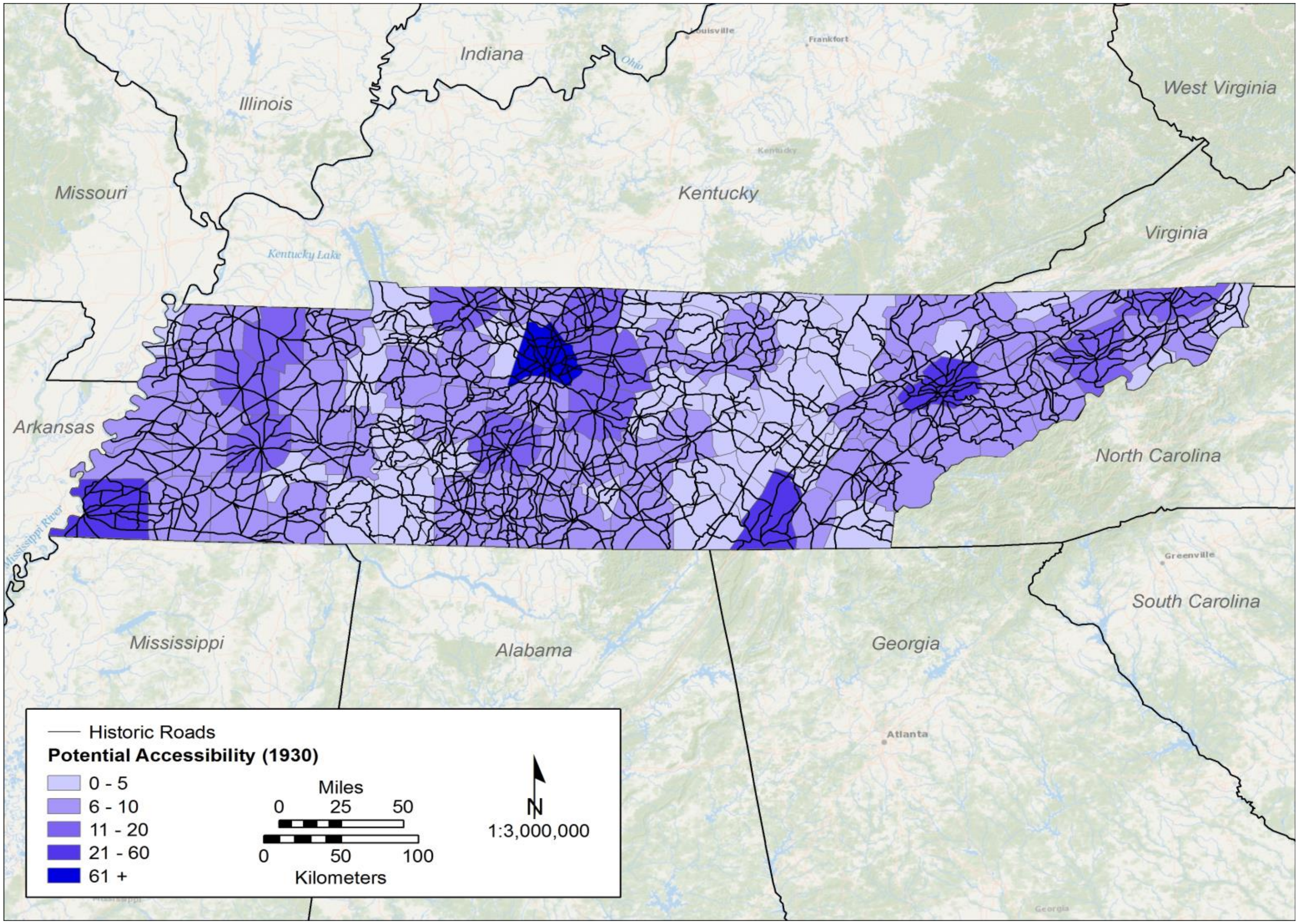

Figure 24. Potential Accessibility During Historic Roads Study Period

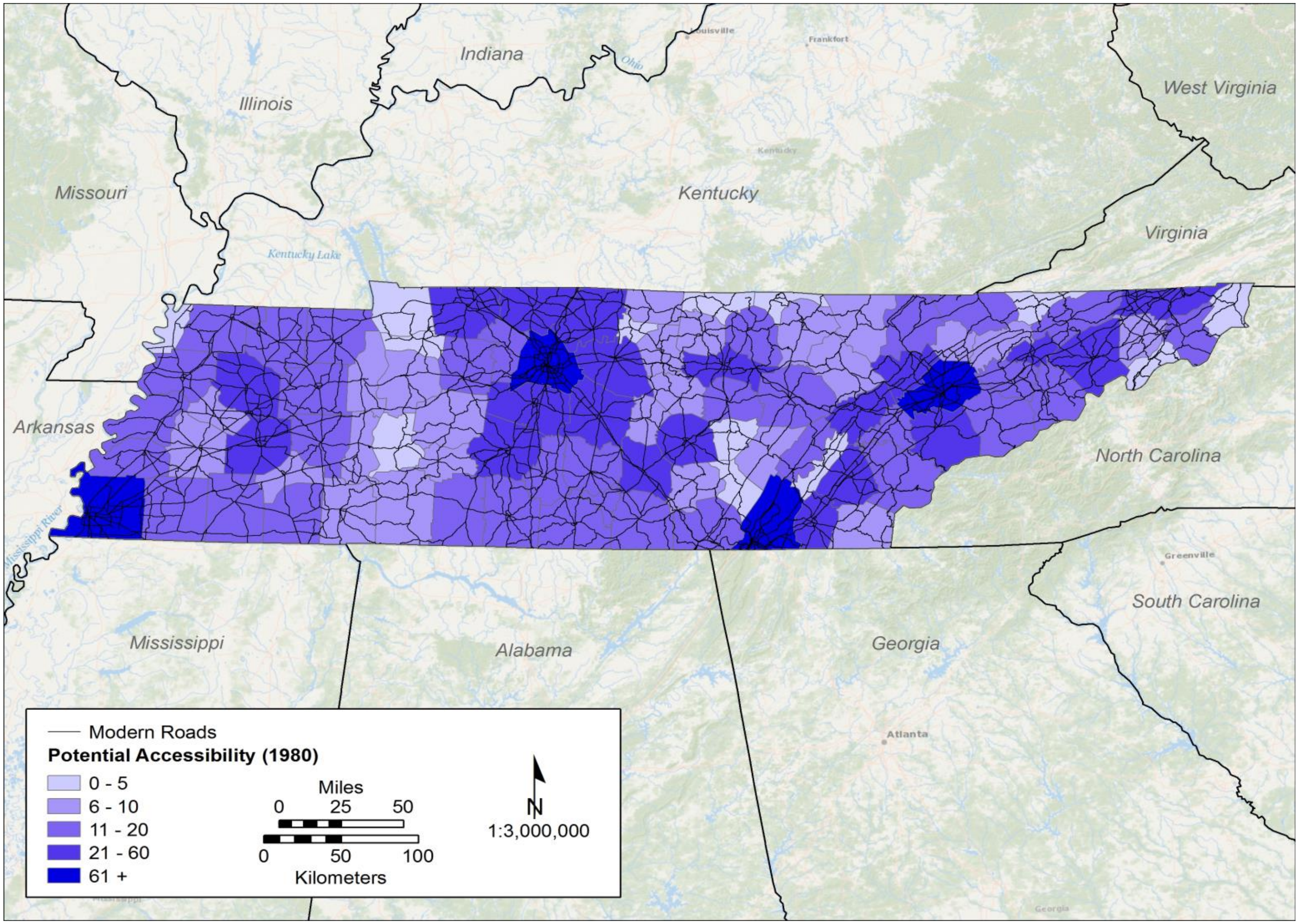

Figure 25. Potential Accessibility During Modern Roads Study Period

4.3 Conclusion

This study has looked at four distinct periods of transportation networks and infrastructure within the state of Tennessee over a 180-year time span in order to assess the potential influence of these networks on population migration at a county level. Additional variables (population share and mean city distance) were input along with A(P) total, with population share proving to be a highly significant indicator of population change in several of the study periods. Lastly, throughout the entire study, the mean geodesic distance to the largest cities remained insignificant as an influence on population growth.

CHAPTER 5: DISCUSSION AND CONCLUSION

Overall, the results only partially support the hypotheses that transport network accessibility in the state of Tennessee was causing the changes in population growth that were taking place throughout the entire study period 1830 through 2010.There were limits to accessible historical data, and even when available the data were often found to be inconsistent. There were limitations in historical data (i.e. port locations or specific construction year of railway sections) that required estimations and assumptions. This chapter concludes by discussing additional data that if included could improve the results of prospective future studies.

Specifically, the results of this study indicate that the importance of network accessibility is related to the analysis methods utilized and the years chosen for the study durations, which may have influenced the outcome of the study overall. The waterways proved to correspond to the expected hypothesis, where potential accessibility proved to be relevant. Railway networks, on the other hand, resulted in accepting the null hypothesis, where starting population share instead became the influential variable by indicating path dependency. In the third study period analyzing historic roads, both potential accessibility and starting population share proved to be significant factors behind population change, although the relationship was opposite to what had been theorized for A(P). It is not possible to determine from the models built whether one or the other has a greater influence or whether the real relationship is reciprocal and reinforcing. Finally, the potential accessibility was also correlated with population change during the modern road transport network. These results suggest that the starting share was also important when modeled independently, but was overruled by the strength of the accessibility potential. Collectively, the results of this study speak to the power and importance of network analysis for analyzing population movement over time. Likewise, while replicating findings in several earlier aforementioned studies (Wang et al. 2014; Kotavaara et al. 2011; Krugman 1991, Pavlínek 2003), this study reinforces the significance of using network analysis rather than a simple proximity analysis.

Overall, this thesis developed a unique dataset for the state of Tennessee over a significant time period that the author intends to be used for future demographic and historic analysis. This study could also be further expanded to include economic or larger-scale analysis within chosen regions of the state if more data such as employment or agricultural agglomeration were to become available across all decades. Another step in expanding this study would be to include network connections outside of the borders of Tennessee. Certain major connections (e.g., Chicago, New York, and St. Louis) most assuredly had influence on migration and growth based on Tennessee's locality. Additionally, a similar approach could be applied to other states or geographic regions outside of Tennessee for academic research or planning purposes.

5.1 Data Limitations and Assumptions

As mentioned previously, historical population data table formatting, data collection methodology and enumerator practice over the years has changed significantly, which proved to be an enormous challenge in data processing and structuring throughout the duration of this thesis. Many historic census sheets were unavailable or inconsistent and therefore often missing pertinent categories that may have further contributed to desirable variable input (i.e. economic, industrial, and population data). Furthermore, the unit of study used (county level data) generally disregarded more detailed demographic information (i.e. agglomerated city or census blocks) that may have provided further insight. Unfortunately, more detailed data was inconsistent and/or unavailable for the entire study duration, leading to the larger units of study used in this research.

Another limitation to the input data was the inability to correspond every newly generated transport feature with a precise construction date. For example, the historic railways were digitized from a georeferenced historic railway map of Tennessee of all railroads that existed by the end of the study period (1920), which lacked specific dates for each individual extension addition. Therefore, all railways were included for the overall study period, slightly skewing the data where perhaps certain transport networks had not actually existed in early railway development. This additionally leaves reason to cautiously accept the findings of the railway study period, as early railway development was slow to progress or ever really take off, unlike the waterways, which had already existed, and the road network, which has grown quite rapidly since the start of motorization.

5.2 Future Research

Future efforts to improve upon this study should include adding variables that could be influencing the population changes, such as economic agglomeration where available, particularly over study periods where predictability remains low. Furthermore, this research can easily be replicated in other states or regions, applying similar methodologies to different geographic regions. As mentioned earlier, the smallest aggregation of census data in a consistently available format stretching the entire length of this study (1830-2010) in regards to population was at the county level. This study could be reassessed over a shorter time frame considering a more recent time range in order to include more readily available census data at a block group level, but this would limit the overall time period encompassed by the study. Another option would be to use dasymetric mapping, which allows for population data to be combined with supplementary geographic layers to generate find-resolution approximations (Nagle et al. 2014). This could provide finer spacial specifics as to where population growth may be coming from and changing. Including economic variables could also provide further insight into factors behind population migrations due to the long-proven connection of economy with community growth driven by employment opportunities (Krugman 1991).

Additionally, including the transport networks outside of the state may be pertinent in understanding population trends throughout this study. This is particularly important, as previously mentioned, because Tennessee in land-locked and is connected to nine separate states, specifically changing as territories expand towards the west. Growth occurring at the peripheries of the state may indeed be experiencing high rates of growth from highly populated areas just outside the state, as is seen in the case of Montgomery County due to the location of Fort Campbell (Muir 2009). Additional work should consider including a secondary accessibility potential by analyzing connections between Tennessee population centers and those population centers outside of the state.

Lastly, it is intended that this study be used as an example of an approach to analyze population change over a significant amount of time based on a theory of potential accessibility. Specifically, this study expanded on the work conducted by Kotavaara et al. (2011), focusing on a region much different geographically and over a much longer time frame. Furthermore, this study helps to pave the way for other researchers to customize and apply similar methodologies in different geographic areas of interest.

BIBLIOGRAPHY

Allison, P.D. 1999. *Multiple regression: a primer.* Thousand Oaks, CA: Pine Forge Press.

Bowes, J.P. 1973. *The Trail of Tears : removal in the south.* Chelea House.

Bugromenko, V. N. 2010. "Modern transportation geography and transportation accessibility," ed. by Geogracom Consulting Firm. *Regional Research of Russia* 1(1): 27-34.

Cherokee Nation. *Treaty of Holston, 1791*. Available at http://www.cherokee.org/AboutTheNation/History/Facts/TreatyofHolston,1791.aspx (last accessed May 27, 2015).

Corlew, R.E., S.J. Folmsbee, and E.L. Mitchel. 1990. *Tennessee: a short history.* University of Tennessee Press.

Dancer, D. and A. Tremayne. 2006. "R-squared and prediction in regression with ordered quantitative response." *Journal of Applied Statistics* 32(5): 483-93.

Ding, C. 2013. "Transport development, regional concentration and economic growth." *Urban Studies* 50(2): 312-28.

Encyclopedia Britannica, Inc. 2012. "South, the vote on secession 1860-61." Encyclopedia Britannica, Inc. Available at http://www.britannica.com/EBchecked/topic/555542/the-South/images-videos/3841/south-the-vote-on-secession-1860-61 (last accessed May 27, 2015).

Erickson, E.P. 1974. "A regression method for estimating population changes of local areas." *Journal of the American Statistical Association* 69(348): 867-75.

Esri. 2013. Available at http://www.arcgis.com (last accessed May 27, 2015).

Garrau, J. 1991. *Edge city: life on the new frontier.* New York, NY: Doubleday.

Geurs, K.T., and J.R. Ritsema van Eck. 2001. "Accessibility measures: review and applications. Evaluation of accessibility impacts of land-use transportation scenarios, and related social and economic impact." (Rijksinstituut voor Volksgezondheid en Milieu RIVM).

Grengs, J. 2014. "Nonwork accessibility as a social equity indicator." *International Journal of Sustaibanle Transportation* 9(1): 1-14.

Hansen, W.G. 1959. "How accessibility shapes land use." *Journal of the American Institute of Planners* 25(2): 73-76.

Huang, R., and Y.D. Wei. 2002. "Analyzing neigborhood accessibility via transit in a GIS environment." *Geographic Information Sciences: A Journal of the Association of Chinese Professionals in Geographic Information Systems* 8(1): 39-47.

Johnston, R. 2009. "Friction of distance." *The dictionary of human geography*. Oxford: Blackwell Publishers.

Kotavaara, O., H. Antikainen, and J. Rusanen. 2011. "Population change and accessibility by road and rail networks: GIS and statistical approach to Finaldn 1970-2007." *Journal of Transport Geography* 19(4).

Kotavaara, O., H. Antikainen, M. Marmion, and J. Rusanen. 2012. "Scale in the effect of accessibility on population change: GIS and a statistical approach to road, air and rail accessibility in Finland, 1990-2008." *The Geographical Journal* 178(4), 366-382.

Krugman, P. 1991. "Increasing returns and economic geography." *Journal of Political Economy* 99(3): 483-99.

Leahy, J.F. 1934. "Railway distance map of the State of Tennessee." *David Rumsey Historical Map Collection*, 109. Chicago: American Hotel Register Co. Available at http://www.davidrumsey.com/luna/servlet/s/85uzv5 (last accessed May 27, 2015).

Long, J.H. 2000. *Atlas of historical county boundaries Tennessee*, ed. by P.T. Sinko. New York: Charles Scribner's Sons.

Lopes, S.B., N.C.M. Brondino, and A.N. Rodriques de Silva. 2014. "GIS-based analytical tools for transport planning: spatial regression models for transportation demand forecast." *ISPRS International Journal of Geo-Information* 3: 565-83.

Majors, W.R. 1980. *Volunteer trails: a program of visual aids for the study of Tennessee.* Meridional Publications.

Mather, M. 2004. *Households and families in Appalachia.* Demographic and Socioeconomic Change in Appalachia, Population Reference Bureau, Appalachian Regional Commission.

Muir, M. 2009. *Fort Campbell*. The Tennessee encyclopedia of history and culture. Austin Peay State University. Available at http://tennesseeencyclopedia.net/entry.php?rec=486 (last accessed May 27, 2015).

Nagle, N.N., B.P. Buttenfield, S. Leyk, and S. Spielman. 2014. "Dasymetric modeling and uncertainty." *Annals of the Association of American Geographers* 104(1): 80-95.

National Historical Geographic Information System. Available at https://www.nhgis.org (last accessed May 27, 2015).

National Map Company. 1927. "Sectional paved road map." *Dave Rumsey Historical Map Collection*, 12-13. Indianapolis, IN: National Map Company. Available at http://www.davidrumsey.com/luna/servlet/s/7aek98 (last accessed on May 27, 2015).

National Waterway Network. 2012. "Waterway Network (Internal)." *ArcGIS Online*. Available at http://www.arcgis.com/home/item.html?id=9884fe3876bd40158034fafd5e23381c (last accessed May 27, 2015).

New Deal Network. 2015. Available at http://newdeal.feri.org (last accessed May 27, 2015).

Ohtani, K. 1994. "The density functions of R^2 and adjusted R^2, and their risk performance under asymmmetric loss in missecified linear regression models." *Economic Modelling* 11(4): 463-71.

Park, C., and M. Allaby. 2013. *A dictionary of environment and conservation (2 ed.).* Oxford University Press.

Pavlínek, P. 2003. "Alternative Theoretical approaches to post-communit transformations in Central and Eastern Europe." *Acta Slavica Iaponica* 20: 85-108.

Pearl, R. "Population and longevity. 1924. " *Nature* 113(2835): 322-23.

Pierce, D. 2010. *Good roads movement*. The Tennessee encyclopedia of history and culture. Western Carolina University. Available at http://tennesseeencyclopedia.net/entry.php?rec=554 (last accessed May 27, 2015).

Rand McNally and Company. 1927. "Kentucky, Tennessee." *David Rumsey Historical Map Collection,* 32-33. Chicago: Rand McNally. Available at http://www.davidrumsey.com/luna/servlet/s/0fxn25 (last accessed on May 27, 2015).

Rothrock, M.U., and S.B. Smith. 1973. *This is Tennessee: a school history*. Nashville: Rothrock and Smith Publishers.

Schnell, G.A., and M.S. Monmonier. 1976. "U.S. population change 1960-70: simplification, meaning, and mapping." *The Journal of Geography* 75(5): 280-91.

Sellers, T. 2010. *Interstate highway system, Tennessee.* The Tennessee encyclopedia of history and culture. Tennessee Department of Transportation. Available at http://tennesseeencyclopedia.net/entry.php?rec=687 (last accessed May 27, 2015).

Sharp, L.N. 2010. *Dixie Highway* Association. The Tennessee encyclopedia of history and culture. Georgia Institute of Technology. Available at http://tennesseeencyclopedia.net/entry.php?rec=382 (last accessed October 2, 2014).

Shell Oil Company. 1956. "Shell highway map of Kentucky, Tennessee." *Dave Rumsey Historical Map Collection*. Chicago: Shell Oil Company. Available at http://www.davidrumsey.com/luna/servlet/s/rytg67 (last accessed May 27, 2015).

State Farm Insurance Companies Travel Bureau, and Rand McNally and Company. 1940. "Road map of Kentucky-Tennessee." *David Rumsey Historical Map Collection*, 46-47. Bloomington, IL: State Farm Insurance Companies Travel Bureau.

Sturtevant, W.C. 1966. *Early American Indian tribes, cultural areas, and linguistic stocks*. Smithsonian Institute.

TNGenWeb Project, Inc. "Territory of the United States south of the River Ohio county lines; 1790-91. Available at http://www.tngenweb.org/maps/census/tn1790.gif (last accessed May 27, 2015).

U.S. Census Bureau. Census of Population and Housing. 2010. Available at http://www.census.gov/prod/www/decennial.html (last accessed May 27, 2015).

Wang, H., et al. 2010. "Self-reported ethnicity, genetic structure and the impact of population stratification in a multiethnic study." *Human Genetics* 128: 167-177.

Wang, Y., A. Monzon, and F.D. Ciommo. 2014. "Assessing the accessibility impact of transport policy by land-use and transport interction model- the case of Madrid." *Computers, Environment and Urban Systems* 49: 126-135.

Wellman, G.C. 2014. "Transportation apartheid: the role of transportation policy in societal inequality." *Public Works Management & Policy* 19(4): 334-39.

Yigitcanlar, T., N. Sipe, R. Evans, and M. Pitot. 2007. "A GIS-based land use and public transport accessibility indexing model." *Australian Planner* 44(3): 30-37.

Yuwei, L and N. Pengfei. 2013. "Externalities, transport networks and the economic growth of urban clusters." *Social Sciences In China* 34(3): 174-94.

Zondag, B., M. de Bok, K.T. Geurs, and E. Molenwijk. 2015. "Accessibility modeling and evaluation: the TIGRIS XL land-use and transport interaction model for the Netherlands." *Computers, Environment and Urban Systems* 49: 115-25.