Factors Influencing Mode Choice for Intercity Travel from Northern New England to Major Northeastern Cities

June 2016

A Research Report from the National Center for Sustainable Transportation

Sean Neely, University of Vermont Transportation Research Center

National Center for Sustainable Transportation

The UNIVERSITY of VERMONT

About the National Center for Sustainable Transportation

The National Center for Sustainable Transportation is a consortium of leading universities committed to advancing an environmentally sustainable transportation system through cutting-edge research, direct policy engagement, and education of our future leaders. Consortium members include: University of California, Davis; University of California, Riverside; University of Southern California; California State University, Long Beach; Georgia Institute of Technology; and University of Vermont. More information can be found at: ncst.ucdavis.edu.

U.S. Department of Transportation (USDOT) Disclaimer

Acknowledgments

This study was funded by a grant from the National Center for Sustainable Transportation (NCST), supported by USDOT through the University Transportation Centers program. The authors would like to thank the NCST and USDOT for their support of university-based research in transportation, and especially for the funding provided in support of this project. Work summarized in this report benefited from the travel survey expertise of Resource Systems Groups (RSG Inc.). Project conception originated with Matthew Coogan of the New England Transportation Institute (NETI) and contributions by Thomas Adler and others of RSG are gratefully acknowledged. The contributions of Dr. Brian H. Y. Lee, Dr. Lisa Aultman-Hall, and Karen Sentoff are acknowledged.

Table of Contents

Factors Influencing Mode Choice for Intercity Travel from Northern New England to Major Northeastern Cities

EXECUTIVE SUMMARY

Long-distance travel from home locations in rural northern New England to major metropolitan areas in the Northeast United States is crucial for rural economies, multimodal planning and quality of life, yet limited research has been conducted on this type of travel. The research presented here is motivated by an interest in this type of travel behavior, and specifically considers travel from home locations in northern New England (Vermont, New Hampshire, Maine, and Massachusetts - excluding the Boston-Cambridge-Quincy Metropolitan Statistical Area), going to Boston, New York City, Philadelphia, and Washington, DC. The research makes use of a unique survey dataset from The Intercity Travel, Information, and Technology Survey Questionnaire. This body of work was co-funded by two USDOT University Transportation Center (UTC) programs.

The first set of results is recorded in a report to the University of Vermont (UVM) Transportation Research Center (TRC): *Intercity Travel in Northeastern Rural Regions of the U.S.* (Neely et al. 2015). The second part is reported here but is also the 2016 thesis by Sean Neely prepared in partial fulfillment of the requirements for the Master of Science degree specializing in civil and environmental engineering at the University of Vermont: *Factors Influencing Mode Choice for Intercity Travel from Northern New England to Major Northeastern Cities*.

The prior report considers the relationship between information access and attitudes about transportation options for this type of travel, using automobile, intercity bus, and passenger rail. The report explores relationships between access to information and attitudes about traveling from Northern New England to major cities in the Northeast US by automobile, intercity bus and passenger rail, accounting for gender, education level, and age group. For traveling to NYC by bus or train, an online planning tool as part of the survey was related to positive attitudes about scheduling flexibility and travel time for certain age and education groups. It was also related to negative attitudes about the ability to get and understand schedules for a bus or train to NYC. Further analysis will aim to better quantify the impacts of access to trip planning information, on attitudes about intercity travel by automobile, bus, and passenger rail.

A multimodal network dataset that was developed for the study region as part of this project is also described in the prior report. The network dataset will be used to develop an accessibility index across the study region. The index will incorporate measures, representing the level of accessibility, to large metropolitan areas, from outside of large metropolitan areas, by multiple modes of travel. The measures will incorporate the availability of each mode of travel, including the existence and frequency of service, number of transfers, and network travel time, calculated using the network dataset, for origins from across the region, going to the four destination cities (Boston, New York City, Philadelphia, and Washington, DC).

This report identifies and quantifies factors that influence mode choice for this type of travel, using automobile, intercity bus, passenger rail, and commercial air travel. Factors taken into consideration include gender, education level, age group, information access, technology use, latent attitudinal variables, and geographic variables, including land use, the built environment, and access to intercity transportation facilities. This research identifies and quantifies factors that influence people's mode choice (automobile, intercity bus, passenger rail, or commercial air travel), when traveling from non-metropolitan northern New England to large metropolitan areas in the Northeast. Factors were found to be significant from four categories: sociodemographic, geographic, attitudinal, and trip-specific. These included land use, distance to urban metropolitan areas, owning a tablet computer, and latent attitudinal factors. Age was only shown to be significant for non-business travel. Gender was only shown to be significant at the 90% confidence level. The only significant personal technology variable was owning a tablet computer. Three latent attitudinal variables strongly contributed to the models for both business and non-business trips: auto dependence, preference for automobile, and comfort with personal space and safety on public transportation.

Next steps in this research should be incorporating time and cost variables into a set of conditional mode choice models with similar remaining variables. Future research work should consider including seasonality and weather conditions, capturing potential general aviation travel, developing more specific attitudinal statements to expand latent factor analysis, and specifying whether the destination city is a final destination, a stop on a trip abroad, or just one leg of a domestic trip. Continuing to improve these types of long-distance intercity mode choice models for traveling from less populated areas to more populated areas will help planners, engineers, and policymakers to make more informed decisions about infrastructure, services, and financial investment for transportation systems in communities and regions around the country that may not have yet been well studied.

Abstract

Long-distance and intercity travel generally make up a small portion of the total number of trips taken by an individual, while representing a large portion of aggregate distance traveled on the transportation system. While some research exists on intercity travel behavior between large metropolitan centers, this report addresses a need for more research on travel behavior between non-metropolitan areas and large metropolitan centers. This research specifically considers travel from home locations in northern New England, going to Boston, New York City, Philadelphia, and Washington, DC. These trips are important for quality of life, multimodal planning, and rural economies. This research identifies and quantifies factors that influence people's mode choice (automobile, intercity bus, passenger rail, or commercial air travel) for these trips.

The research uses survey questionnaire data, latent factor analysis, and discrete choice modeling methods. Factors include sociodemographic, built environment, latent attitudes, and trip characteristics. The survey, designed by the University of Vermont Transportation Research Center and the New England Transportation Institute, was conducted by Resource Systems Group, Inc. in 2014, with an initial sample size of 2560. Factor analysis was used to prepare 6 latent attitudinal factors, based on 70 attitudinal responses from the survey statements. The survey data were augmented with built environment variables using geographic information systems (GIS) analysis. A set of multinomial logit models, and a set of nested logit models, were estimated for business and non-business trip mode choice.

Results indicate that for this type of travel, factors influencing mode choice for both business and non-business trips include trip distance; land use; personal use of technology; and latent attitudes about auto dependence, preference for automobile, and comfort with personal space and safety on public transportation. Gender is a less significant factor. Age is only significant for non-business trips.

The results reinforce the importance and viability of modeling long-distance travel from less populated regions to large metropolitan areas, and the significant roles of trip distance, built environment, personal attitudes, and sociodemographic factors in how people choose to make these trips for different purposes. Future research should continue to improve these types of long-distance mode choice models by incorporating mode specific travel time and cost, developing more specific attitudinal statements to expand latent factor analysis, and further exploring built environment variables. Improving these models will promote better planning, engineering, operations, and infrastructure investment decisions in many regions and communities across the United States which have not yet been well studied, possibly impacting levels of service.

1. Introduction

Significant research exists on travel behavior within large metropolitan centers, with considerably much less research on travel between large metropolitan areas (Miller, 2004). More limited still, is the amount of research on travel behavior for trips from less populated areas to large metropolitan areas. There is a need for more research on travel behavior between non-metropolitan areas and large metropolitan centers, because of its impacts on quality of life, multimodal planning, and rural economies. The research presented here is motivated by an interest in this type of travel behavior, and specifically considers travel from home locations in northern New England (Vermont, New Hampshire, Maine, and Massachusetts - excluding the Boston-Cambridge-Quincy Metropolitan Statistical Area), going to Boston, New York City, Philadelphia, and Washington, DC. When residents of northern New England plan a trip to major cities in the Northeast United States, there are often several transportation mode options to consider. This work identifies and quantifies factors that influence mode choice for this type of travel, using automobile, intercity bus, passenger rail, and commercial air travel. Factors taken into consideration include gender, education level, age group, information access, technology use, latent attitudinal variables, and geographic variables, including land use, the built environment, and access to intercity transportation facilities. This research identifies and quantifies factors that influence people's mode choice (automobile, intercity bus, passenger rail, or commercial air travel), when traveling from non-metropolitan northern New England to large metropolitan areas in the Northeast.

Travel behavior among certain age groups looks different now than in the past (Frändberg and Vilhelmson 2014; Hjorthol et al. 2010; Pooley et al. 2005). Many of today's older people are traveling more than prior cohorts did (Frändberg and Vilhelmson 2014; Hjorthol et al. 2010). Younger people today are traveling less than prior cohorts did (Frändberg and Vilhelmson 2014; McDonald 2015; Polzin et al. 2014).

Millennials (the generation born approximately during the years 1981-2000) have been shown to make fewer overall trips and to use automobiles less than other generations (McDonald 2015; Polzin et al. 2014). It is uncertain how this behavior might change in the future, which poses a challenge in terms of public investment for future transportation infrastructure (McDonald 2015; Polzin et al. 2014). Because of this, there has been a growing interest in the differences in travel behavior among age groups and generations. Some research exists on the general travel behavior of millennials compared with other generations, but there is a strong need to further study the differences in travel behavior between these groups. There is a need for more research on the intercity travel behavior of millennials compared with other generations.

Differences in travel behavior, including mode choice, have been shown between males and females (Frändberg and Vilhelmson 2014; Mattson et al. 2010; Presser and Hermsen 1996). Mattson et al. (2010) showed that gender significantly influences mode choice for rural intercity transportation, with males more likely to choose automobile, and females more likely to choose train or van for rural intercity travel. Frändberg and Vilhelmson (2014) showed a persistent, yet converging gap in long-distance travel between genders, with women traveling more and men traveling less than they did in 1978.

This research work compares the mode choices among age groups and between genders, for intercity travel during 2014, going from home locations in less populated areas of northern New England (Vermont, New Hampshire, Maine, and Massachusetts - excluding the Boston-Cambridge-Quincy Metropolitan Statistical Area), to large metropolitan areas (Boston, New York, Philadelphia, and Washington, DC).

Chapter 2 of this report presents the relevant background and literature review. Understanding long-distance and intercity travel behavior plays an important role in the planning, engineering, and operations of the transportation system, as well as the associated decision-making for long-term financial investment. These trips are important for quality of life, multimodal planning, and rural economies, creating a need to better understand intercity and long-distance travel from less populated areas to large metropolitan ones. Developing discrete mode choice models is a means to this end. Incorporating latent attitudinal variables in mode choice models can help to improve them (Ashok et al. 2002; Daly et al. 2012; Mattson et al. 2010; Popuri et al. 2011; Walker and Ben-Akiva 2002). There is also an opportunity to better understand the intercity travel behavior of different age groups and generations, including travel from less populated areas to large metropolitan ones.

Chapter 3 of this report will introduce and describe the survey instrument and survey sample which provide the source of the unique primary dataset used in this research. The Intercity Travel, Information, and Technology Survey Questionnaire was part of a project by the University of Vermont's Transportation Research Center (UVM TRC) and the New England Transportation Institute (NETI), with funding from the US Department of Transportation (USDOT) University Transportation Center Program. The survey was conducted, on behalf of the UVM TRC and NETI, by Resource Systems Group, Inc. (RSG, Inc.) in May 2014.

The survey addressed trips from northern New England to four major cities in the Northeast: Boston, New York City, Philadelphia, and Washington, DC. The survey had questions about actual trips taken, a hypothetical trip to New York City, attitudes about traveling in general, and attitudes about traveling specifically by automobile, intercity bus, and passenger rail. The survey data include revealed preferences for the respondents' most recent trip taken, and other trips taken during the previous year, to Boston, New York City, Philadelphia, and Washington, DC, by automobile, intercity bus, passenger rail, and air. The network distance between home locations and the most recent destination city ranged from 27 miles to 848 miles. The survey data also include stated preferences about traveling to New York City by automobile, intercity bus and passenger rail.

Figure 1-1 displays the study origin area, destination cities, other regional cities, airport hubs, Amtrak railways, and interstate highways. While airport hubs are seen in each of the four destination cities, they are sparsely distributed across the study origin area. Once in the corridor from Boston to Washington, DC, Amtrak railways connect each of the destination cities with each other, but there are only a few Amtrak railways running through the study origin area. Interstate highways are also prevalent within the corridor connecting the destination cities with each other, while only a few freeways exist through the study origin area. In general, transportation options appear rich for travel between the destination cities, while fewer transportation options appear for travel from the study origin area to the destination cities.

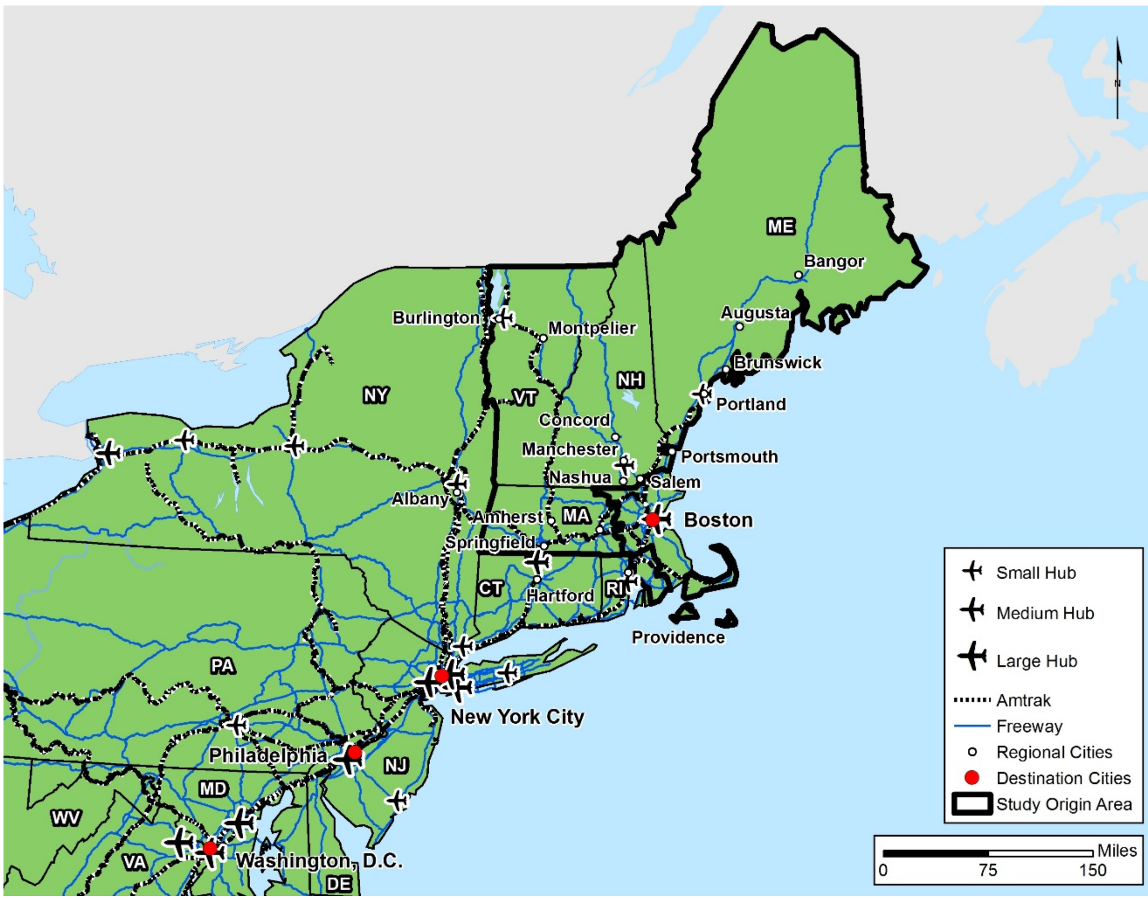


Figure 1-1: Study Area: Origin Area and Destination Cities

Sources: Zip Codes and Cities from ESRI; Background data from © OpenStreetMap contributors

This unique dataset, of survey responses from respondents across northern New England, contains interesting variables that include a series of 70 attitudinal statements about different aspects of the transportation experience. Making use of this distinctive dataset, including latent attitudinal variables, identified through factor analysis, described in Chapter 4, added substantial value to this analysis.

Chapter 4 describes data tabulation, validation, and augmentation performed by colleagues and by the author. Preliminary descriptive analysis of the survey response data, such as the distributions of travel modes for respondents' most recent trip to one of the destination cities, and consideration of which respondents and choices to include in the analysis are described. Data augmentation includes adding supplemental sociodemographic, land use, and transportation facility access variables with geographic information systems (GIS) analysis, as well as latent attitudinal factor analysis to reduce attitudinal survey statement responses to a smaller set of latent attitudinal factors for use as explanatory variables in primary mode choice model building.

Chapter 5 presents the modeling methodology used to answer the research questions. This includes discrete mode choice model building to identify and quantify the influence of numerous factors on mode choice, for traveling from northern New England to large metropolitan areas in the Northeast. A set of multinomial logit (MNL) models, and a set of nested logit (NL) models were estimated for business and non-business trips.

Contributions of this report, presented in Chapter 6, are the identification of factors and quantification of their influence, using a unique dataset, on mode choice for intercity travel originating from homes in less populated areas of northern New England, going to large metropolitan destinations in the Northeast. This helps to improve our understanding of mode choice for multimodal planning efforts, with potential benefits to rural economies and ultimately quality of life. Factors were found to be significant from four categories: sociodemographic, geographic, attitudinal, and trip-specific. These included land use, distance to urban metropolitan areas, owning a tablet computer, and latent attitudinal factors. Age was only shown to be significant for non-business travel. Gender was only shown to be significant at the 90% confidence level. The only significant personal technology variable was owning a tablet computer. Three latent attitudinal variables strongly contributed to the models for both business and non-business trips: auto dependence, preference for automobile, and comfort with personal space and safety on public transportation.

Another contribution of this research is a better understanding of the differences in mode choices between genders and among age groups, in the context of intercity travel from homes in less populated areas in the study region, going to large metropolitan areas in the Northeast. These differences were looked at both with and without controlling for other factors. Slightly more males made business trips than did females. Over forty percent more females made non-business trips than males did. Males and females chose automobile close to the same amount for both business and non-business trips. Males chose airplane almost twice as often as females did for business trips, and bus slightly more than females. Females chose rail almost twice as often as males did for business trips. Females and males chose airplane in about equal amounts for non-business trips. More females chose both intercity bus and intercity rail than males did for non-business trips, with the difference being greatest for bus.

Next steps in this research should be incorporating time and cost variables into a set of conditional mode choice models with similar remaining variables. Future research work should consider including seasonality and weather conditions, capturing potential general aviation travel, developing more specific attitudinal statements to expand latent factor analysis, and specifying whether the destination city is a final destination, a stop on a trip abroad, or just one leg of a domestic trip. Continuing to improve these types of long-distance intercity mode choice models for traveling from less populated areas to more populated areas will help planners, engineers, and policymakers to make more informed decisions about infrastructure, services, and financial investment for transportation systems in communities and regions around the country that may not have yet been well studied.

2. Background and Literature Review

There are several ways to define long-distance travel, without a current consensus among those in the research community. Definitions vary from using a distance threshold, ranging from 50 to 200 miles, to only including intercity or interregional travel (Aultman-Hall et al. 2015), or only considering non-routine travel. For some people though, a trip of 50 miles might not even cover their one-way workday commute, or long-distance travel might just be a regular part of their routine. Most definitions of long-distance travel would likely consider trips originating in home locations in northern New England (Maine, New Hampshire, Vermont, and Massachusetts outside of the Boston Metro Area), and going to Boston, New York City, Philadelphia, and Washington, DC, to be made up mostly of long-distance travel. The network distance between home locations and the most recent destination city in this study visited by respondents, which is likely to be the trip most accurately recalled, ranged from 27 miles to 848 miles.

Long-distance travel generally makes up a small percentage of total trips taken by an individual, but it makes up a large percentage of aggregate distance traveled, with significance in terms of travel demand and modeling, economic and environmental impacts, including congestion and emissions (Axhausen 2000; Bierce and Kurth 2014; LaMondia et al. 2014; Moeckel et al. 2013). As such, long-distance travel plays an important role in the planning, engineering, and operations of the transportation system, as well as the associated decision-making for long-term financial investment. In recent years, the US Department of Transportation (USDOT) Federal Highway Administration (FHWA) has been interested in better understanding the factors that influence mode choice for long-distance travel (Anderson and Simkins 2012; Outwater et al. 2015). There remains a need for studying long-distance travel behavior and accounting for differences in behavior. Developing better long-distance mode choice modeling is a means to this end, which has been limited by the shortage of data, a main barrier for modeling and planning (Lamondia et al. 2014; Moeckel et al. 2013).

Residents in less populated areas generate more long-distance trips than residents in more populated areas (Bierce 2014; Daly et al. 2012; Polzin et al. 2014). Anderson and Simkins (2012) show that mode choice for overall long-distance travel in the US is heavily influenced by fixed attributes like home location and demographics, compared with factors more prone to fluctuation, like the costs and benefits of each mode. Mattson et al. (2010) show that rising fuel costs do impact stated preference mode choice for long-distance travel originating in rural areas, resulting in a mode-shift to public transportation modes, like intercity bus and rail. For states such as those in this research, with a large amount of rural area, like Maine, New Hampshire, Vermont and Massachusetts, understanding factors that influence mode choice for long-distance travel is important for long-term transportation planning and engineering of the regional transportation system, including passenger mode infrastructure and levels of service, and to support quality of life in the study area for a changing population, including an aging demographic.

Modeling Mode Choice

Anderson and Simkins (2012) built a set of multinomial logistic regression models to better understand the factors that influence mode choice for nationwide long-distance travel, using the 2001 National Household Travel Survey (NHTS) as a primary data source. Explanatory factors they considered include: demographic, trip duration, purpose, stations, terminals, highways, and mode availability. Their consideration of explanatory factors showed emphasis on socioeconomic factors such as age, income, gender, and urban/rural designation, as well as land-use factors, and a tendency to consider how the relationship between these factors and mode choice changes with trip purpose. Separate models were estimated for each trip purpose: business, pleasure, and personal business. Model results predicted personal auto and commercial air service rather well, but did not predict intercity bus and rail use very well. This is an indication that the survey data may not have been sufficient for predicting these modes (Anderson and Simkins 2012). The NHTS long-distance data include personal travel attributes and traveler demographics at the national level, with half of long-distance trips using the state level for origin-destination, and the highest resolution level being the Metropolitan Statistical Area (MSA). They do not include mode-specific attributes like travel time and cost, and origin-destination data are aggregated. These traits of the NHTS data lead researchers to supplement the data with other sources, or limit analysis for MSAs only (Anderson and Simkins 2012).

Ashiabor et al. (2007) built nested and mixed logit models for intercity transportation throughout the US, for auto and commercial air travel. Because it is a nested logit model, more transportation modes can be added, as survey data are available. Models were built separately for business and nonbusiness travel. The models used utility functions, with travel time, travel cost, and household income as independent variables. They built a set of nested and mixed logit models, using the 1995 American Travel Survey (ATS). The ATS data include long-distance trips greater than 100 miles in one direction. The data contain 556,026 records for person-trips, with 348 variables. This large dataset does not have travel time and travel cost, so these variables were generated synthetically for the study. ATS mode share data indicate people prefer faster travel modes for greater distances, with income level being a contributing factor (Ashiabor et al. 2007).

Moeckel et al. (2013) developed an aggregate discrete mode choice nested logit model for long-distance travel, using the 2001 NHTS long-distance dataset. Automobile modes were nested together, and transit modes were nested together. The model includes independent variables for travel costs, distance, accessibility to transit, frequency of service, amount of transfers, and costs of parking. Modes included were single-occupancy vehicle, ride-sharing for 2, 3, and 4+ passengers, regional bus, rail, and commercial air service. A sensitivity analysis is described for increasing fuel costs, bus service improvements, and to make sure no individual parameter determines the outcome of the model (Moeckel et al. 2013). Their model is also applied to the North Carolina Statewide Transportation Model (NCSTM).

Moeckel et al. (2013) emphasized advantages of using a nested logit model versus a more traditional multinomial logit model, including mitigating for the Independence of Irrelevant Alternatives (IIA). The model developed in their study was built with the goal of overcoming limitations in transferring existing models for long-distance applications. Because comprehensive statewide data for long-distance travel in North Carolina were not available, parameter values were determined heuristically, instead of econometrically. The observed mode shares have high constants, putting limits on their policy sensitivity, which concern the authors, even though the constants are still lower than those found in comparable long-distance mode choice models from the literature. These relatively high constants may result from bias in the observed mode split data, which are difficult to collect for long-distance travel (Moeckel et al. 2013).

Lee et al. (2015) built a series of discrete mode choice models to examine rideshare mode potential in non-metropolitan areas of the northeastern US. Models included binomial logistic regression, multinomial logit, and nested logit models. Variables examined include sociodemographic, attitudinal, and built environment factors. They used survey questionnaire data with 1,795 total participants from Maine, New Hampshire (outside of Boston commuter shed), Vermont, and Upstate New York (outside of New York commuter shed). The emphasis was on rideshare commuters, and potential rideshare commuters, for home-to-work trips. Results suggested that people who currently rideshare, compared to people who could benefit from a formal ridesharing program, are different in terms of home and work locations.

Incorporating Attitudes

Behavioral scientists have criticized traditional simplified discrete mode choice models for not accounting for decisions that appear irrational, except as imperfect knowledge or unobserved utility (Daly et al. 2012). Ashok et al. (2002) indicate that attitudes, as latent variables, are as important of a factor in discrete choice as are the attributes of the alternatives. Attitudes take time to form, and are influenced by personal experiences and external factors, like socioeconomic factors (Walker and Ben-Akiva 2002).

Researchers have used various methods to incorporate attitudes of decision makers into discrete mode choice models. Mattson et al. (2010) were unsuccessful when trying to incorporate attitudinal variables in their mixed logit model of long-distance travel, as it resulted in unreliable estimates. Instead, a set of binary logit models was developed for each mode, with the outcome equal to 1 for choosing the mode, and equal to 0 for not choosing the mode. All attributes describing the traveler, the trip, and the mode were added as explanatory variables, in addition to the 28 survey responses on attitude. Attitudinal variables had Likert scale values from 1 to 10, with greater values showing more agreement with the survey statement (Mattson et al. 2010).

Daly et al. (2012) applied attitudinal and choice models to a passenger rail transportation study, focusing on the significance of latent attitudinal variables, which is difficult because they are not directly measurable as are socio-demographic variables. Their primary contribution is the use of ordered logit models to incorporate the ordinal nature of the attitudinal variables, compared to the typical use of continuous attitudinal variables. The work aims to continue improving on traditional simplified discrete mode choice models, by incorporating attitudinal variables. Latent attitudinal variables can be derived from indicators, such as survey responses about attitudes, based on a Likert scale. The authors define attitudes as a reflection of latent variables which correspond to the decision-makers' characteristics, reflecting their "needs, values, tastes, and capabilities". Models incorporating latent attitudes have not been used very much in transportation or other fields, possibly because the theoretical work ranges across many fields of study. The models in the Daly et al. study use simultaneous estimation of the latent attitude model and the choice model, for consistency and efficiency of estimates. The survey data they used captured stated choice preference between different scenarios of rail travel, concerning levels of security, wait time, cost, etc. Attitude questions focused on policies with effects on personal privacy.

In the work of Popuri et al. (2011), factor analysis was used to reduce attitudinal responses, from 23 survey statements about daily travel to work, to six factors. The data source was a survey of daily commuter transportation in northeastern Illinois, with outcome choices of transit or auto, and a final sample of 868 respondents included in the analysis. The six attitudinal factors were combined with explanatory variables for socioeconomics, travel times, and travel costs, to build a binary logistic regression model for mode choice of transit or auto. Finally, the attitudinal factors were ranked in decreasing order of computed elasticities of transit mode choice to each attitudinal factor. Results indicated that the attitudinal factors helped improve the model, in terms of intuitiveness and goodness-of-fit. Also, levels of stress and productive use of travelers' time were shown to be important.

There are two reasons for not using the original survey attitudinal statements as variables in a mode choice model (Popuri et al. 2011). The first reason is the large amount of correlation between statements. The second reason is the negative effect on model parsimony. The factor analysis used to reduce the attitudinal statements to a smaller number of latent factors began with computing pairwise Pearson correlation coefficients between each of the 23 survey statements. Principal component analysis was then used to estimate factor loadings. This resulted in six factors retained based on both professional judgement and amount of total variance in the original variables explained by each factor. The factors were rotated, using the Varimax technique (Kim and Mueller 1978), resulting in loading each variable heavily onto a single factor, enabling easier interpretation. Factor loadings were used to compute the factor scores; the six resulting attitudinal factors were then used in the mode choice model as explanatory variables (Popuri et al. 2011). In summary, attitudinal factors are important explanatory variables for mode choice models and factor analysis is the key to their incorporation.

Age Groups and Generations

Variance in travel behavior among age groups looks different now than it has in the past (Frändberg and Vilhelmson 2014; Hjorthol et al. 2010; Pooley et al. 2005). Many of today's older people are traveling more than prior cohorts did (Frändberg and Vilhelmson 2014; Hjorthol et al. 2010). Younger people today are traveling less than prior cohorts did (Frändberg and Vilhelmson 2014; McDonald 2015; Polzin et al. 2014). It is unclear due to limited data if this trend applies to long-distance travel.

The trends shown by millennials are changing US demographics, and will impact travel behavior moving forward (Polzin et al. 2014). Planning for future transportation infrastructure, which can take over a decade for large-scale projects, and often takes into account forecasts of travel demand for 25 to 50 years or more, requires understanding future travel behavior of millennials, especially because most cohorts have traveled most during middle age (Polzin et al. 2014).

Millennials have been shown to use automobiles less than other generations, but also to make fewer trips overall (McDonald 2015). The differences of millennials, compared with other generations, are correlated with differences in vehicle miles traveled (VMT), and other factors that include: residential location, race, employment, education, income, household characteristics, marital status, driver licensing, vehicle availability, values, and technology use (Polzin et al. 2014). Numerous millennials are staying in their parents' residences longer, waiting to become licensed drivers, waiting to marry and/or have children, and using technology as substitutes for travel (Polzin et al. 2014). If millennials are only holding off on significant milestones such as marriage, and raising children, until a later time, there is greater uncertainty as to what travel decisions will be made, in terms of trip generation and mode choice, as more millennials choose to raise families. This uncertainty poses a challenge in terms of public investment for future transportation infrastructure (McDonald 2015).

Polzin et al. (2014) addressed several research questions, including: How do travel behaviors of millennials differ from prior generations' behaviors? Their work was based heavily on the National Household Travel Survey (NHTS) data, but incorporated various data sources and time periods from 2001 to 2009. They recognized that better quality data and more data would help to support more rigorous statistical analysis in this research area. Alternative data sources, like state household travel surveys, would be valuable to help improve understanding of millennials' travel behavior, compared with other generations, moving forward (Polzin et al. 2014).

Gender

The literature shows an emphasis in considering gender among the factors influencing mode choice (Anderson and Simkins 2012; Frändberg and Vilhelmson 2014; Mattson et al. 2010; Presser and Hermsen 1996). Presser and Hermsen (1996) showed differences among genders for work-related overnight trips, regarding job attributes and other background characteristics. Mattson et al. (2010) showed that gender significantly influences mode choice for rural intercity transportation, with males more likely to choose automobile, and females more likely to choose train or van for rural intercity travel. Frändberg and Vilhelmson (2014), considering 30 years of Swedish travel data, including long-distance trips, showed a persistent, yet converging gap in travel between genders, with women traveling more now and men traveling less now than they did in 1978.

Aims of Current Research

This research aims to contribute to the literature by filling the gap in research on long-distance travel originating in less populated areas and going to large metropolitan areas, considering differences among age groups and between genders, incorporating latent attitudinal variables into logit discrete choice models of mode choice.

3. Survey Data

3.1 Survey Sampling Strategy

The Intercity Travel, Information, and Technology Survey Questionnaire was a project by the University of Vermont's Transportation Research Center (UVM TRC) and the New England Transportation Institute (NETI). The survey was conducted, on behalf of the UVM TRC and NETI, by Resource Systems Group, Inc. (RSG, Inc.) in May 2014. The survey instrument can be found in Appendix A. Data from the survey were used for the research presented in this report and are summarized in Appendix B. The survey concerned trips from northern New England to four major cities in the Northeast: Boston, New York City, Philadelphia, and Washington, DC. Surveying took place from May 1 through May 16, 2014. Respondents were recruited via email by Research Now, an online research firm based in Plano, Texas, and directed to RSG, Inc.'s survey platform.

The travel survey sampling protocol relied on respondent panels from Research Now to include residents from four New England states: Maine, New Hampshire, Vermont, and Massachusetts, outside of the Boston metropolitan area (Boston-Cambridge-Quincy Metropolitan Statistical Area (MSA)). There is no indication that a sample strategy other than random sampling was used. A total of 2560 valid survey responses were collected.

3.2 Survey Instrument

The survey instrument, found in Appendix A, had questions on household demographics, actual trips taken, a hypothetical trip to New York City, and attitudes about traveling in general and about traveling by automobile, airplane, intercity bus, and passenger rail. There were a total of 98 questions plus a home zip code question that determined respondent eligibility for inclusion in the survey. The survey was organized into four parts, shown in Table 3-1.

Table 3-1: Survey Structure

Part		
1	Recent intercity travel trips and general travel preferences	
	1-A	Questions about recent trips
	1-B	Questions about the survey respondent's most recent trip to: Boston, New York City, Philadelphia, or Washington, DC
	1-C	General travel and communication questions about the survey respondent and their household
2	Travel preferences	
3	A hypothetical trip to New York City – stated preference	
4	Demographics about the survey respondent and their household	

Part 1 of the survey asked 13 questions about recent intercity travel trips and general travel preferences. For many questions, respondents were able to "select all" relevant answers from a list. For example, selecting which modes of transportation they have used for recent trips. Other questions allowed respondents to choose a relevant frequency or quantity (e.g., the number of trips to each city in the last twelve months, or the number of people and licensed drivers living in their household).

Part 2 included a list of 35 statements, shown in Table 3-2 about intercity travel preferences, many regarding a specific utility or disutility pertaining to a certain mode. Respondents were asked to select how much they agree or disagree with each statement on a Likert scale from 1 (completely agree) to 7 (completely disagree). This method elicits beliefs and values, in order to explore foundations of attitudes; it comes from the Theory of Planned Behavior (TPB) (Ajzen, 1991). Statements were randomized for each respondent and shown ten at a time.

Table 3-2: General Mode Attitudinal Survey Statements for Factor Analysis

I feel I am less dependent on cars than my parents are/were.
I need to drive my car to get where I need to go.
I love the freedom and independence I get from owning one or more cars.
It would be hard for me to reduce my driving mileage.
For me to be able to leave the driving to someone else (e.g., a bus driver) would be desirable.
It would be desirable for my household to be able to have fewer cars.
Being able to freely perform tasks, including using a laptop, tablet, or smartphone is an important reason for me to choose bus or train travel.
Having reliable WiFi internet access while I travel on a bus or train is important to me.
When taking a bus or train, being able to plan my trip and buy tickets online is important to me.
It would be important to me to receive email or text message updates about my bus or train trip.
I find tablet or smartphone apps for travel and trip planning to be helpful.
When the government tries to improve things, it never works.
If everyone works together, we could improve the environment and future for the earth.
People like me take the bus or the train.
I would be willing to pay more when I travel if it would help the environment.
I tend to use the fastest form of transportation, regardless of cost.
For me, the whole idea of being on a bus or train with other people I do not know seems uncomfortable.
I enjoy being out and about and observing people.
I don't mind traveling with people I do not know.
Having my privacy is important to me when I travel.
When I choose a home, I value having adequate space for parking two or more cars.
When I choose a neighborhood to live in, I like to be able to walk to a commercial or village center.
Living in a multiple family building (e.g., apartment, condo) wouldn't give me enough privacy.
I like living in a neighborhood where there is a lot going on.
I am confident that if I want to, I can do things that I have never done before.
I worry about crime or other disturbing behavior on buses and trains, or while walking in and around the stops/stations.
It is important to me to control the radio and the air conditioning in the car.
I feel really stressed when driving for a long time in congestion in and around big cities.
I prefer to use the most comfortable transportation mode regardless of cost or time.
Having a low-stress trip is more important than reaching my destination quickly.
I get very annoyed being stuck behind a slow driver.
I am usually in a hurry when I make a trip.
With my schedule, minimizing time spent traveling is very important to me.
I would use the bus or train more often if it were cheaper to ride.
Rather than owning a car, I would prefer to borrow, share, or rent a car just for when I need it.

Part 3 presented a hypothetical scenario, in which someone has asked the respondent to travel from their home to Manhattan, in New York City (NYC), for an important appointment during the following month, and the respondent has decided to go. They would stay one night at a hotel and travel alone. The host would pay for the hotel costs, but not for travel. The respondent would be responsible for all costs of gas, parking, or any fares. The respondent was asked to assume that, for one reason or another, they had already decided that they would not take any part of the trip by plane. They would then need to choose between taking the entire trip by car (whether or not it was their own vehicle) and taking at least part of the trip by intercity bus or train.

All respondents were asked to select what mode(s) of transportation they thought were available to them for this trip to NYC, how likely they would choose to take a bus or train for a trip like this to NYC, and whether learning that no WiFi or electrical outlets were available on the bus or train would make them less likely to choose a bus or a train for this trip.

At this point, approximately halfway through completion of the survey, respondents were randomly selected to be in one of two groups: a control group or a test group. Random bias was checked to select an even split within each state of residence. The test group had access to an intercity travel planning web tool, designed with this survey by RSG, Inc. The tool had scheduling options for traveling to New York City by intercity bus and rail. The control group did not have access to the planning tool.

The test group was then provided a link to review the web tool related to their hypothetical trip to NYC. After having reviewed the web tool, respondents were asked to close the web tool and proceed with the remainder of the survey. The test group was then provided with four statements about travel options and information availability, and asked to select how much they agree or disagree with each statement on a similar Likert scale as earlier. Next, both groups were asked to continue imagining the trip to NYC, and were given another series of 35 statements, shown in Table 3-3 about attitudes related to intercity travel, to select their level of agreement on the same scale. These 35 attitudinal statements were included as attributes in the factor analysis for this report. The test and control groups are mentioned here because they were an important component of the survey (Neely et al. 2015). However, they were not used in this research. The survey technical report explores differences in attitudes and stated preferences overall, and broken down by gender, education level, and age group, for each section of the survey, as a result of access to information and technology, based on the control and test groups (Neely et al. 2015).

Table 3-3: NYC Trip Attitudinal Survey Statements for Factor Analysis

When I drive long distances (like from my home area to NYC), I can get tired and stressed.
I worry about the difficulty in finding a parking space at a reasonable cost when I get to NYC.
I am concerned that the schedule of the bus or train only lets me travel a few times per day, and I need to be flexible.
I could deal with the limited schedules offered by a bus or train for this trip from my home to NYC.
I like the idea that I might see and meet new people on a bus or train to NYC.
I don't like the idea of riding with a lot of people that I don't know on a bus or train.
If I took a bus or train to NYC, I might have to be with people whose behavior I find unpleasant.
I could be with other people who share my values when I take a bus or train on a trip like this.
I think that taking a BUS to NYC would take a lot longer than driving.
I think that taking a TRAIN to NYC would take a lot longer than driving.
Without thinking about it much, I would guess that the cost of taking the trip by BUS would be less than the cost of the car trip (including gas, tolls, and parking).
Without thinking about it much, I would guess that the cost of taking the trip by TRAIN would be less than the cost of the car trip (including gas, tolls, and parking.)
It would be really important to me to minimize costs when I plan this trip to NYC next month.
I really want to minimize the time I spend on the trip to NYC, even if that means more stress or higher costs.
Being able to use my laptop, tablet, or smartphone when traveling makes me more interested in taking a bus or train to NYC.
I am the kind of person who would take my own car to NYC.
Most people whose opinions I value would approve of my taking this trip by bus or train.
My family would think that I should take this kind of trip by car or plane.
My colleagues would likely think that it is strange not to go by a car or plane to NYC.
When my friends go to NYC, they always take a bus or train.
When my family members go to NYC, they always take a bus or train.
It might be unsafe to make this trip by bus or train.
The experience at the NYC bus or train station would be so unpleasant that I would try to avoid it.
It would be easy for me to get the schedules for a bus or train between here and NYC, and I would understand them.
I like the idea of taking a bus or train instead of driving for this trip to NYC.
I think that the most RATIONAL choice would be to take a bus or train instead of a car.
I think that the most PLEASURABLE choice would be to take a bus or train instead of a car.
I think that the most STRESSFUL choice would be to take a bus or train instead of a car.
All other things being equal, if a bus was cheaper, but less reliable than a train, I would choose to take a bus.
I am confident that if I wanted to, I could take a bus or train for such a trip to NYC next month.
I would make an effort to choose a bus or train for such a trip to NYC next month.
For me to take a bus or train for such a trip to NYC the next month would be impossible.
In this imaginary situation, I would plan to take a bus or train for this trip to NYC next month.
I would trust the person who invited me to NYC to recommend how I should travel.
I don't know all the things I NEED to do to make this trip work by bus or train.

Respondents were then asked how likely they were, on the seven-point Likert scale, to choose intercity bus or train for a trip to NYC the next month, like the one described in the hypothetical situation. For test group members who indicated a different level of likeliness to take intercity train or bus to NYC, than they had earlier, they were asked to comment on the reasons why, and were provided an open-ended comment field. Respondents were then asked how seriously they would consider taking a bus or train to NYC, in real life.

Part 4 included five questions about what personal technology devices respondents own, and their demographics: age group, gender, level of education, and annual household income level.

3.3 Survey Sample

The survey sample was made up of residents from northern New England, including Maine, New Hampshire, Vermont, and Massachusetts, outside of the Boston metropolitan area (Boston-Cambridge-Quincy Metropolitan Statistical Area (MSA)). Figure 3-1 shows the study area, made up of zip code locations for survey respondents, and the four destination cities.

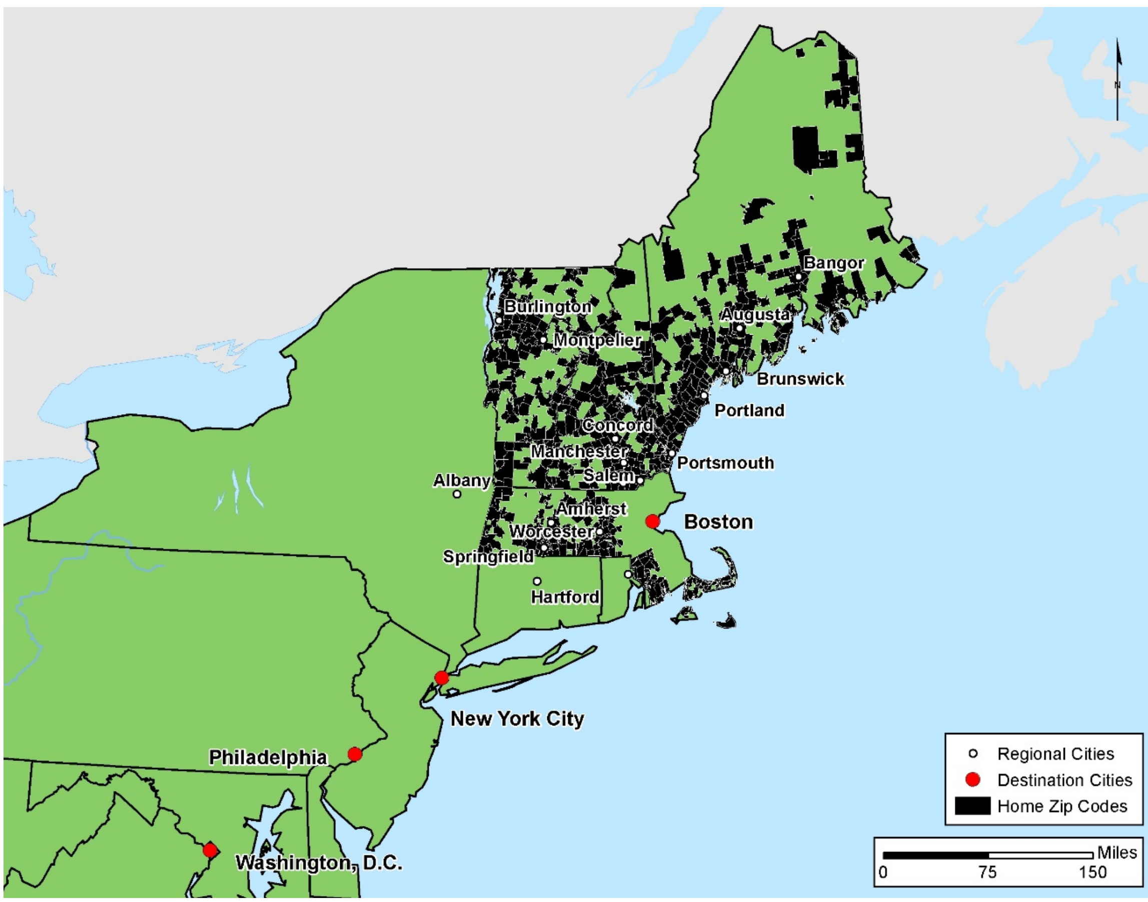


Figure 3-1: Survey Sample: Home (Origin) Zip Codes & Destination Cities

Sources: Zip Codes and Cities from ESRI; Background data from © OpenStreetMap contributors

Table 3-4 shows the number of respondents from each state, with balanced coverage compared with state population (US Census, 2010). Massachusetts had the highest number of respondents, followed by New Hampshire, Maine, and Vermont, respectively.

Table 3-4: Responses by State

State of Residence	Number of Responses
Maine	521
Massachusetts	937
New Hampshire	727
Vermont	375
Total	**2560**

Table 3-5 shows the percentage of respondents from each age group compared with the percentage of the study population (US Census, 2010) from those age groups. Ages 55-64 years had the highest number of respondents. The sample had limited coverage for ages less than 25 years and for ages 75 years and older.

Table 3-5: Distribution of Age Groups

Age (years)	Percentage of Sample	Percentage of Population
18-24	3%	12%
25-34	14%	14%
35-44	15%	17%
45-54	22%	21%
55-64	27%	17%
65-74	17%	10%
75-84	3%	6%
85 or older	0%	3%
Total	**100%**	**100%**

The distribution of responses by education level are compared between the sample and the study population (US Census, 2010) in Table 3-6. Two differences in the sample stand out. The sample had good coverage of more highly educated individuals. The least educated were not well covered.

Table 3-6: Distribution of Education Levels

Education Level	Percentage of Sample	Percentage of Population
Less than high school diploma	1%	11%
High school diploma or equivalent	10%	31%
Some college or associate degree	27%	29%
Bachelor's degree or higher	63%	28%
Total	**100%**	**100%**

The distribution of responses by income level are compared between the sample and the study population (US Census, 2010) in Table 3-7. The survey sample has weak coverage of lower income levels, compared with the study population. Overall, there are groups within the study region that are not well represented in the sample. Prior research suggests younger/older, less educated and lower income individuals do less travel (Schimek 1996; Mallett 2001; Brueckner 2003; McDonald 2015). However, it is unclear the modal implications. Therefore, the total amount of travel may be overstated in this study.

Table 3-7: Distribution of Income Levels

<table>
<tr><th>Income Level</th><th>Percentage of Sample</th><th>Income Level</th><th>Percentage of Population</th></tr>
<tr><td>Under $25,000</td><td>6%</td><td>Under $25,000</td><td>22%</td></tr>
<tr><td>$25,000 - $49,999</td><td>16%</td><td>$25,000 - $49,999</td><td>24%</td></tr>
<tr><td>$50,000 - $74,999</td><td>22%</td><td>$50,000 - $74,999</td><td>19%</td></tr>
<tr><td>$75,000 - $99,999</td><td>20%</td><td>$75,000 - $99,999</td><td>14%</td></tr>
<tr><td>$100,000 - $149,999</td><td>23%</td><td>$100,000 - $149,999</td><td>13%</td></tr>
<tr><td>$150,000 - $199,999</td><td>8%</td><td>$150,000 - $199,999</td><td>4%</td></tr>
<tr><td>$200,000 - $249,999</td><td>2%</td><td rowspan="3">$200,000 or more</td><td rowspan="3">3%</td></tr>
<tr><td>$250,000 or more</td><td>2%</td></tr>
<tr><td>Missing values</td><td>1%</td></tr>
<tr><td>Total</td><td>100%</td><td>Total</td><td>100%</td></tr>
</table>

Respondents Traveling with Other People

Table 3-8 through Table 3-12 show respondents who traveled alone and with others, overall and by trip purpose, mode, destination city, and home state. Of the 2560 survey respondents, 1996 respondents traveled to one of Boston, New York City, Philadelphia, or Washington, DC in the previous year. Overall, the majority of these trips were made by automobile. More than five times as many respondents traveled with overall compared with those who traveled alone. Broken down by trip purpose, twice as many respondents traveled alone for non-business trips compared with those who traveled alone for business trips. More than nine out of ten respondents who traveled with others did so for non-business trips compared with business trips. Over half of the respondents who traveled alone chose automobile, while over three-quarters of respondents who traveled with others chose automobile. This is expected, as traveling by automobile can be convenient with groups of people compared with other passenger modes. About two-thirds of respondents who traveled alone went to Boston, compared with more than 8 out of 10 respondents who traveled with others and went to Boston. The breakdown of respondents traveling alone or with others by home state looks similar to the breakdown overall of respondents by home state.

Table 3-8: Respondents Who Traveled Alone or with Others on Most Recent Trip to Boston, New York, Philadelphia, or Washington, DC

Traveled Alone		Traveled With Others		Total Who Traveled	
301	15%	1695	85%	1996	100%

Table 3-9: Respondents Who Traveled Alone or with Others on Most Recent Trip by Trip Purpose

Purpose	Traveled Alone		Traveled With Others		Total	
Business	124	34%	146	7%	270	11%
Non-Business	239	66%	1857	93%	2096	89%
Total	**363**	**100%**	**2003**	**100%**	**2366**	**100%**

Table 3-10: Respondents Who Traveled Alone or with Others on Most Recent Trip by Mode

Mode	Traveled Alone		Traveled With Others		Total	
Personal Auto	191	58%	1422	77%	1613	74%
Rental Car	9	3%	49	3%	58	3%
Intercity Bus	34	10%	97	5%	131	6%
Intercity Train	36	11%	156	8%	192	9%
Airplane	49	15%	66	4%	115	5%
Other	13	4%	61	3%	74	3%
Total	**332**	**100%**	**1851**	**100%**	**2183**	**100%**

Table 3-11: Respondents Who Traveled Alone or with Others on Most Recent Trip by Destination City

City	Traveled Alone		Traveled With Others		Total	
Boston	197	65%	1370	81%	1567	79%
New York City	56	19%	202	12%	258	13%
Philadelphia	10	3%	36	2%	46	2%
Washington, DC	38	13%	87	5%	125	6%
Total	**301**	**100%**	**1695**	**100%**	**1996**	**100%**

Table 3-12: Respondents Who Traveled Alone or with Others on Most Recent Trip by Home State

State	Traveled Alone		Traveled With Others		Total	
Maine	64	21%	315	19%	379	19%
New Hampshire	84	28%	511	30%	595	30%
Vermont	51	17%	179	11%	230	12%
Massachusetts	102	34%	690	41%	792	40%
Total	**301**	**100%**	**1695**	**100%**	**1996**	**100%**

Figure 3-2 through Figure 3-4 show distributions of respondents traveling with other adults, with other children, and with other people in general. Most respondents traveled with other people, generally with one or two adults, and without children.

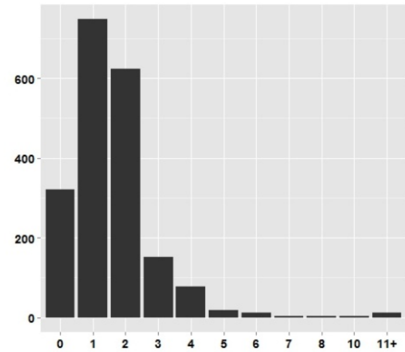


Figure 3-2: Respondents Traveling with Other Adults

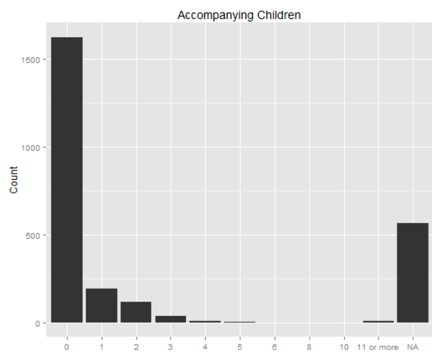


Figure 3-3: Respondents Traveling with Children

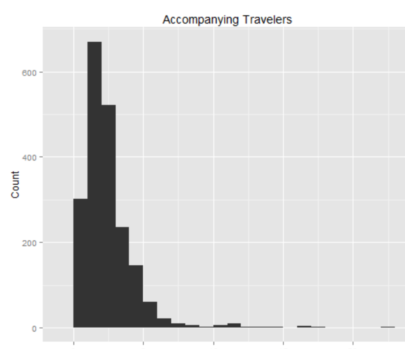


Figure 3-4: Respondents Traveling with Other People

<u>Data Summary</u>

The survey data used in this report included questions and statements about household demographics, personal technology use, recent trips taken to the destination cities of interest, attitudes about traveling in general, and attitudes about traveling by automobile, airplane, intercity bus, and passenger rail. The survey sample was balanced in coverage among states, compared with state population. The sample coverage was limited for ages below 25 years and for ages 75 years and above. The more highly educated were covered by the sample well, but the least educated were not covered well. The sample did not cover lower income levels well. Overall, some groups were not well represented, possibly resulting in travel being overstated for the study. The majority of respondents traveled with other people, generally with one or two adults, and no children. While most respondents chose automobile, this is especially true for those traveling with others.

4. Data Tabulation

This chapter describes data tabulation, including validation, augmentation, preliminary descriptive analysis of the survey response data, distributions of travel modes (Section 4.4) for respondents' most recent trip to one of the destination cities, consideration of which respondents to include in the analysis (Section 4.5), and the determination of assuming a universal mode choice set (Section 4.6). Data augmentation includes adding 25 supplemental sociodemographic, land use, and transportation facility access variables with geographic information systems (GIS) analysis (Section 4.1). Data augmentation also includes performing latent attitudinal factor analysis to reduce 70 attitudinal survey statement responses to a smaller set of 6 latent attitudinal factors for use as explanatory variables in mode choice model estimation (Sections 4.2-4.3). The factors represent attitudes about preference for automobile, dependence on automobile, personal space and safety on public transportation, personal ability to plan for long-distance trips by bus or train, and the opinions of social networks on taking different modes for long-distance intercity travel (automobile, commercial airplane, intercity bus, or intercity train).

The survey data were validated prior to analysis. Data validation included checking for the number of responses for each question, missing values, unique values, assessing the frequency distributions of the dataset, and screening the amount of time taken to complete the survey, for each respondent (Neely et al. 2015).

4.1 Data Augmentation: Geographic Information Systems Analysis

Additional variables were added using available data and GIS analysis, for each home zip code. These included demographic information, land use, distances to destination cities, distances to the nearest urbanized areas within a metropolitan area, and distances to airports, rail stations, and bus stations of different sizes and types.

Table 4-1 displays variables considered for measuring access and accessibility to transportation modes, and the rationale used to consider them. Many of these variables were calculated by Karen Sentoff of the UVM TRC during the analysis for the technical report describing for the survey dataset (Neely et al. 2015).

Table 4-1: Variables Added by GIS Analysis

	Variable Label	Rationale
Access to Transportation Facilities	Network distance from zip code centroid to nearest intercity bus facility	Measure of access to the intercity bus system.*
	Network distance from zip code centroid to nearest transit bus facility	Measure of access to the transit bus system, often connected to the intercity bus system.*
	Network distance from zip code centroid to nearest intercity rail station	Measure of access to the intercity rail system.*
	Network distance from zip code centroid to nearest commuter rail station	Measure of access to the commuter rail system.*
	Network distance from zip code centroid to nearest commercial service airport	Measure of access to the commercial air transportation system.*
	Network distance from zip code centroid to nearest medium hub or greater commercial service airport	Measure of accessibility by air travel; medium hubs or greater provide accessibility to more destinations than smaller airports.*
	Network distance from zip code centroid to nearest large hub commercial service airport	Measure of accessibility by air travel, as large hubs will provide accessibility to the greatest number of destinations.*
	Network distance from zip code centroid to nearest intercity transportation facility	Measure of access to non-automobile intercity modes.
	Ratio of network distance to intercity bus facility, to network distance to destination	Represents the relationship between how far one must travel to get to the intercity passenger mode, and how far the (known) destination is.
	Ratio of network distance to intercity rail station, to network distance to destination	
	Ratio of network distance to nearest commercial service airport, to network distance to destination	
	Number of intercity bus terminals within 25 network miles	Measure of access, assuming that more facilities within the buffer increases access to this mode. For bus and train, multiple facilities may only access the same route, but travelers may choose to pair auto with these modes, for varying portions of trip length.
	Number of intercity train stations within 60 network miles	
	Number of commercial service airports within 75 network miles	
	Dummy variable indicating intercity bus facility located within home zip code	Indicator of high proximity to base level of access to these passenger modes.
	Dummy variable indicating intercity rail facility located within home zip code	
	Dummy variable indicating commercial service airport located within home zip code	
	Dummy variable indicating any intercity transportation passenger mode facility located within home zip code	Indicator of high proximity to base level of access to some intercity passenger mode.
	Network distance to closest intermodal transportation facility	Measure of access to interconnected passenger modes.

[* variable prepared by Karen Sentoff of the UVM TRC (Neely et al. 2015)]
Data sources: Esri, US Census (2010) Data, US Department of Transportation: Bureau of Transportation Statistics

Table 4-1 (Continued): Variables Added by GIS Analysis

	Variable Label	Rationale
Auto Access	Ratio of vehicles to licensed drivers in household	"Car availability index" (Limtanakool et al 2006). Describes availability of personal auto for nested logit models.
Demographics	Zip code location population according to the 2010 census	May serve as a proxy for access to intercity passenger travel modes.*
Land Use	Rural-urban commuting area (RUCA) code from USDA for 2010 census	
	Network distance from zip code centroid to centroid of urban area within metropolitan area	
	Network distance from zip code centroid to centroid of urban cluster within micropolitan area	
	Number of urban cluster areas within home zip code area (2010 census)	
	Urban/rural designation	May serve as a proxy for access to intercity passenger travel modes. Urban designation may increase access to passenger travel modes.

[* variable prepared by Karen Sentoff of the UVM TRC (Neely et al. 2015)]

Data sources: Esri, US Census (2010) Data, US Department of Transportation: Bureau of Transportation Statistics, US Department of Agriculture (USDA): Economic Research Service

4.2 Data Augmentation: Latent Factor Analysis for Attitudinal Variables

Factor analysis was used to prepare a set of six final latent factors related to attitudes towards transportation (Table 4-2), based on the responses to 70 attitudinal statements from the survey (Table 3-2 and Table 3-3). The factors represent attitudes about preference for automobile, dependence on automobile, personal space and safety on public transportation, personal ability to plan for long-distance trips by bus or train, and the opinions of social networks on taking different modes for long-distance intercity travel (automobile, commercial airplane, intercity bus, or intercity train). This sub-section describes the factor analysis modeling framework, the steps used in the factor analysis procedure, the resulting factors, and their interpretation.

Table 4-2: Summary of Factors

	Factor Name	No. of Variables	SS Loadings	Proportion Variance	Cumulative Variance
1	Preference for Auto	10	4.30	0.18	0.18
2	Comfort with Personal Space and Safety on Bus or Train	5	2.61	0.11	0.29
3	Social Networks: No Bus or Train	2	1.37	0.06	0.34
4	Logistics & Effort: Bus or Train	3	1.28	0.05	0.40
5	Social Networks: No Car or Plane	2	1.21	0.05	0.45
6	Auto Dependence	2	0.86	0.04	0.48

Factor analysis identifies latent variables that cannot be measured directly, but act as constructs that produce composite variables that can be measured, where the factors are random variables that combine linearly to denote a larger number of the measurable variables (Rencher and Christensen 2012). This report used factor analysis to identify latent attitudinal variables by using responses to attitudinal survey statements.

Maximum-likelihood factor analysis was used because it is sufficiently robust to perform well, even when the data are not normally distributed, as is the case of the set of attitudinal variables for this study (Fuller and Hemmerle 1966, Nwabueze et al. 2009). The factor model represents the set of responses to 70 attitudinal statements from the survey, referred to as response variables (x_1, x_2,…,x_n), as linear combinations of a smaller number of common factors and a unique factor for each response variable, expressed by equation 1.1 (Afifi and Clark 1996; Rencher and Christensen 2012).

$$x_P - \mu_P = l_{P1}F_1 + l_{P2}F_2 + \cdots + l_{Pm}F_m + e_P \qquad (1.1)$$

$$For\ P = 1 \dots n$$

The model assumes that $\mu_1,\dots,\mu_p$ represent the mean for each response variable; *m* represents the number of common factors; $F_1,\dots,F_m$ represent the common factors; l_{ij} represent the coefficients of the common factors, known as factor loadings (e.g., l_{ij} is the loading of the *i*th response variable on the *j*th common factor); $e_1,\dots,e_p$ represent unique factors for each of the original response variables (Afifi and Clark 1996).

Nine steps were used in the factor analysis procedure, and are outlined in this section (Kim and Mueller 1978). These steps include: 1) exploring the attitudinal statements with correlation analysis; 2) selecting attitudinal statements associated with mode choice based on univariable analysis; 3) fitting exploratory factor models with an iterative process that involves 4) identifying and removing attitudinal variables not contributing substantially to the factors; 5) evaluating factor models by comparing multiple criteria to determine the number of factors to include, and selecting the model with the most consistent number of factors indicated among multiple criteria; 6) transforming the resulting factors to simplify understanding and interpretation using factor rotation; 7) computing factor scores; 8) interpreting the factor scores to determine what is indicated by high and low scores from each factor; and 9) naming factors based on what underlying constructs were identified.

Step 1: Prepare correlation matrix

Pairwise Spearman rank correlation coefficients, which do not assume normality, were computed pairwise among the 70 attitudinal statements. These were computed in order to explore the data. There is low correlation overall; only 41 pairs out of the 2415 possible statement pairs have a correlation coefficient greater than 0.5.

Step 2: Perform univariable analysis

The 70 attitudinal statements were assessed by univariable analysis, with individual multinomial logit models for mode choice of auto, commercial air, intercity bus, or intercity rail, for the most recent trip. Variables with a p-value of 0.25 or below, from the likelihood ratio test, were selected for inclusion in a factor analysis. This univariable analysis was performed in order to restrict attitudinal variables included in the factor analysis to those associated with mode choice. This resulted in 49 variables selected for inclusion in the next step of the factor analysis.

Step 3: Extract initial factors: exploratory factor model

A series of exploratory factor models were fit, starting with 2 factors, and increasing incrementally by 1 factor, until a non-significant p-value came from the Chi Square test, failing to reject the hypothesis that the number of factors are sufficient. This was the initial criteria used to determine the number of factors to include, for this stage of the process, with additional criteria used in Step 5 (Spicer 2005).

Step 4: Remove variables not contributing substantially (return to step 3)

Factor loadings, representing correlation among the variables and factors, were estimated (Spicer 2005). Attitudinal variables not substantially contributing to the model (i.e., with loadings below the cutoff of 0.40, in absolute value, on all factors) were identified, removed, and the model was refit. This process was iterated, repeating Step 3 and Step 4 until four preliminary factor models with different numbers of attitudinal variables were fit. These are referred to as Fit 1, Fit 2, Fit 3, and Fit 4.

Step 5: Select the preferred non-rotated factor model

The four models fit with Step 3 and Step 4 were evaluated by comparing indicators from multiple criteria used to determine the most reasonable number of factors with substantial significance. The model with the most consistent number of factors indicated among multiple criteria, Fit 3, was selected as the preferred non-rotated model. Figure 4-1 presents these selection criteria for factor inclusion, including the Chi Square test for significance (the initial criteria used in Step 3); the Kaiser test for eigenvalues; the criterion of substantive importance, using 2% minimum variance explained by a factor (100/n percent); and non-graphical solutions to the Cattell's Scree Test for the correlation matrix. Typical graphs are included with scree test eigenvalues, parallel analysis, the scree test optimal coordinate, and the scree test acceleration factor.

Step 6: Rotate factors

The preferred model was rotated to simplify understanding and interpretation of the results. The factors in Fit 3, the preferred non-rotated factor analysis model, were rotated using the Varimax technique, a type of orthogonal rotation that can lead to clearer understanding and interpretation (i.e., simple structure) (Kim and Mueller 1978). This resulted in loading most variables onto a single factor each, for simpler interpretation. Twelve factors were extracted, with a p-value of 0.231, failing to reject the null hypothesis that 12 factors are sufficient. Figure 4-2 presents multiple selection criteria for factor inclusion for the rotated factor model, with 6 factors selected for consideration in the discrete choice model estimation.

Step 7: Compute factor scores

Factor models calculate factor scores, as continuous values, for each individual respondent. Factor scores are the values for each factor and are interpreted in Step 8.

Step 8: Interpret factor scores

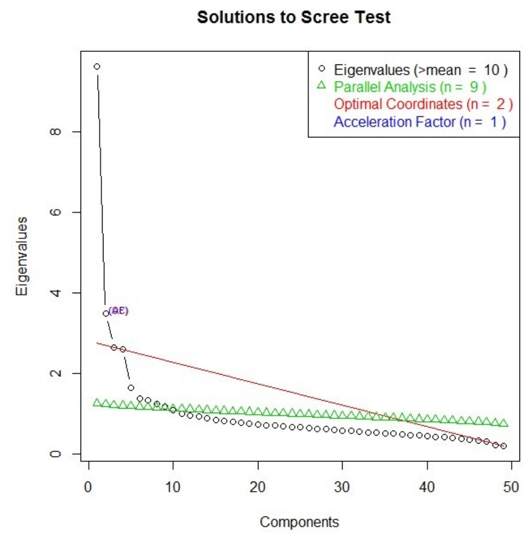
Factor scores were interpreted by examining attitudinal statements loaded onto each factor, and evaluating what responses correspond with high and low scores for the respective factor. This is presented in section 4.3.

Step 9: Name factors

The resulting factors were named after underlying constructs, based on the variables that loaded more heavily on each factor (Rummel 1988). Factors were considered for inclusion in the discrete mode choice models based on how well each factor explains the amount of total variance in the original variables. Table 4-2 shows the resulting factors. The factors are listed in descending order of importance, based on the loadings and proportion of variance from attitudinal survey statements explained by factor analysis. The first factor has the greatest loading of attitudinal statement variables and proportion of variance explained, with lower amounts shown for each successive factor.

Figure 4-1 has multiple parts. The first row in Figure 4-1 shows 25 factors extracted. Twenty-one of the 49 variables were identified as not contributing substantially (loadings of at least 0.40 for substantial factors) to the factor analysis. The model was refit using the 28 remaining variables. The second row in Figure 4-1 shows 14 factors were extracted. Four variables were identified as not contributing substantially. When the factor model was refit using the 24 remaining variables, 12 factors were extracted (third row in Figure 4-1). Four variables were identified as not contributing substantially. The factor model was refit using the 20 remaining variables. The fourth row in Figure 4-1 shows 9 factors were extracted.

Fit 1	
Factor Inclusion Criteria	**No. of Factors**
Chi Square Test	25
Kaiser Test	10
Criterion of Substantive Importance	10
Parallel Analysis	9
Optimal Coordinate	2
Acceleration Factor	1
Variables	**No. of Variables**
In Factor Model	49
With at least one loading ≥ 0.4	28



Fit 2	
Criteria for Factor Inclusion	**Number of Factors**
Chi Square Test	14
Kaiser Test	6
Criterion of Substantive Importance	3
Parallel Analysis	5
Optimal Coordinate	2
Acceleration Factor	1
Variables	**Number of Variables**
In Factor Model	28
With at least one loading ≥ 0.4	24

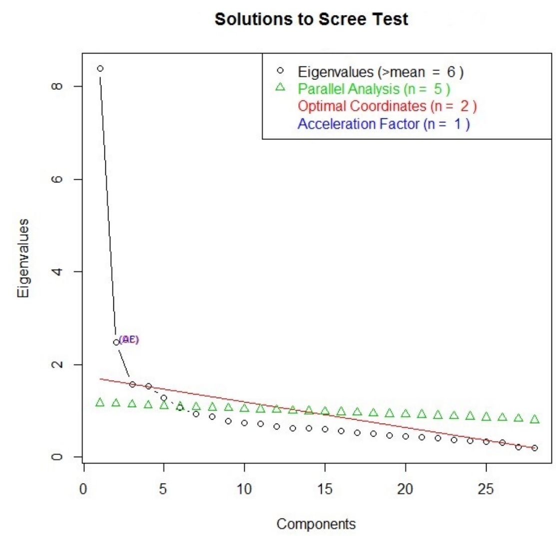
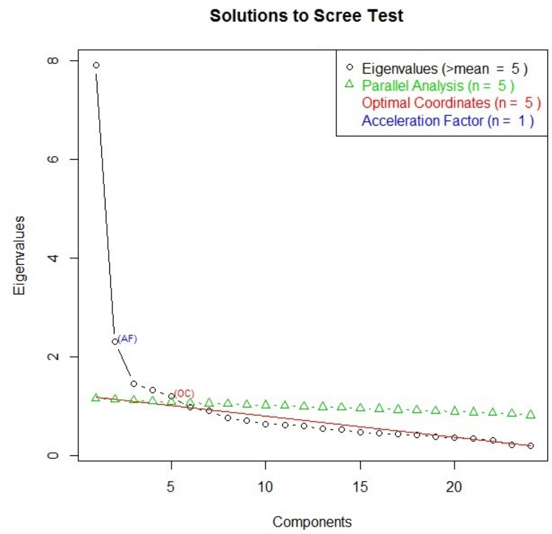


Figure 4-1: Numbers of Factors Supported by Inclusion Criteria

Fit 3	
Criteria for Factor Inclusion	**Number of Factors**
Chi Square Test	12
Kaiser Test	5
Criterion of Substantive Importance	4
Parallel Analysis	5
Optimal Coordinate	5
Acceleration Factor	1
Variables	**Number of Variables**
In Factor Model	24
With at least one loading ≥ 0.4	20



Fit 4	
Criteria for Factor Inclusion	**Number of Factors**
Chi Square Test	9
Kaiser Test	5
Criterion of Substantive Importance	2
Parallel Analysis	4
Optimal Coordinate	4
Acceleration Factor	1
Variables	**Number of Variables**
In Factor Model	20
With at least one loading ≥ 0.4	20

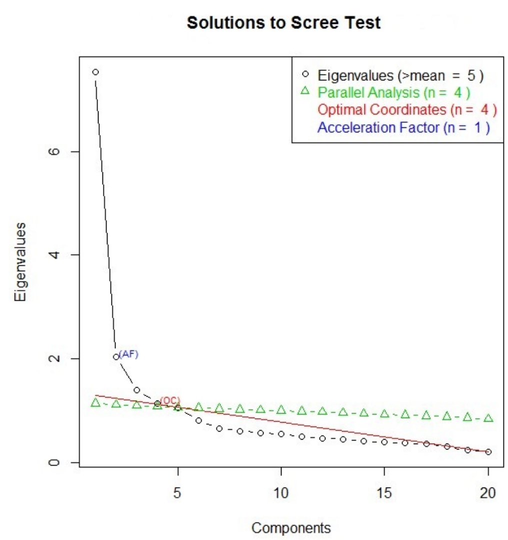


Figure 4-1 (Continued): Numbers of Factors Supported by Inclusion Criteria

Table 4-3 summarizes results for the four non-rotated factor analysis models described above. Fit 3 shows the most consistent number of factors among the different criteria. For this reason, Fit 3 is selected as the preferred non-rotated factor model.

Table 4-3: Number of Factors Supported by Multiple Inclusion Criteria Summary

Criteria for Factor Inclusion	Number of Factors			
	Fit 1	Fit 2	Fit 3	Fit 4
Chi Square Test	25	14	12	9
Kaiser Test	10	6	5	5
Criterion of Substantive Importance	10	3	4	2
Parallel Analysis	9	5	5	4
Optimal Coordinate	2	2	5	4
Acceleration Factor	1	1	1	1

Criteria for Factor Inclusion	**Number of Factors**
Chi Square Test	12
Kaiser Test	5
Criterion of Substantive Importance	6
Parallel Analysis	5
Optimal Coordinate	5
Acceleration Factor	1

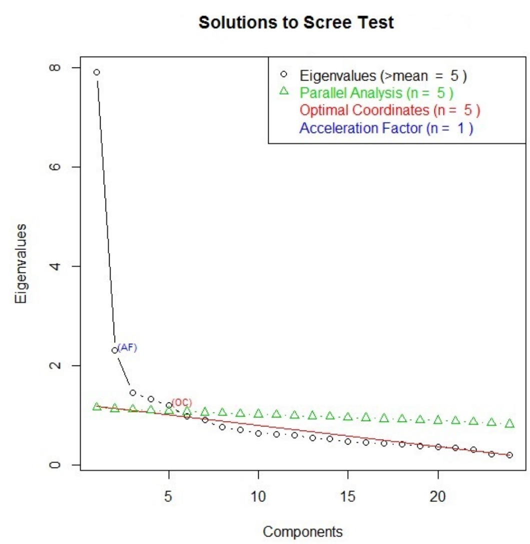


Figure 4-2: Number of Factors Supported by Inclusion Criteria

Recall Table 4-2 shows the resulting summary of factors. Interpreting and naming factors from the rotated factor analysis model was more straight-forward than the model before rotation. This is illustrated by comparing the factor loadings, with a cutoff value of 0.3, before and after Varimax rotation, displayed in Table 4-4 and Table 4-5, respectively. The attitudinal statements are shown in descending order based on loading. For factors before rotation, there are 14 attitudinal statements that load onto multiple factors, with three statements loading onto three factors, and one statement loading onto four factors. For factors after Varimax rotation, only two attitudinal statements load onto multiple factors, two each. The rotation resulted in close to what's referred to as 'simple structure'; the attitudinal statements load onto factors in a way that is readily interpretable, in clearly delineated simple groupings, ideally with limited multiple loadings (Kim and Mueller, 1978).

Table 4-5: Factor Loadings After Varimax Rotation (Cutoff Value 0.3)

Attitudinal Statements	Factors					
	1	2	3	4	5	6
I think that the most RATIONAL choice would be to take a bus or train instead of a car.	X					
I would trust the person who invited me to NYC to recommend how I should travel.	X					
I think that the most PLEASURABLE choice would be to take a bus or train instead of a car.	X					
For me to take a bus or train for such a trip to NYC the next month would be impossible.	X					
I think that the most STRESSFUL choice would be to take a bus or train instead of a car.	X					
I don’t like the idea of riding with a lot of people that I don’t know on a bus or train.		X				
For me, the whole idea of being on a bus or train with other people I do not know seems uncomfortable.		X				
The experience at the NYC bus or train station would be so unpleasant that I would try to avoid it.		X				
It might be unsafe to make this trip by bus or train.		X				
I am the kind of person who would take my own car to NYC.	X					
I could deal with the limited schedules offered by a bus or train for this trip from my home to NYC.	X					
All other things being equal, if a bus was cheaper, but less reliable than a train, I would choose to take a bus.	X	X				
When I drive long distances (like from my home area to NYC), I can get tired and stressed.	X					
When my friends go to NYC, they always take a bus or train.			X			
When my family members go to NYC, they always take a bus or train.			X			
It would be easy for me to get the schedules for a bus or train between here and NYC, and I would understand				X		
I am confident that if I wanted to, I could take a bus or train for such a trip to NYC next month.	X			X		
For me to take a bus or train for such a trip to NYC the next month would be impossible.				X		
My family would think that I should take this kind of trip by car or plane.					X	
My colleagues would likely think that it is strange not to go by a car or plane to NYC.					X	
I feel I am less dependent on cars than my parents are/were.						X

4.3 Data Augmentation: Interpreting Factor Scores and Naming Factors

Factor scores were calculated by the factor model process, as continuous values for each survey respondent. To interpret factor scores, the respective attitudinal statement responses were examined for what responses correspond to high scores for the respective factor, and what responses correspond to low scores for each respective factor. Before this assessment was made, univariable multinomial logit models were fit for each factor to check for significance in mode choice. Five of the six factors showed significance at the 95 percent confidence level, excluding Factor 4: Logistics & Effort: Bus or Train. Factor 4 was not shown to be significant for any outcome modes. Thus, Factor 4 was excluded from initial consideration for model inclusion.

Table 4-6 displays the attitudinal variables loaded onto Factor 1, with associated values from each variable that correspond to high and low factor scores. Higher factor scores correspond with having the possibility of taking a bus or train, although preferring to take one's own automobile. Lower factor scores correspond to the opposite of this. Thus, Factor 1 is named "Preference for Auto".

Table 4-6: Interpreting Factor 1 Scores

Factor 1 - Preference for Auto				
			High Score: 3.273	Low Score: -2.196
Attitudinal Statements	Loadings	Statement	Likert	Likert
tripstatements_26_1	0.79	I think that the most RATIONAL choice would be to take a bus or train instead of a car.	7 – Completely Disagree	1 – Completely Agree
tripstatements_34_1	0.78	I would trust the person who invited me to NYC to recommend how I should travel.	7 – Completely Disagree	1 – Completely Agree
tripstatements_27_1	0.74	I think that the most PLEASURABLE choice would be to take a bus or train instead of a car.	7 – Completely Disagree	1 – Completely Agree
tripstatements_32_1	0.74	For me to take a bus or train for such a trip to NYC the next month would be impossible.	7 – Completely Disagree	1 – Completely Agree
tripstatements_28_1	0.72	I think that the most STRESSFUL choice would be to take a bus or train instead of a car.	7 – Completely Disagree	1 – Completely Agree
tripstatements_16_1	-0.56	I am the kind of person who would take my own car to NYC.	1 – Completely Agree	Agree Less
tripstatements_4_1	0.46	I could deal with the limited schedules offered by a bus or train for this trip from my home to NYC.	Neutral, mixed	Agree more, mixed
tripstatements_29_1	-0.46	All other things being equal, if a bus was cheaper, but less reliable than a train, I would choose to take a bus.	Agree more	7 – Completely Disagree
tripstatements_1_1	0.41	When I drive long distances (like from my home area to NYC), I can get tired and stressed.	Agree Less	Agree More
tripstatements_31_1	0.35	I would make an effort to choose a bus or train for such a trip to NYC next month.	Agree More	Agree More

Table 4-7 displays the attitudinal variables loaded onto Factor 2, with associated values from each variable that correspond to high and low factor scores. Higher factor scores correspond with feeling more comfortable with one's personal space and safety taking a bus or train, while lower factor scores correspond to the opposite. Thus, Factor 2 is named "Comfort with Personal Space and Safety on Bus or Train".

Table 4-7: Interpreting Factor 2 Scores

Factor 2 - Comfort with Personal Space and Safety on Bus or Train				
			High Score: 2.481	Low Score: -2.956
Attitudinal Statements	Loadings	Statement	Likert	Likert
tripstatements_6_1	0.77	I don't like the idea of riding with a lot of people that I don't know on a bus or train.	7 – Completely Disagree	1 – Completely Agree
travelpreferences_17_1	0.7	For me, the whole idea of being on a bus or train with other people I do not know seems uncomfortable.	7 – Completely Disagree	1 – Completely Agree
tripstatements_23_1	0.62	The experience at the NYC bus or train station would be so unpleasant that I would try to avoid it.	7 – Completely Disagree	1 – Completely Agree
tripstatements_22_1	0.59	It might be unsafe to make this trip by bus or train.	7 – Completely Disagree	1 – Completely Agree
tripstatements_29_1	0.38	All other things being equal, if a bus was cheaper, but less reliable than a train, I would choose to take a bus.	Agree Less	Agree More, mixed

Table 4-8 displays the attitudinal variables loaded onto Factor 3, with associated values from each variable that correspond to high and low factor scores. Higher factor scores correspond with feeling like friends and family would not be likely to take a bus or train to New York City, while lower factor scores correspond to feeling like friends and family would take a bus or a train for this type of trip. Thus, Factor 3 is named "Social Networks: No Bus or Train".

Table 4-8: Interpreting Factor 3 Scores

Factor 3 - Social Networks: No Bus or Train				
			High Score: 2.186	Low Score: -2.473
Attitudinal Statements	Loadings	Statement	Likert	Likert
tripstatements_20_1	0.85	When my friends go to NYC, they always take a bus or train.	7 – Completely Disagree	1 – Completely Agree
tripstatements_21_1	0.63	When my family members go to NYC, they always take a bus or train.	7 – Completely Disagree	1 – Completely Agree

Table 4-9 displays the attitudinal variables loaded onto Factor 4, with associated values from each variable that correspond to high and low factor scores. Higher factor scores correspond with feeling like it would not be logistically feasible to take a bus or train to New York City, while lower factor scores correspond to feeling like it would be logistically feasible to take a bus or train for this type of trip. Thus, Factor 4 is named “Logistics & Effort: Bus or Train”.

Table 4-9: Interpreting Factor 4 Scores

Factor 4: Logistics & Effort: Bus or Train				
			High Score: 3.283	Low Score: -2.084
Attitudinal Statements	Loadings	Statement	Likert	Likert
tripstatements_24_1	0.70	It would be easy for me to get the schedules for a bus or train between here and NYC, and I would understand them.	7 – Completely Disagree	1 – Completely Agree
tripstatements_31_1	0.55	I am confident that if I wanted to, I could take a bus or train for such a trip to NYC next month.	7 – Completely Disagree	1 – Completely Agree
tripstatements_33_1	-0.34	For me to take a bus or train for such a trip to NYC the next month would be impossible.	1 – Completely Agree	7 – Completely Disagree

Table 4-10 displays the attitudinal variables loaded onto Factor 5, with associated values from each variable that correspond to high and low factor scores. Higher factor scores correspond with feeling like family and colleagues would think that one should not take a car or plane to New York City. Conversely, lower factor scores correspond to feeling like family and colleagues would think they should take a car or plane for this type of trip. Thus, Factor 5 is named “Social Network: No Car or Plane”.

Table 4-10: Interpreting Factor 5 Scores

Factor 5: Social Network: No Car or Plane				
			High Score: 2.055	Low Score: -2.375
Attitudinal Statements	Loadings	Statement	Likert	Likert
tripstatements_18_1	0.64	My family would think that I should take this kind of trip by car or plane.	7 – Completely Disagree	1 – Completely Agree
tripstatements_19_1	0.54	My colleagues would likely think that it is strange not to go by a car or plane to NYC.	7 – Completely Disagree	1 – Completely Agree

Table 4-11 displays the attitudinal variables loaded onto Factor 6, with associated values from each variable that correspond to high and low factor scores. Higher factor scores correspond with feeling more dependent on automobiles in general, and personal automobiles in particular. Conversely, lower factor scores correspond to feeling less dependent on automobiles in general, and owning automobiles in particular. Thus, Factor 6 is named "Auto Dependence".

Table 4-11: Interpreting Factor 6 Scores

Factor 6: Auto Dependence				
			High Score: 1.225	Low Score: -2.643
Attitudinal Statements	Loadings	Statement	Likert	Likert
travelpreferences_1_1	0.64	I feel I am less dependent on cars than my parents are/were.	7 – Completely Disagree	1 – Completely Agree
travelpreferences_35_1	0.55	Rather than owning a car, I would prefer to borrow, share, or rent a car just for when I need it.	7 – Completely Disagree	1 – Completely Agree

Factor scores were interpreted by considering what attitudinal statement responses correspond to high scores and low scores for the respective attitudinal statement response variables loading onto each factor. Each factor was named based on what underlying construct could be identified, based on which attitudinal response variables loaded onto the factor, and how the factor scores were interpreted for the factor. Five of the six factors were used for full logit model estimation as described in Chapter 6.

4.4 Distribution of Travel Modes

Revealed preference data from the survey were extracted for most recent trips made overall and by destination city, by purpose, by gender, and by age group, for each mode. Modes here include any automobile (personal automobile, rental/borrowed car, or car service), intercity bus, intercity rail, and airplane.

Table 4-12 shows the distribution of travel modes overall and by destination city. The share for personal auto, for overall trips, is largest for Boston, the closest destination city, and descends in order of distance, for New York City, Philadelphia, and Washington, DC. The share for intercity bus, for overall trips, is largest for New York City by twofold, followed by Boston, Philadelphia, and Washington, DC. The share for intercity rail, for overall trips, using one or multiple modes, is again largest for New York City by twofold, followed by Washington, DC, Boston, and Philadelphia. The share for airplane, for overall trips, using one or multiple modes, is largest for Washington, DC, the farthest destination city, and descends in reverse order of distance, for Philadelphia, New York City, and Boston.

Table 4-12: Distributions of Travel Modes by Destination City

	Any Auto		Intercity Bus		Intercity Rail		Airplane		N
Overall	**1661**	**85%**	**112**	**6%**	**120**	**6%**	**73**	**4%**	**1963**
Boston	1377	89%	77	5%	82	5%	8	1%	1545
New York City	166	66%	26	10%	31	12%	27	11%	250
Philadelphia	34	76%	1	2%	2	4%	8	18%	45
Washington	83	68%	6	5%	5	4%	29	24%	123

Table 4-13 shows aggregate distributions of travel modes for business and non-business trips. There were far more recent trips reported for non-business purposes than for business purposes, by more than sixfold. Auto was chosen for the majority of both business and non-business trips. For business trips, airplane was the second most chosen mode, with almost equal portions of intercity bus and intercity rail chosen. For non-business trips, intercity rail was chosen slightly more than intercity bus, and airplane was chosen least overall, about half as often as intercity bus or rail. It is expected that air travel is more common for business trips, as the expense may be covered and more easily justified, compared with having to pay the expense oneself.

Table 4-13: Distributions of Travel Modes by Purpose

Auto	Airplane	Intercity Bus	Intercity Rail	**Total**
Business				
205	25	17	16	**263**
77.9%	9.5%	6.5%	6.1%	**100.0%**
Non-Business				
1455	47	94	104	**1700**
85.6%	2.8%	5.5%	6.1%	**100.0%**

Table 4-14 shows differences in mode choices, for business and non-business trips, between genders, for traveling from home locations in northern New England to large metropolitan areas in the Northeast. More males made business trips than did females. For business trips, the majority of both females and males chose automobile, in equal proportions. Males chose airplane almost twice as much as females did for this kind of travel. Males also chose bus slightly more often than females. On the other hand, females chose rail more than three times as much as males did for business trips.

More females made non-business trips than did males. For non-business trips, mode share for auto was close, with males choosing auto slightly more often than females. Females and males chose airplane in approximately equal amounts. More females chose intercity bus and intercity rail than males did, with the difference greater for intercity bus.

Table 4-14: Distributions of Travel Modes Between Genders

	Auto	Airplane	Intercity Bus	Intercity Rail	Totals
			Business		
Females	95	8	7	12	**122**
	77.9%	6.6%	5.7%	9.8%	**100.0%**
Males	110	17	10	4	**141**
	78.0%	12.1%	7.1%	2.8%	**100.0%**
			Non-Business		
Females	843	29	65	64	**1001**
	84.2%	2.9%	6.5%	6.4%	**100.0%**
Males	612	18	29	40	**699**
	87.6%	2.6%	4.2%	5.7%	**100.0%**
					1963

Table 4-15 shows differences in mode choices, for business and non-business trips, among age groups, for traveling from home locations in northern New England to large metropolitan areas in the Northeast. For business trips, the percent of each age group choosing auto is between 78% and 83%. Similar percentages of people ages 35-54 years and 65 years or older chose airplane for this type of business trip (about 7%), while people below age 35 years chose airplane more often and people ages 55-64 years chose airplane most often. About 5-7% of each age group, except for ages 65 years or older, chose intercity bus, while 10% of those age 65 years or older chose this mode. Nobody age 65 years or older chose intercity rail for business travel, with ages 18-34 years and 55-64 years choosing intercity rail no more than 5%, and ages 35-54 years choosing intercity rail about twice as often.

For non-business trips, the percentage of all age groups choosing auto is higher overall, than for business trips (81-90%), with the highest percentage choosing auto from ages 35-54 years, and the lowest percentage from ages 65 years or older. The percentage of all age groups choosing airplane for non-business trips is between 2% and 3%, except for ages 65 years or older, who chose airplane for 6% of these trips. All age groups, except for ages 35-54 years, chose intercity bus for 6-7% of non-business trips, while those ages 35-54 years chose intercity bus for less than 4% of non-business trips. Ages 35-54 years are some of the prime child-raising years, and intercity bus may be the least conducive to traveling with groups of children. Those ages 35-54 years also chose intercity rail less than other age groups, although the difference is smaller.

Table 4-15: Distributions of Travel Modes Among Age Groups

	Auto	Airplane	Intercity Bus	Intercity Rail	Total
			Business		
18-34 years	32	4	2	2	**40**
	80.0%	10.0%	5.0%	5.0%	**100.0%**
35-54 years	80	7	6	10	**103**
	77.7%	6.8%	5.8%	9.7%	**100.0%**
55-64 years	68	12	6	4	**90**
	75.6%	13.3%	6.7%	4.4%	**100.0%**
65+ years	25	2	3	0	**30**
	83.3%	6.7%	10.0%	0.0%	**100.0%**
			Non-Business		
18-34 years	264	5	22	22	**313**
	84.4%	1.6%	7.0%	7.0%	**100.0%**
35-54 years	563	10	24	31	**628**
	89.7%	1.6%	3.8%	4.9%	**100.0%**
55-64 years	372	13	29	28	**442**
	84.2%	2.9%	6.6%	6.3%	**100.0%**
65+ years	256	19	19	23	**317**
	80.8%	6.0%	6.0%	7.3%	**100.0%**

Figure 4-3 through Figure 4-8 show distributions of travel modes by attitudes, based on factor scores from latent attitudinal factor models. Figure 4-3 shows a box and whisker plot of Factor 1 scores and mode choice. In box and whisker plots, the box includes the inter-quartile range of data points, made up of the second and third quartiles, the dark line indicates the median, and the ends of the whiskers indicate the minimum of the first quartile and the maximum of the fourth quartile, respectively. The circles outside of the whiskers indicate outliers in the data, which are either smaller than the first quartile or larger than the third quartile, by at least 1.5 times the interquartile range in either case. Figure 4-3 does not indicate any statistical difference in factor scores for *Preference for Auto* by mode.

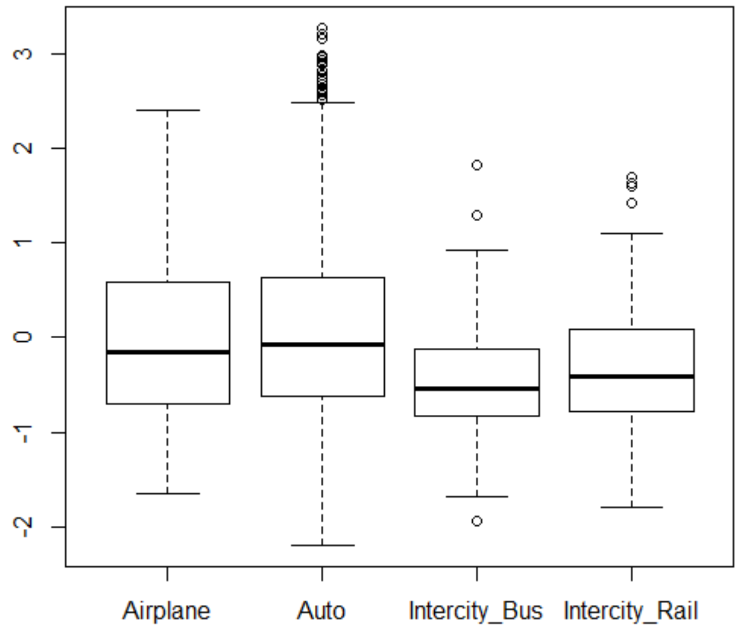


Figure 4-3: Mode Choice Vs. Factor 1 – Preference for Auto

Figure 4-4 shows a box plot of Factor 2 scores and mode choice. It does not indicate any statistical difference in factor scores for *Comfort with Personal Space and Safety on Bus or Train* by mode.

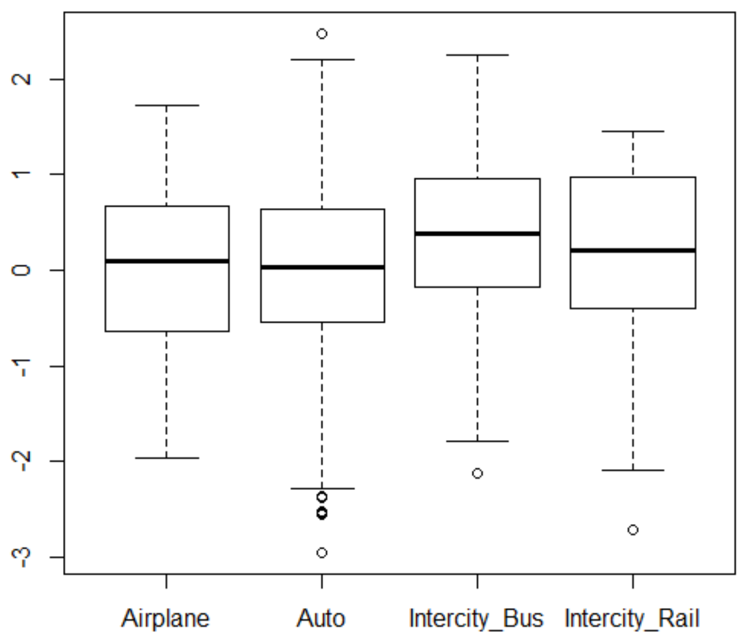


Figure 4-4: Mode Choice Vs. Factor 2 - Comfort with Personal Space and Safety on Bus or Train

Figure 4-5 shows a box plot of Factor 3 scores and mode choice. It does not indicate any statistical difference in factor scores for *Social Network: No Bus or Train* by mode.

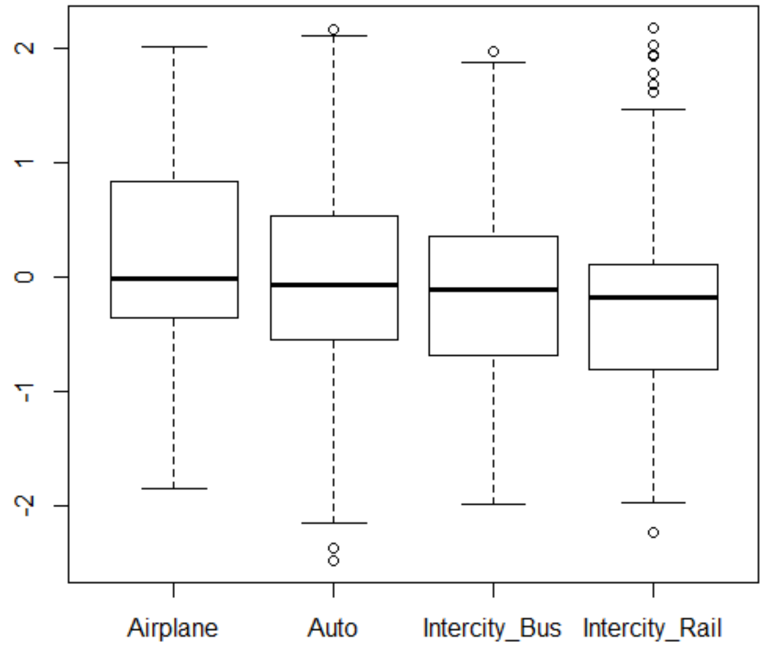


Figure 4-5: Mode Choice Vs. Factor 3 – Social Network: No Bus or Train

Figure 4-6 shows a box plot of Factor 4 scores and mode choice. It does not indicate any statistical difference in factor scores for *Logistics & Effort: Bus or Train* by mode.

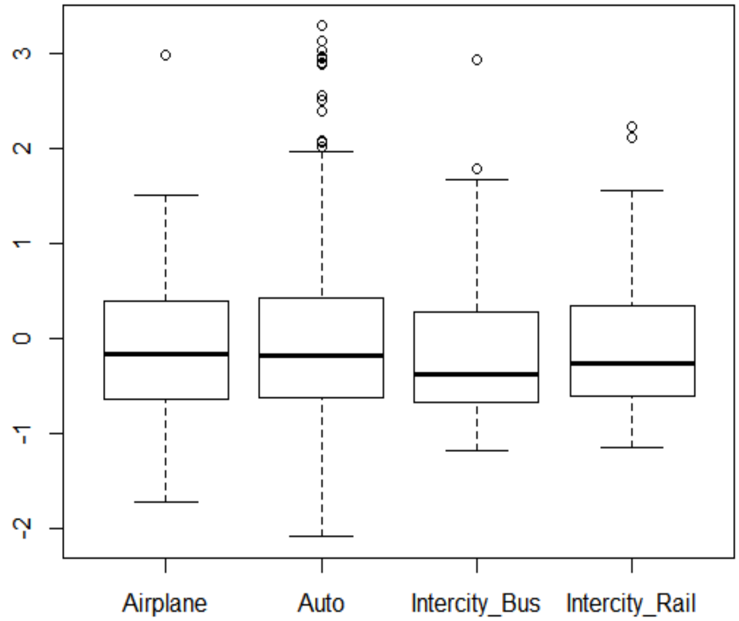


Figure 4-6: Mode Choice Vs. Factor 4 - Logistics & Effort: Bus or Train

Figure 4-7 shows a box plot of Factor 5 scores and mode choice. It does not indicate any statistical difference in factor scores for *Social Network: No Car or Plane* by mode.

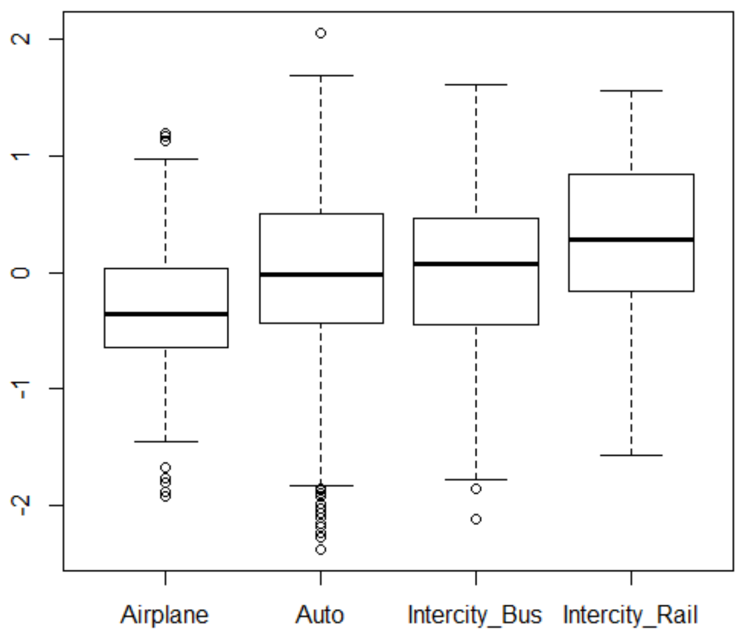


Figure 4-7: Mode Choice Vs. Factor 5 - Social Network: No Car or Plane

Figure 4-8 shows a box plot of Factor 6 scores and mode choice. It does not indicate any statistical difference in factor scores for *Auto Dependence* by mode.

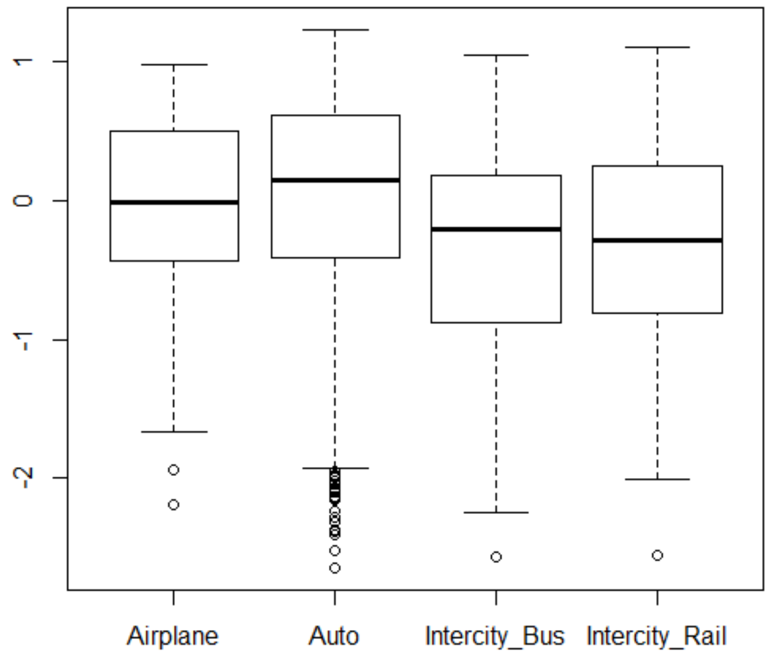


Figure 4-8: Mode Choice Vs. Factor 6 - Auto Dependence

Assessing latent factors one at a time with box and whisker plots had limited findings and does not have the same capability of models that control for multiple variables.

4.5 Respondents to Include in Analysis

Of the 2560 survey respondents, 564 did not visit any of the destination cities in the previous year. Removing these from the analysis, there are 1996 remaining respondents who visited at least one of the destination cities in the previous year.

Respondents Without a License

The vast majority of respondents (98.9%) who visited at least one of the destination cities in the previous year have a driver's license. There were 23 respondents (1.1%) making these trips who do not have a license. Due to the small number, these respondents were left in the analysis. Table 4-16 shows the age distribution of these respondents without a license. Over half of these respondents without a license are below the age of 35 years.

Table 4-16: Respondents (who visited at least one destination city) Without a Driver's License by Age

Age (years)	Count
18-24	3
25-34	11
35-44	2
45-54	4
55-64	1
65-74	2
75-84	0
85+	0

Table 4-17 shows the distribution of respondents by state of residence for those respondents without a license. The distribution follows a similar shape as the distribution of total number of respondents from each state.

Table 4-17: Respondents Without a Driver's License by State of Residence

Residence	Count
Massachusetts	12
New Hampshire	5
Maine	3
Vermont	3

Final sample size

Table 4-18 shows the selection of respondents that were included in this analysis. Of the 1996 respondents who visited at least one of the destination cities in the previous year, 1963 respondents chose at least one of the study outcome modes of interest (automobile, commercial airplane, intercity bus, intercity rail) for their most recent trip. These 1963 respondents comprise the final sample used.

Table 4-18: Respondents to Include in Analysis

Total survey respondents	2560
Respondents who visited ≥ 1 city	1996
Respondents who chose ≥ 1 study mode (auto, airplane, intercity bus, intercity rail)	1963

Figure 4-9 displays the zip code polygons of the final sample of 1963 respondents who visited at least one of the destination cities in the previous year, and chose at least one of the modes included in the study. It shows a map symbolizing the number of respondents originating from each zip code, using a shading gradient. The zip codes with higher numbers of respondents are aligned with some of the more populated cities, as well as the commuter shed outside of the Boston metropolitan area. There appears to be lower numbers of respondents within the individual rural zip codes, but a larger number of zip codes with low numbers of respondents, reflecting the general geographic distribution of the population.

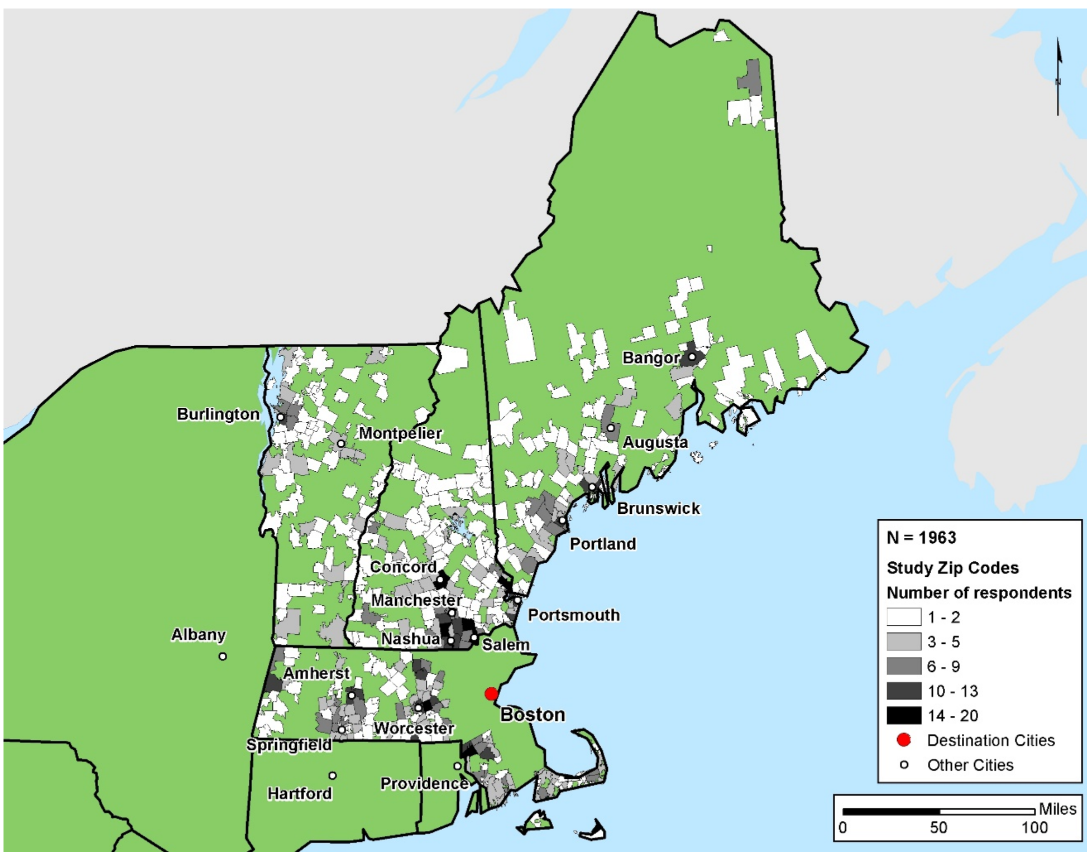


Figure 4-9: Study Sample: Number of Respondents from Home (Origin) Zip Codes
Sources: Zip Codes and Cities from ESRI; Background data from © OpenStreetMap contributors

Figure 4-10 displays the frequency distribution of respondents by destination city. The figure shows that the majority of the most recent trips made by the respondents went to Boston, followed by New York City, third being Washington, DC, and Philadelphia having the least. Boston may attract the largest portion of these trips, as it is the closest major metropolitan area for all respondents. These destination cities may have been the final destination for most respondents, while others may have just stopped through for a layover on their way to a final destination, domestic or international. The survey did not capture this distinction.

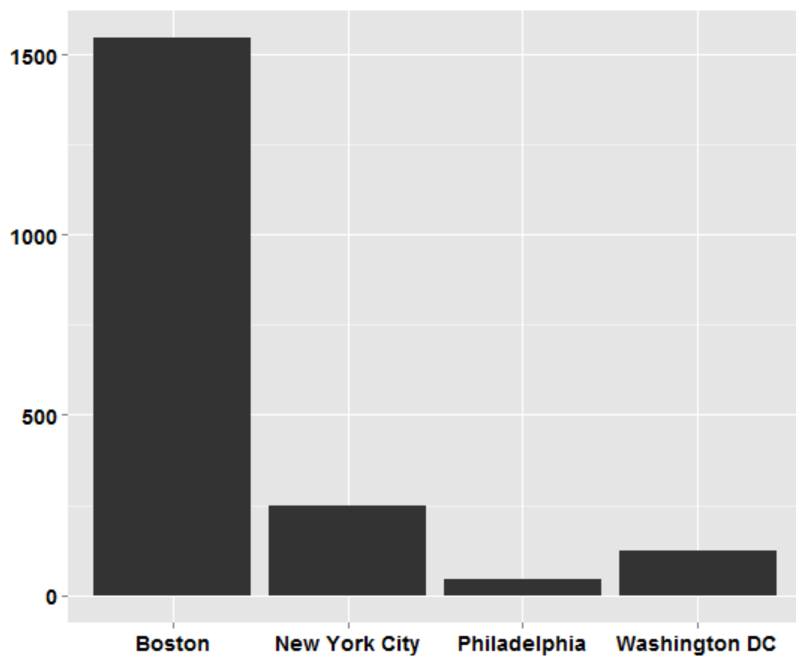


Figure 4-10: Distribution of Respondents by Destination City

4.6 Choice Set Generation

Figure 4-11 through Figure 4-13 display distributions of distances from each respondent's zip code to intercity passenger facilities, overall and for those who chose that specific mode for each destination city, for intercity bus, intercity rail, and commercial air service. Based on the shape of the long tails of these distributions, and the calculated 90th and 95th percentiles, cutoff distances for each type of transportation facility were selected to measure access to that type of facility. These cutoff distances are displayed in Table 4-19. Based on these cutoff distances, a universal choice set, which assumes all respondents had access to each of the four mode outcomes (automobile, commercial airplane, intercity bus, and intercity rail), was used in the analysis, as it captures 90 to 95% of respondents.

Table 4-19: Cutoff Distances to Each Type of Intercity Transportation Facility

Facility Type	Distance
Intercity Bus	25 miles
Intercity Rail	60 miles
Commercial Air Service	75 miles

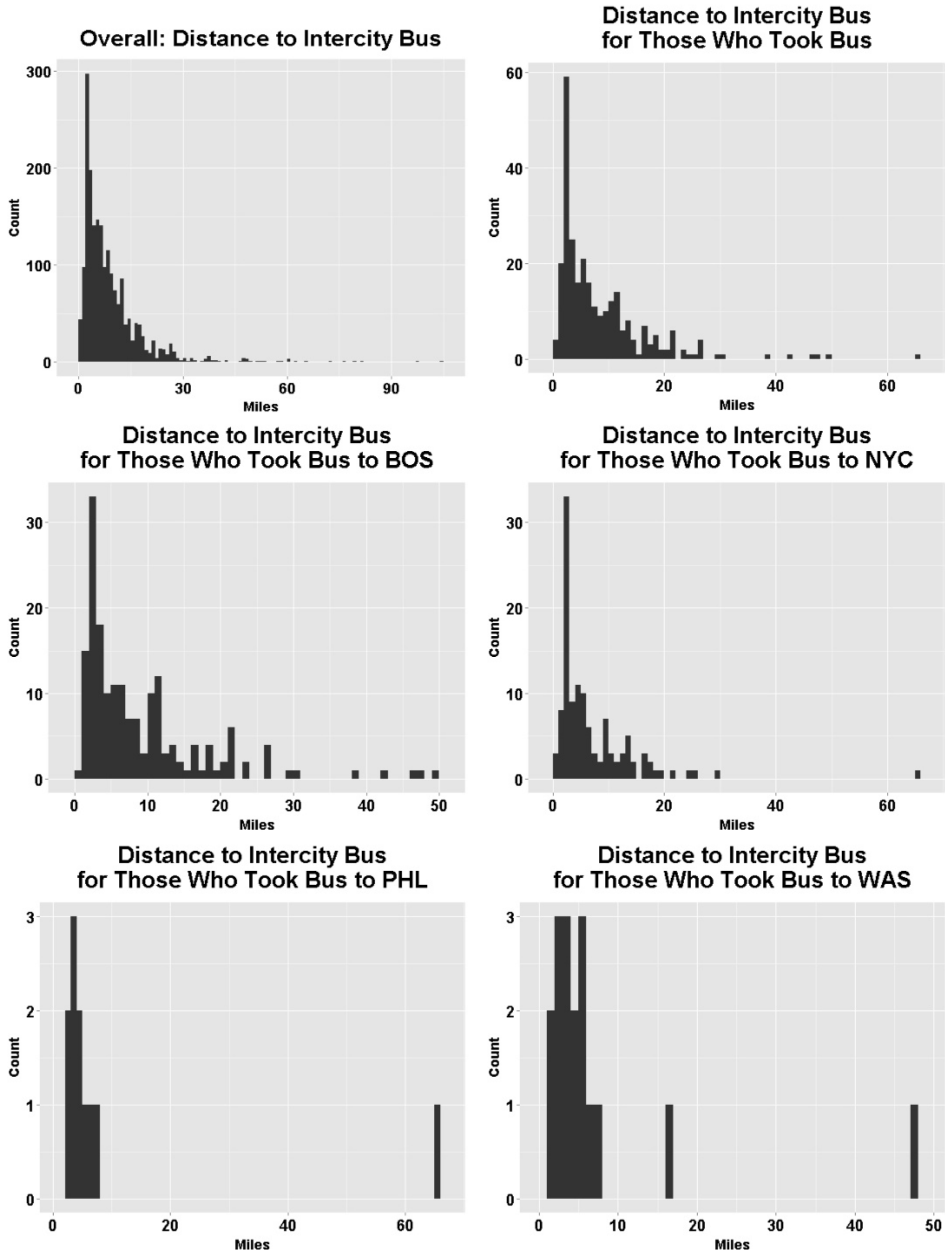


Figure 4-11: Distributions of Distances from Respondents' Home Zip Code Centroid to Closest Intercity Bus Facility

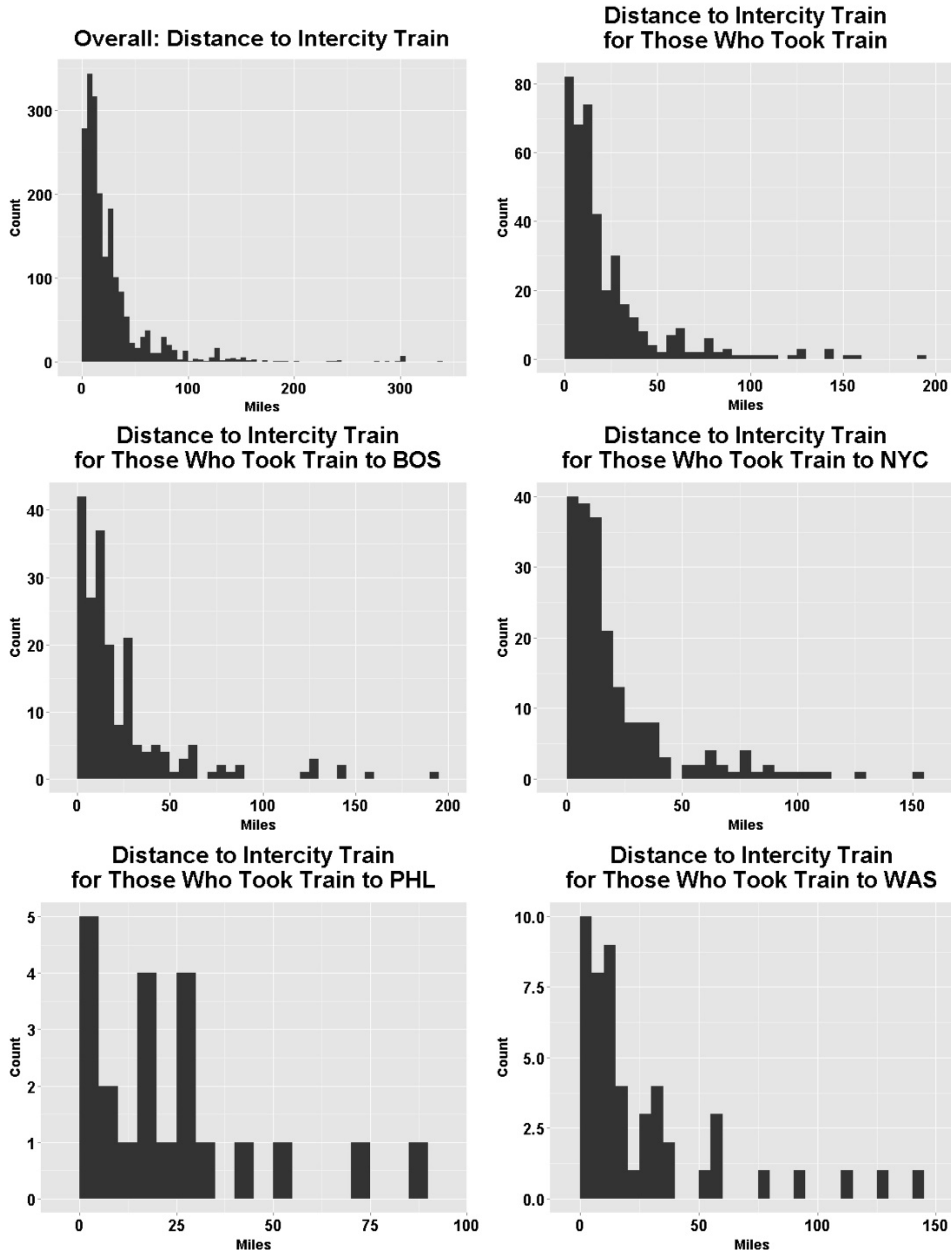


Figure 4-12: Distributions of Distances from Respondents' Home Zip Code Centroid to Closest Intercity Train Facility

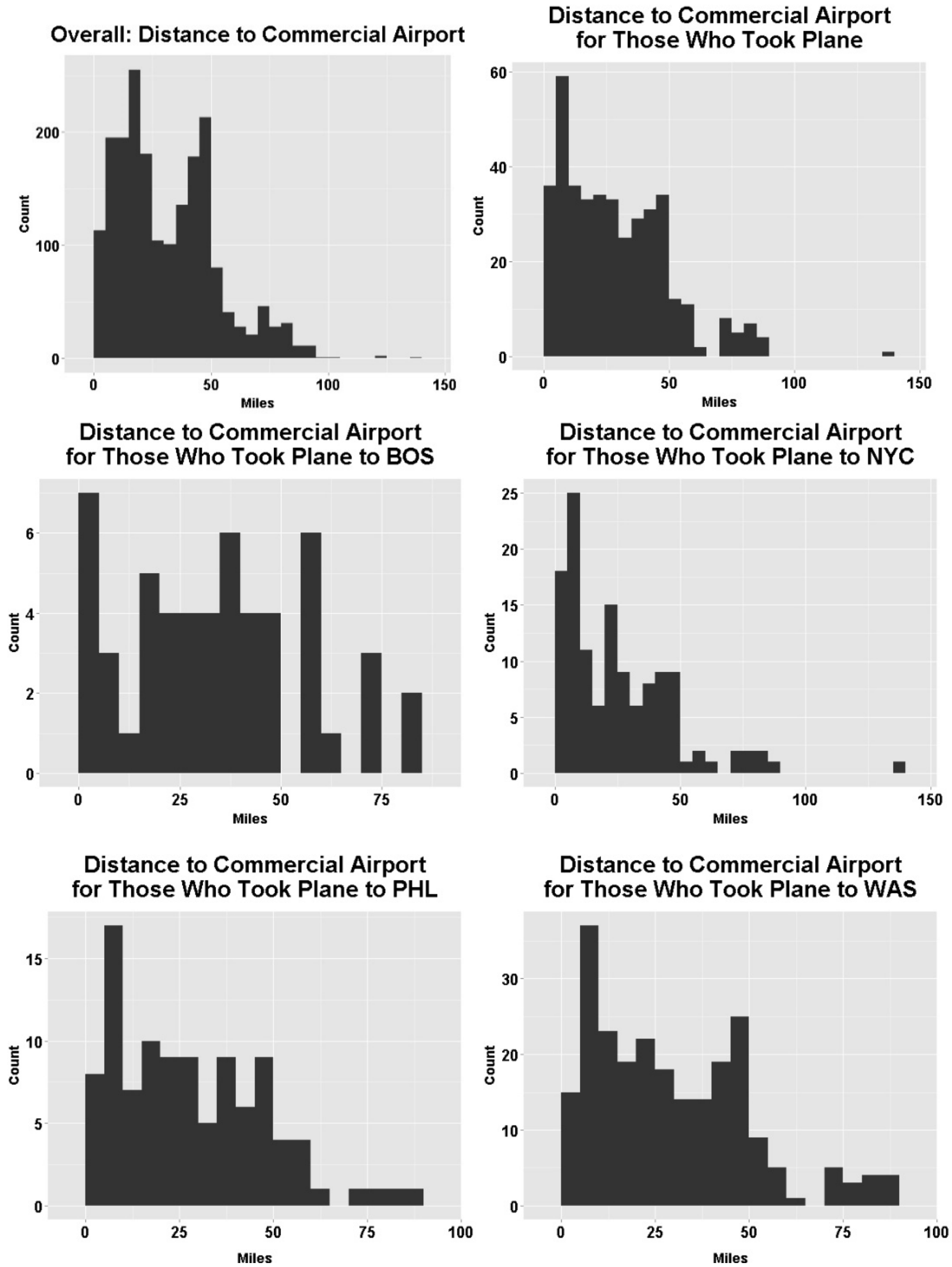


Figure 4-13: Distributions of Distances from Respondents' Home Zip Code Centroid to Closest Primary Commercial Service Airport

Figure 4-14 through Figure 4-17 display distributions of the ratio of distances to different transportation facilities, to the distance to each destination city. This describes how far one has to travel, in order to get to a certain kind of transportation facility, relative to the total distance to the destination city. The distributions smooth out as the distance to the destination city increases past a certain threshold. Boston shows the most prominent fluctuation in this ratio. As the distance to the destination city increases, the distance to the different facilities stays constant, and makes it easier to travel a greater portion of the trip on an alternative mode to automobile. The greatest fluctuation is for airports, while the least is for bus stations, reflecting corresponding distributions of distances to transportation facilities. The greatest fluctuation among cities and modes is for the ratio of the distance to a commercial airport over the distance to Boston. The smoothest distributions are for bus stations and trips to New York City, Philadelphia, and Washington, D. C. These differences in fluctuations reflect components of the decision making framework for people choosing a mode for traveling from their home locations in northern New England to major cities in the Northeast, leading to the discrete choice model approach.

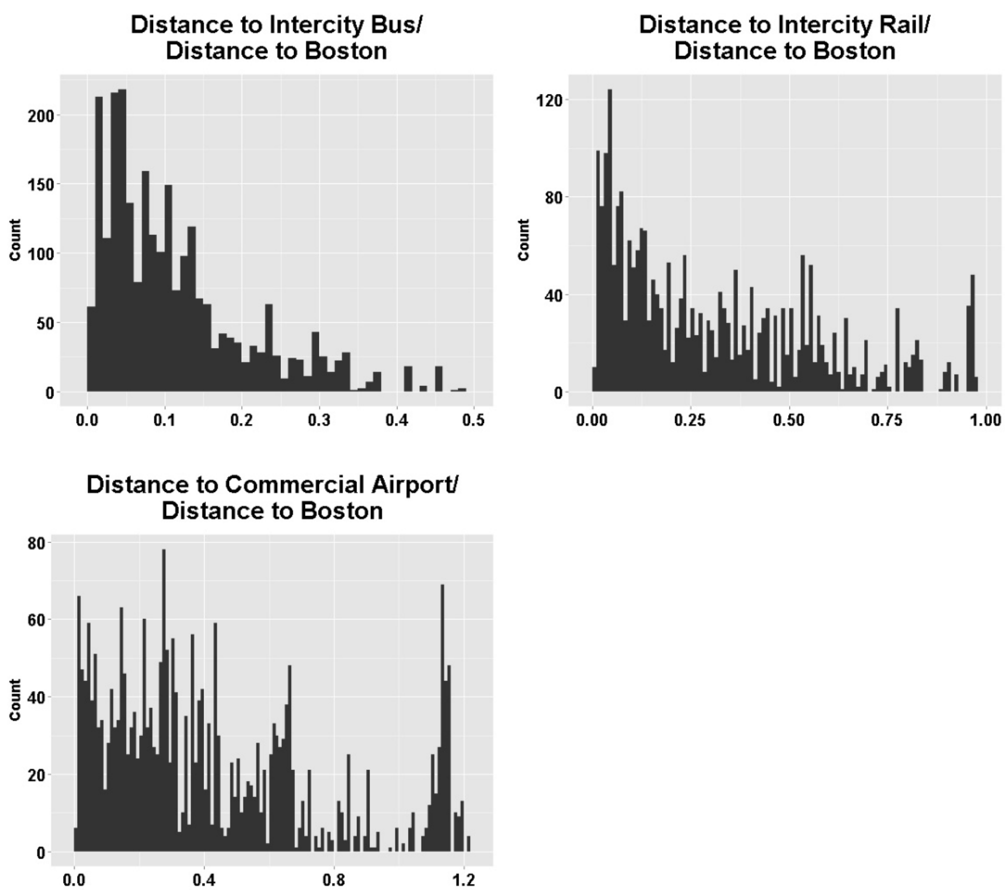


Figure 4-14: Distributions of Ratios of Distances from Respondents' Home Zip Code Centroid to Transportation Facilities to Distance to Boston

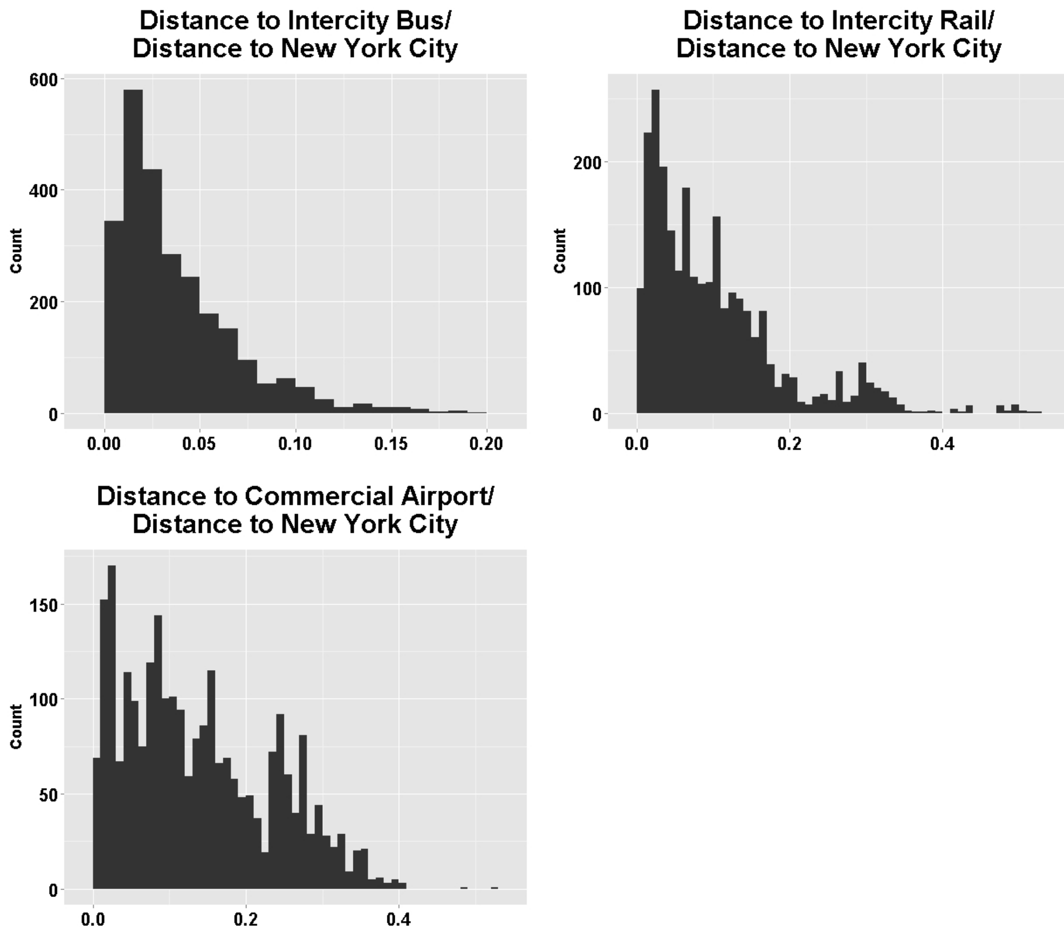


Figure 4-15: Distributions of Ratios of Distances from Respondents' Home Zip Code Centroid to Transportation Facilities to Distance to New York City

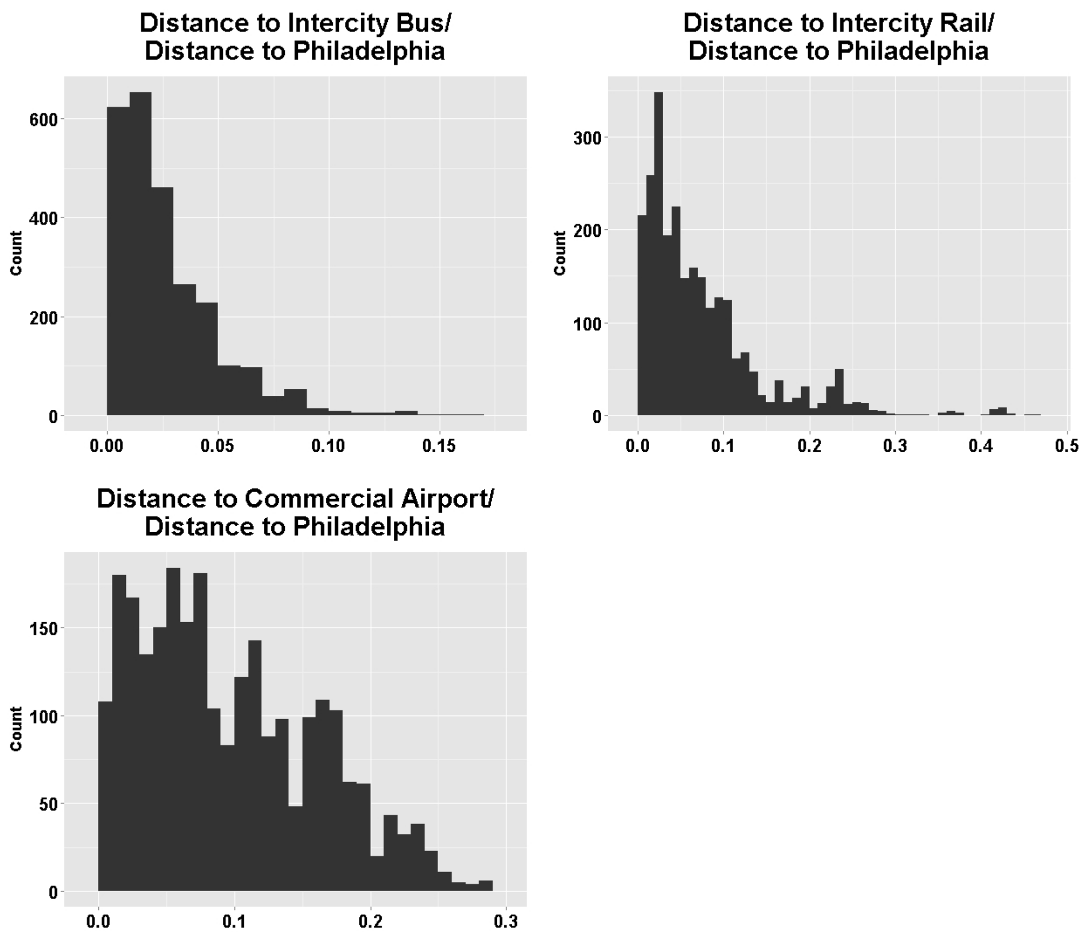


Figure 4-16: Distributions of Ratios of Distances from Respondents' Home Zip Code Centroid to Transportation Facilities to Distance to Philadelphia

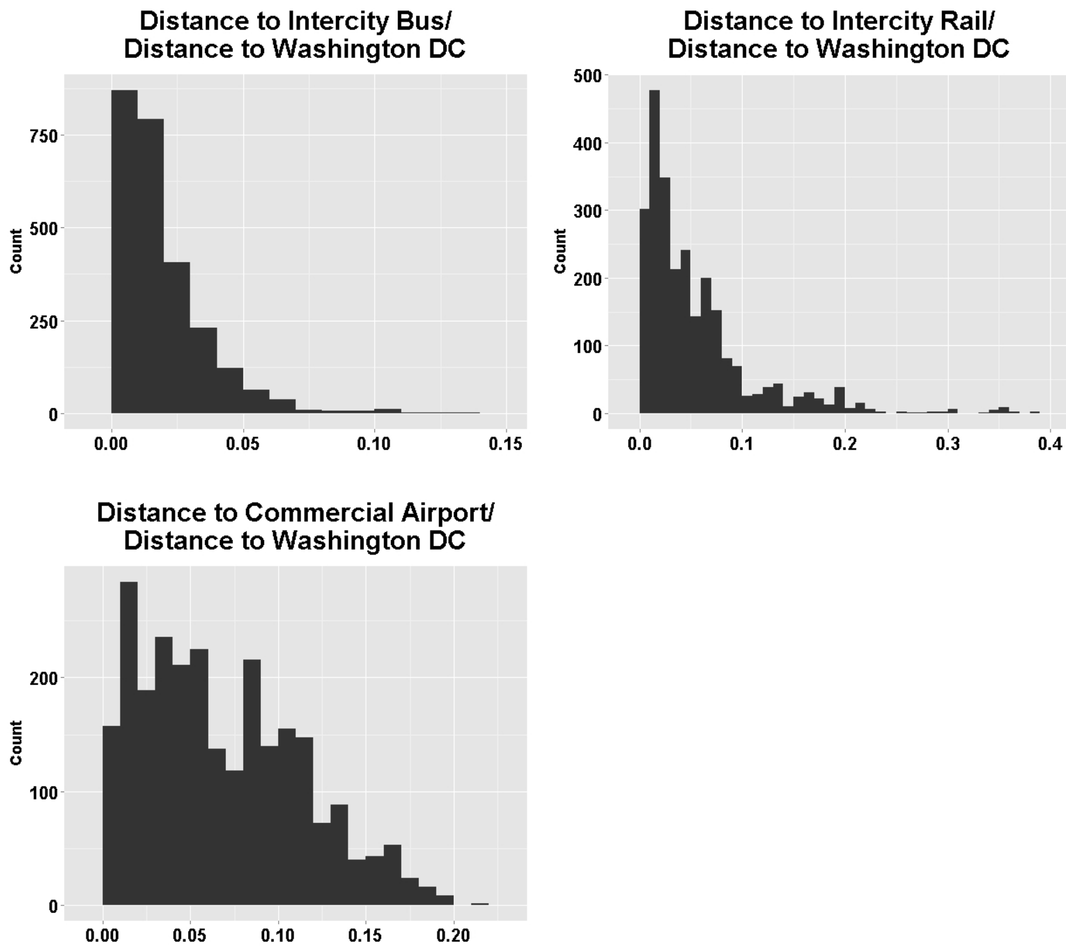


Figure 4-17: Distributions of Ratios of Distances from Respondents' Home Zip Code Centroid to Transportation Facilities to Distance to Washington, DC

5. Modeling Approach

Discrete mode choice modeling methods were used to estimate a set of multinomial and nested logit models to identify factors that influence people's travel mode choice (automobile, commercial airplane, intercity bus, or intercity rail), for traveling from home locations in northern New England (Maine, New Hampshire, Vermont, and Massachusetts outside of the Boston metropolitan area), going to Boston, New York City, Philadelphia, or Washington, DC, for business and non-business trip purposes. Discrete choice models are rooted in a framework of human decision making processes and statistical assumptions. In this context, when faced with a decision, an individual considers available mutually exclusive alternatives (choice set), and chooses one alternative based on a decision rule (Ben-Akiva and Lerman 1985; Koppelman and Bhat 2006). This choice process includes four elements: a decision maker (including the associated attributes of the decision maker and their environment), a choice set, attributes of the alternatives within that choice set, and a decision rule. This chapter describes the logit choice modeling framework used, how it was applied to the Intercity Travel, Information, and Technology Survey data, and estimation of multinomial logit and nested logit models.

5.1 The Logit Choice Modeling Framework

The framework of human decision making processes and statistical assumptions which leads to the discrete choice modeling process is based on random utility maximization theory. The choice set (C_n) of the individual decision maker (n) is the set of alternatives available and known to the decision maker, assuming the alternatives are independent, mutually exclusive, and finite (Ben-Akiva and Lerman 1985). An individual is assumed to choose the alternative (*i*) from their respective choice set (C_n) that has the most utility (i.e. amount of usefulness). The researcher does not know with certainty what these utilities are, so they are represented in choice models as random variables, such that [equation 1.2 (from Ben-Akiva and Lerman 1985)] the probability of choosing alternative *i*, given the choice set C_n, is equal to the probability that its utility (U_{in}) is greater than or equal to the utility of the other available alternatives (U_{jn}) (Ben-Akiva and Lerman 1985; Train 1986; Koppelman and Bhat 2006):

$$P(i|C_n) = P[U_{in} \geq U_{jn}, all\ j \in C_n] \qquad (1.2)$$

An alternative's utility, U_{in}, includes [equation 1.3 (from Train 1986; Koppelman and Bhat 2006)] a component that is measurable and known to the analyst, V_{in}, as well as a component that is unknown and unmeasured, $\varepsilon_{in,}$ represented by a random variable (Ben-Akiva and Lerman 1985; Train 1986; Koppelman and Bhat 2006):

$$\mathrm{U}_{in} = \mathrm{V}_{in} + \varepsilon_{in} \qquad (1.3)$$

V_{in}, the component of utility that is known, is expressed as linear in parameters equation [equation 1.4 (from Train 1986)]:

$$V_{in} = \boldsymbol{\beta} V(\boldsymbol{z}_{in}, \boldsymbol{s}_n) \quad (1.4)$$

The vector-valued function V_{in} of observed attributes of each alternative, $\mathbf{z}_{in}$, and observed characteristics of the decision maker, $\mathbf{s}_n$, is multiplied by a vector containing parameters, $\boldsymbol{\beta}$, estimated using maximum likelihood techniques (Ben-Akiva and Lerman 1985; Train 1986; Koppelman and Bhat 2006).

For logit models, the unobserved (error) component of utility, ε_{in}, is assumed to be identically and independently distributed (IID), following a Gumbel (Type I Extreme Value) distribution [equation 1.5 (from Ben-Akiva and Lerman 1985)]:

$$Gumbel\ Probability\ Density\ Function : f(\varepsilon) = \mu e^{-\mu(\varepsilon-\eta)} e^{-e^{-\mu(\varepsilon-\eta)}} \ (1.5)$$

$$\mu\ is\ the\ scale\ parameter\ for\ variance$$

$$\eta\ is\ the\ parameter\ for\ mode$$

$$Mean = \eta + \frac{0.577}{\mu}$$

$$Variance = \frac{\pi^2}{6\mu^2}$$

In a discrete choice model, if the ε term is assumed to follow a normal distribution, it would lead to the multinomial probit (MNP) choice model (Koppelman and Bhat 2006). Using MNP models for discrete choice, particularly when these choices are unordered, has challenges associated with computation and interpretation (Ben-Akiva and Lerman 1985; Borooah 2002: Koppelman and Bhat 2006). By using an assumption that the unobserved components of utility are identically and independently distributed (IID) among alternatives and individuals following a Gumbel distribution, the result is a closed-form probabilistic multinomial logit model with gains in computation ease, using maximum likelihood estimation, so probability calculations can be made without numerical integration or simulation (Ben-Akiva and Lerman 1985;

Koppelman and Bhat 2006). Multinomial logit models have been found effective and useful for transportation mode choice problems (Ben-Akiva and Lerman 1985).

Multinomial logit models for discrete choice analysis follow the form shown in equation 1.6 (from Ben-Akiva and Lerman 1985), with the probability of each choice outcome being chosen, for each individual ($P_n(i)$), represented by the ratio of the exponentiated observable component of the utility of that alternative, for that individual (V_{in}), divided by the sum of the exponentiated utilities of all choice outcomes available in the choice set for that individual, with the probability being between zero and one for all alternatives in the choice set (C_n), and the sum of probabilities for choosing each alternative in the choice set equal to one. The probability of choosing each alternative is a function of its share of the utility of all alternatives (Koppelman and Bhat 2006). The utility of each choice outcome, for each individual, is broken into two parts (equation 1.6) a vector of estimated parameters (**β´**), and a vector of factors ($\boldsymbol{x}_{jn}$ and $\boldsymbol{x}_{in}$) representative of the attributes of the alternative and of the individual. The parameters of the reference alternative are set equal to zero, such that the exponentiated utility of the reference is equal to one.

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} = \frac{e^{\beta' X_{in}}}{\sum_{j \in C_n} e^{\beta' X_{jn}}} \quad (1.6)$$

$$0 \leq P_n(i) \leq 1, \quad for\ all\ i \in C_n$$

$$\sum_{i \in C_n} P_n(i) = 1$$

An alternative way to express the multinomial logit model (equation 1.7) is in terms of the log odds that the chosen alternative is *j* in a set of *J* alternatives, with the *Jth* alternative being the reference alternative (Agresti 1996):

$$log\left(\frac{P(j)}{P(J)}\right) = \boldsymbol{\beta}'\boldsymbol{X}_{jn}, \quad j = 1, \dots, J-1 \quad (1.7)$$

The parameter estimates are interpreted as the change in the log odds for a one unit change in the explanatory factor if all other factors are held constant (Agresti 1996).

Conditional multinomial logit models are used when the explanatory factors are specific to each choice outcome, whether alternative specific (e.g., frequency of transit) or generic factors applying to all alternatives (e.g., travel time, travel cost), while unconditional or generalized multinomial logit models are used when the explanatory factors are specific to the individual (e.g., sociodemographic, built environment, attitudinal) and/or the trip itself (e.g., trip distance) (Borooah 2002; Koppelman and Bhat 2006; Anderson and Simkins 2012).

An important property of the multinomial logit model is the independence of irrelevant alternatives (IIA), asserting that for any person, the ratio of the probabilities of making a choice for two alternatives [equation 1.8 (from Ben-Akiva and Lerman 1985)] is independent of the existence or traits of all other alternatives, meaning that other alternatives do not have an impact on the choice between any other two alternatives (Ben-Akiva and Lerman 1985; Train 1986; Borooah 2002; Koppelman and Bhat 2006). The IIA property can simplify adding or removing alternatives without changing model structure or parameters, although because it causes a uniform change in the probability of choosing all present alternatives if a new alternative is introduced, which comes from the assumption of IID unobserved utilities, it can be unrealistic if the new alternative is not completely independent from existing alternatives (Borooah 2002; Koppelman and Bhat 2006). This may be a problematic assumption for mode choice when a new alternative is introduced that is related or similar to one or more existing alternatives.

$$\frac{P_n(i)}{P_n(l)} = \frac{e^{V_{in}} / \sum_{j \in C_n} e^{V_{jn}}}{e^{V_{ln}} / \sum_{j \in C_n} e^{V_{jn}}} = \frac{e^{V_{in}}}{e^{V_{ln}}} = e^{V_{in} - V_{ln}} \quad (1.8)$$

An alternative logit model that can overcome the IIA property is the nested logit (NL) model, which nests alternatives together that are more similar to each other than they are to others, by using probabilities conditional on choosing the corresponding nest (e.g., probability of choosing bus, given that transit (nest) has been chosen), effectively relaxing the IIA assumption between nests while maintaining it within nests. The NL model follows similar assumptions as the multinomial logit model, including the Gumbel distribution of the error term for each alternative, except that the error for each alternative is broken down into one component specific to the nest, and another component specific to the alternative (Ben-Akiva and Lerman 1985; Koppelman and Bhat 2006). Likewise, the observed utility for each alternative is broken down into a component specific to the nest, and another component specific to the alternative.

Maximum likelihood is used to estimate the parameters, which involves determining the value of the parameters that indicates the highest likelihood of the full joint sample making the observed choices (Ben-Akiva and Lerman 1985; Koppelman and Bhat 2006). The likelihood function [equation 1.9 (from Ben-Akiva and Lerman 1985; Train 1986; Koppelman and Bhat 2006)] refers to the joint probability density function of the sample's chosen alternatives, and parameter values are estimated for the maximum value of the likelihood function by taking the log-likelihood function [equation 1.10 (from Koppelman and Bhat 2006)] and setting its first derivative [equation 1.11 (from Koppelman and Bhat 2006)] equal to zero.

$$L(\beta) = \prod_{n \in N} \prod_{j \in C_n} P_{in}^{\delta_{in}} \quad (1.9)$$

$$\delta_{in}(chosen\ indicator) = \begin{cases} 1, if\ j\ chosen\ by\ individual\ n \\ 0, otherwise \end{cases}$$

$$P_{in}\ is\ probability\ that\ individual\ i\ chooses\ alternative\ n$$

$$LL(\beta) = Log\big(L(\beta)\big) = \sum_{n \in N} \sum_{j \in C_n} \delta_{in} \times \ln(P_{in}(\beta)) \quad (1.10)$$

$$\frac{\partial(LL)}{\partial \beta_k} = \sum_{n \in N} \sum_{j \in C_n} \delta_{in} \times \frac{1}{P_{in}} \times \frac{\partial P_{in}(\beta)}{\partial \beta} \quad \forall k \quad (1.11)$$

The models used in this report also assume that the choice set contains a finite number of alternatives, that only one of the alternatives in the choice set can be chosen, that each individual considers all alternatives, and that the choice set includes the chosen alternative (Train 1985). The assumption of a finite number of alternatives is met in this research. The assumption that only one of the alternatives can be chosen is not met, because respondents were able to choose all modes used for their most recent trip. It is not known if each respondent actually considered all alternatives. The assumption that the choice set includes the chosen alternative is met because respondents were only included in the model if they chose one of the alternatives.

Multinomial logit models for business and non-business trips were estimated using the methods described here, with Log-likelihood and McFadden R^2 values considered as measures for goodness of fit. McFadden R^2 is a pseudo-R^2 measure that cannot be interpreted the same way as an R^2 value in ordinary least squares (OLS) regression. McFadden R^2 is also known as ρ^2 (rho-squared), with values ranging from 0.2 to 0.4 indicating model fit being very good (Louviere et al. 2000). McFadden R^2 is also helpful for comparing models using the same sample, trying to estimate the same outcome.

5.2 Applying the Logit Model to the Intercity Travel, Information, and Technology Survey Data

This analysis used unconditional, or generalized, logit models, assuming that mode choice is determined by factors describing the traveler and/or the trip (Anderson and Simkins 2012). Choice outcome variables were the travel modes, which included automobile, commercial airplane, intercity bus, and intercity rail. Separate models were estimated for business and non-business trip purposes, using revealed preference data for the most recent trip taken by survey respondents over the past year to Boston, New York City, Philadelphia, or Washington, DC.

Using the most recent trip is expected to result in more accurate models, based on the assumption that data collected about the most recent trip will be more accurate than data collected about other trips taken over the past year. Some data variables were only included in the survey for the most recent trip, including trip purpose and number of passengers. Nested logit models were also explored, with nests for different types of automobile modes and different types of transit modes. Automobile modes included personal auto, rental/car share/borrowed car, and car service. Transit modes included intercity bus, intercity rail, and commercial airplane. The models estimated here are unconditional, which some prior research has suggested is more suitable for long-distance travel in the US overall (Anderson and Simkins 2012).

The modeling methods for this study sought to include the independent variables that would result in the “best” model, considering available data, and the context of the study interests (Hosmer et al. 2013). To accomplish this, the research model-building methodology used a plan for variable selection and adequate model assessment, regarding each variable and overall model performance, with elements of purposeful selection, and stepwise selection, as described by Hosmer et al. (2013).

Four categories of explanatory factors were considered, including sociodemographic factors, geographic factors, attitudinal factors, and trip-specific factors. Factors included for initial consideration as independent variables were selected based on the literature, survey data, latent attitudinal factors prepared from the survey data, available supplemental data, and preliminary analysis. Table 5-1 through Table 5-4 present factors considered for model inclusion. The categories are: sociodemographic, built environment, latent attitudinal, and trip-specific variables. Original attitudinal statements from the survey were also considered, and although included in the estimation process, were not found to be significant in these models.

Table 5-1: Initial Factors Considered for Model Inclusion – Traveler (Sociodemographic)

Description	Code/Values	Name
No. of registered vehicles in HH	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+	Vehicles
Driver's license	1 = Yes 0 = No	License
Ratio of vehicles to licensed drivers in HH	Continuous; greater than or equal to zero	veh.per.lice.hh
No. of adults in HH	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+	Num_Adult
No. of adults in HH w/ license	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+	Num_Adult_Lic
No. of children in HH (under 18)	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+	Num_Child
No. of children in HH w/ license	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+	Num_Child_Lic
No. of licensed drivers in HH	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+	Num_Drivers
Internet access method(s)	0 = None 1 = Home 2 = School 3 = Work 4 = Public internet service (e.g., library) 5 = Mobile device with data plan 6 = Other	Internet
Technology owned	0 = None 1 = Desktop computer 2 = Laptop 3 = Smartphone 4 = Tablet 5 = Standalone GPS device	Tech
Age group (years)	1 = 18-24 2 = 25-34 3 = 35-44 4 = 45-54 5 = 55-64 6 = 65-74 7 = 75-84 8 = 85+	Age
Millennial status (age <35 years; age >= 35 years)	0 = Non-Millennial 1 = Millennial	Millennial
Gender	1 = Male 0 = Female	Gender
Education level	1 = Less than high school diploma 2 = High school diploma or equivalent 3 = Some college, no degree 4 = Associate degree 5 = Bachelor's degree 6 = Graduate or professional degree	Education
Income level	1 = Under $25,000 2 = $25,000 - $49,999 3 = $50,000 - $74,999 4 = $75,000 - $99,999 5 = $100,000 - $149,999 6 = $150,000 - $199,999 7 = $200,000 - $249,999 8 = $250,000+	Income

Table 5-2: Initial Factors Considered for Model Inclusion – Traveler (Built Environment)

Description	Code/Values	Name
State of residence	1 = Massachusetts 2 = New Hampshire 3 = Maine 4 = Vermont	State
Urban/rural	1 = Urban 0 = Rural	Urban
RUCA code	1 = Metropolitan area core 2 = Metropolitan area high commuting 3 = Metropolitan area low commuting 4 = Micropolitan area core 5 = Micropolitan area high commuting 6 = Micropolitan area low commuting 7 = Small town core 8 = Small town high commuting 9 = Small town low commuting 10 = Rural area	RUCA
Network distance from zip code centroid to centroid of Urban Cluster within Metropolitan Area	Continous	Net_Met
Network distance from zip code centroid to centroid of Urban Cluster within Micropolitan Area	Continous	Net_Mic
Network distance from zip code centroid to centroid of urban area within metropolitan area	Continous	dist_N_uam
Network distance from zip code centroid to centroid of urban cluster within micropolitan area	Continous	dist_N_ucm
Zip Code location population according to the 2010 census	Continous	Population
Network distance from zip code centroid to commercial service airport	Continous	dist_N_air
Network distance from zip code centroid to nearest medium hub or greater commercial service airport	Continous	dist_N_air.m
Network distance from zip code centroid to nearest large hub commercial service airport	Continous	dist_N_air.l
Network distance from zip code centroid to intercity rail station	Continous	Net_Rail
Network distance from zip code centroid to commuter rail station	Continous	Net_Rail_C
Network distance from zip code centroid to intercity bus station	Continous	Net_Bus_I
Network distance from zip code centroid to transit bus station	Continous	Net_Bus_T
Network distance from zip code centroid to nearest intercity	Continous	Dist_N_fac

transportation facility		
Network distance to closest intermodal transportation facility	Continous	Dist_N_intermodal
Ratio of network distance to intercity bus facility, to network distance to destination	Continous	dfdd.bus
Ratio of network distance to intercity rail station, to network distance to destination	Continous	dfdd.rr
Ratio of network distance to nearest commercial service airport, to network distance to destination	Continous	dfdd.air
Number of intercity bus terminals within 25 network miles	Integer	Count_bus
Number of intercity train stations within 60 network miles	Integer	Count_train
Number of commercial service airports within 75 network miles	Integer	Count_air
Dummy variable indicating intercity bus facility located within home zip code	1 = Yes 0 = No	Busint_dum
Dummy variable indicating intercity rail facility located within home zip code	1 = Yes 0 = No	Rrint_dum
Dummy variable indicating commercial service airport located within home zip code	1 = Yes 0 = No	Airint_dum
Dummy variable indicating any intercity transportation passenger mode facility located within home zip code	1 = Yes 0 = No	any.fac.zip
Number of urban cluster areas within home zip code area (2010 census)	1 = 1 urban cluster area within home zip code area 2 = 2 urban cluster areas within home zip code area 3 = 3 urban cluster areas within home zip code area	Count_UC

Table 5-3: Initial Factors Considered for Model Inclusion – Traveler (Latent Attitudinal)

Description	**Code/Values**	**Name**
Preference for auto	Continuous	Factor 1
Comfort with personal space and safety on bus or train	Continuous	Factor 2
Social networks: no bus or train	Continuous	Factor 3
Social networks: no car or plane	Continuous	Factor 5
Auto dependence	Continuous	Factor 6

Table 5-4: Initial Factors Considered for Model Inclusion – Trip

Description	Code/Values	Name
Destination city	1 = Boston 2 = New York City 3 = Philadelphia 4 = Washington, DC	Destination
Straight line distance to destination city	Continous	Dist_Dest_S
Network distance to destination city	Continous	Dist_Dest_N
Trip purpose	1 = Business 0 = Non-Business	
Group size	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11+	Group
Planning method	1 = Airline, bus, or train website 2 = Travel website (e.g., Orbitz.com) 3 = Called airline, bus company or train line 4 = Travel agency 5 = Friend or family member booked it 6 = Other	Planning
Nights stayed for most recent trip	0, 1, 2, 3, 4, 5, 6, 7+	Nights

The survey data did not include variables specific to each mode choice such as travel time and cost, number of transfers, or wait time, nor did they include data about respondents' explicit reasons for choosing one mode over another for a given trip. In order to employ a conditional multinomial logit model, which assumes mode choice is determined by attributes about each modal option (e.g., travel time and cost), these mode-specific attributes would first need to be calculated (Anderson and Simkins 2012). Attempts were made during the course of this research to acquire and include variables describing travel time and cost. For example, calculating airfares from the nearest commercial service airport to each destination city, using the Air Travel Price Index, a measure of changes in the cost of commercial air travel between airport pairs by quarter annual intervals (Anderson and Simkins 2012). Based on the timeline, geographic area of interest, and available resources for this research work, it was ultimately deemed prohibitive to acquire these variables, so conditional multinomial logit models were not estimated.

The diagram shown in Figure 5-1 shows a schematic representation for a simple multinomial logit model, based on the methods of Anderson and Simkins (2012), similar to those estimated in this research work. Four categories of explanatory factors are shown on the left: sociodemographic, geographic, attitudinal, and trip factors. The explanatory factors influence the mode choice made by the individual, resulting in one of four mode choice outcomes: auto, (intercity) bus, (intercity) rail, or (commercial) air.

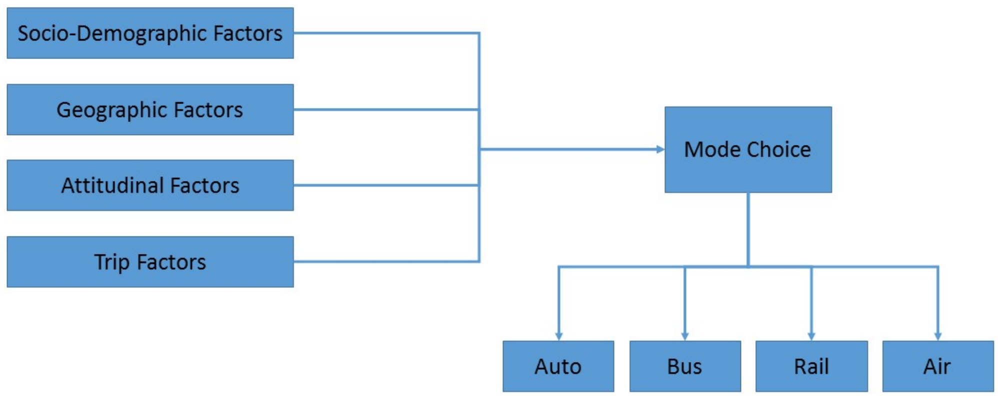


Figure 5-1: Simple Multinomial Logit Model

The diagram shown in Figure 5-2 shows a schematic representation for a nested logit model, based on the methods of Moeckel et al. (2013), similar to those estimated in this research work. Four categories of explanatory factors are again shown on the left: sociodemographic, geographic, attitudinal, and trip factors. The explanatory factors influence the mode choice made by the individual, resulting in first, a choice of nest (auto or transit), and second, one of the mode choice outcomes within that nest, with three choice outcomes in each nest. Here, the nest for auto contains three distinct kinds of auto mode: personal auto, rental or borrowed car, or car service. The nest for transit contains three distinct kinds of transit mode: (intercity) bus, (intercity) rail, or (commercial) airplane. The modes within each nest are assumed to be correlated to each other, compared with modes outside of, or in a different nest.

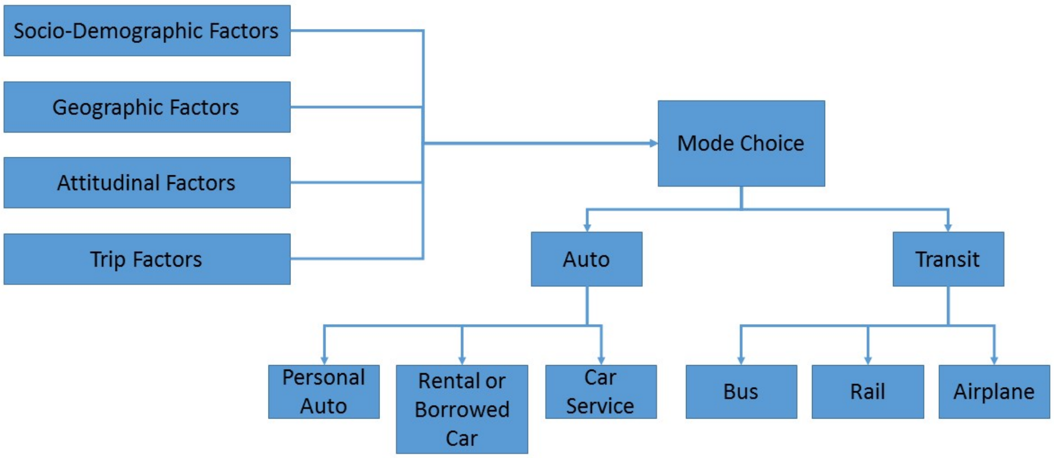


Figure 5-2: Simple Nested Logit Model

5.3 Multinomial Logit Model Estimation

Univariable analysis of all 115 potential independent variables resulted in 102 variables with a p-value of less than 0.25 from a likelihood ratio test, the entry level of significance chosen to be more inclusive at this stage. Four variables that did not meet the entry level of significance were still included, based on assumptions of subject matter relevance. These variables included ownership of a laptop, whether or not a respondent was a millennial, network distance from zip code centroid to nearest intercity bus station, and network distance from zip code centroid to nearest intercity transportation facility of any type. Table 5-5 summarizes the numbers of variables considered from each category for model inclusion.

Table 5-5: Summary of Univariable Analysis of Initial Factors Considered for Model Inclusion

Table	No. Variables	Variables with p-value < 0.25	No. Variables Selected
Traveler: Socio-Demographic	25	16	18
Traveler: Latent Factors	6	5	5
Traveler: Attitudinal Statements	27	27	27
Traveler: Built Environment	39	27	39
Trip	18	17	17
Total	**115**	**102**	**106**

A pairwise Spearman rank correlation matrix was prepared with potential independent variables, to check for correlation between independent variables. For independent variables with correlations between them, only one variable was considered for inclusion, to prevent collinearity in the model. Many variables were shown to be correlated with each other. The decision for which correlated variables to choose for consideration was based on context, and how well the variables appeared to capture relevant aspects of the decision-making process. A variable was created for the distance to the most recent destination city, to replace separate variables with distances to each destination city. After the correlation analysis, the list of variables to include in an initial multivariable model included 79 variables.

A multivariable model was fit with the initial variables identified above. During the course of building the first multivariable model, 39 variables were identified as causing a singular matrix, indicating correlation, in combination with other variables in the model, although this was not evident during univariable analysis. These 39 variables were left out of this multivariable model. The remaining 40 variables were included in the first multivariable model. The p-value for each variable was assessed from its test statistic.

Each variable not indicating significance at the 5% level was removed, one at a time, and the reduced model was compared to the full model using the likelihood ratios. The null hypothesis was that the reduced model is the 'right' model. During the course of removing these variables and performing likelihood ratio tests, the tests often failed, giving errors that indicated differences in sample sizes. This was due to missing values included for many of the variables. Sometimes there were only a small number of missing values, sometimes a large number. Attempting to perform likelihood ratio tests with models fit to different sample sizes violates assumptions of the test. Data imputation was considered, as it is recommended for dealing with missing data values by Hosmer et al. (2013). Imputation was not used though, due to potential complications that might occur for this dataset. Instead, all variables were checked for missing values. The multivariable model building process was restarted, excluding any variable with missing values. Removing individual records containing missing values was considered, but was not attempted at this point, in order to maintain a larger sample size.

Upon restarting multivariable model building, a stepwise selection procedure was followed. This is an alternative to purposeful selection (Hosmer et al. 2013) that seemed to provide more transparency in the model-building process, in terms of identifying issues introduced by different variables, whether due to causing a singular matrix, or differences in sample size. It also seemed to more clearly show the effects of adding each variable to the model during the process. When stepwise selection was employed, as described by Hosmer et al. (2013), building towards a preliminary main effects model was slow and tedious. An initial multivariable model was built with a stepwise selection procedure, from the initial variables identified above. The results did not provide a model which adequately addressed the research question.

At this point, the multivariable model building process was restarted once again. This time, a hybrid method was employed, going through a stepwise work flow, but relying more on judgment as an analyst, as to which explanatory variables make the most sense from a subject matter and preliminary analysis perspective, and relying less exclusively on statistical indicators for selection of variables. Each category of independent variables was assessed to select those appearing to be the most relevant and promising from each category, considering results of univariable analysis, cross-tabulations, and the literature. Those variables selected in this process were included first, then remaining variables were added to check for contributions in terms of significance and confounding. The resulting models were more satisfactory for determining influential factors for the purpose of answering the research question.

5.4 Nested Logit Model Estimation

Once a set of preferred multinomial logit models was selected, nested logit (NL) models were estimated using the variables included in the preferred multinomial logit models. Numerous nesting configurations were tried, with several resulting in errors indicating computational singularity. It appeared that NL models were more sensitive to this than multinomial logit models. The nesting configuration that provided the best model output included two nests: one for personal automobile and commercial airplane, with a second nest for intercity bus and intercity rail. All NL models that provided actual output had at least one nesting coefficient less than zero or greater than one, both of which are inconsistent with random utility maximization, suggesting that nesting was not needed (Koppelman and Bhat 2006). Thus, all nested logit models were rejected.

6. Results

Results indicate that for this type of travel, factors influencing mode choice for both business and non-business trips, from homes in northern New England to Boston, New York City, Philadelphia, or Washington, DC, include trip distance; land use; personal use of technology; and latent attitudes about auto dependence, preference for automobile, and comfort with personal space and safety on public transportation. Gender is a less significant factor. Age is only significant for non-business trips.

Table 6-1 displays the results for Model 1a – Business Trips MNL. Independent variables are included from each category---sociodemographic, built environment, attitudinal, and trip-specific variables. The model has a McFadden R^2 value of 0. 277. This is a pseudo-R^2 measure that cannot be interpreted the same way as an R^2 value in ordinary least squares (OLS) regression. McFadden R^2 is also known as ρ^2 (rho-squared), with values ranging from 0.2 to 0.4 indicating model fit being very good (Louviere et al. 2000). McFadden R^2 is helpful for comparing models using the same sample, trying to estimate the same outcome. In this regard, Model 1a had a higher McFadden R^2 value than other models for business trips explored during estimation. The Likelihood Ratio chi-square test for goodness-of-fit indicates this model fits significantly better than the null model. There were no alternative-specific variables, so each independent variable was included as distinct interactions with the three alternatives (airplane, intercity bus, and intercity rail) in comparison to automobile, the reference alternative. All independent variables in this model, except for gender, are shown to be statistically significant for one or more alternatives, at the 95 percent confidence level, with expected signs. Variable interactions were explored among independent variables. None were included in the preferred model however, as they did not appear to contribute to the interpretability of the model or individual variables.

Table 6-1: Model 1a Recent Business Trip MNL[1]

	Airplane		Intercity Bus		Intercity Rail	
Variables	**Coeff.**	**t-val.**	**Coeff.**	**t-val.**	**Coeff.**	**t-val.**
Intercept	-6.472	**-4.17**	-21.564	-0.01	-1.818	**-2.33**
Size of closest airport hub (small)[2]	0.623	1.11	2.048	**2.64**	0.073	0.12
Gender (male)[2]	0.566	1.05	0.264	0.42	-1.174	-1.85
Rural area (yes)[2]	0.647	0.53	16.542	0.01	-1.361	**-2.04**
Number of urban cluster areas within home zip code area (2010 census)	0.072	0.12	1.291	**2.00**	-0.908	-0.81
Distance to destination city	0.008	**5.71**	0.001	0.72	0.003	1.72
Owns tablet (yes)[2]	1.245	**2.01**	-0.760	-1.25	0.030	0.05
Preference for Auto	0.066	0.24	-1.307	**-2.74**	-0.835	**-2.23**
Comfort with Personal Space and Safety on Bus or Train	0.074	0.25	0.996	**2.66**	-0.501	-1.64
Auto Dependence	-0.092	-0.26	-1.158	**-3.01**	-0.255	-0.74
Log-likelihood	-145.5					
McFadden R^2	0.277					

[1] Automobile is reference category
[2] Binary variables presented when value = 1; otherwise value = 0

The only sociodemographic variables included in Model 1a were gender and ownership of a tablet. Gender is shown to be statistically significant at the 90 percent confidence level, indicating that males are less likely to choose intercity rail, for this type of long-distance business travel, compared with automobile. Owning a tablet computer shows a positive relationship with choosing airplane, compared with automobile, for business trips. Perhaps this is related to valuing the ability to multitask during travel, or maybe it is a proxy for income. This was the only technology type to consistently show significance in mode choice among the models developed for each purpose. Age was not shown to be a significant factor in mode choice for business trips. The fact that gender was only shown to be significant at the 90 percent confidence level in mode choice for business trips may be indicative of a narrowing gender gap, in employment and in travel (Frändberg and Vilhelmson 2014).

Built environment variables in this model include a dummy variable indicating whether or not a rural area is designated within the home zip code area, according to the 2010 US Census, and a variable indicating the number of urban cluster areas (with 2,500-49,999 people) for the home zip code area, according to the 2010 US Census. The results suggest that people living in zip codes with a designated rural area are less likely to choose intercity rail for this type of travel. This makes sense, as rural areas tend to have less access to rail facilities, compared with more urban areas. People living in zip codes with more urban cluster areas are more likely to choose intercity bus for this type of travel, compared with people living in zip codes with fewer urban cluster areas. This makes sense, as intercity bus facilities are more prevalent in areas with denser development.

Three latent attitudinal factors, prepared during the factor analysis, are included in this model. The first, 'Preference for Auto', indicates that people who, although feel capable of taking a bus or a train, have a strong preference for their car, are indeed, less likely to choose bus or train for this type of travel, compared with choosing automobile. The second, 'Comfort with Personal Space and Safety on Bus or Train', shows a positive relationship with choosing intercity bus compared with automobile for this type of travel. Being more comfortable with personal space and safety while riding a bus would make one more willing to choose this mode. The third, 'Auto Dependence', indicates that people who identify themselves as being more dependent on automobiles compared with other people, are less likely to choose intercity bus for this type of travel, compared with choosing automobile. This seems intuitive, but shows the significance of personal attitudes in choosing automobile among travel modes. The results for these attitudinal factors in the model are consistent with expectations that personal attitudes are significant in discrete mode choice (Ashok et al. 2002; Daly et al. 2012; Popuri et al. 2011).

The only trip-specific variable included in this model is network distance to destination city. The variable was included as a continuous predictor due to the smooth shape of its distribution. It indicates that as the distance to one's destination city increases, the more likely one is to choose airplane for traveling there. This makes sense, as the benefits of air travel, in terms of time and cost, are greater for longer trips. That is, as trip distance increases, more time is being saved by air travel, and cost becomes more reasonable. For business trips in particular, the amount of time saved can be crucial, and when the cost of the trip is a business expense, rather than a personal expense, it may become easier to justify.

Table 6-2 shows results from the best nested logit model built for business trips, using the same variables as the multinomial logit model for business trips. It has nesting coefficients greater than one for both nests. This indicates no correlation between modes within each nest, and is inconsistent with random utility maximization theory, suggesting nesting is not needed so the model is rejected (Koppelman and Bhat, 2006).

Table 6-2: Model 1b Recent Business Trip NL[1]
Airplane and Auto are nested together; Bus and Rail are nested together

	Nesting Coefficient	**t-val.**
Auto_Air	15.570	1.42
Transit	8.421	1.01
Log-likelihood	-149.2	
McFadden R2	0.259	

[1] Automobile is reference category

Table 6-3 displays the results for Model 2a – Non-Business Trips MNL. Independent variables are included from each category---sociodemographic, built environment, attitudinal, and trip-specific variables. The model has a McFadden R^2 value of 0.138. This pseudo-R^2 value cannot be directly compared with that from Model 1a, as models have considerable differences in sample size (263 business trips vs. 1700 non-business trips). McFadden R^2 is helpful for comparing models using the same sample, trying to estimate the same outcome. In this regard, Model 2a had a higher McFadden R^2 value than other models for non-business trips explored during estimation. The Likelihood Ratio chi-square test for goodness-of-fit indicates this model fits significantly better than the null model. There were no alternative-specific variables, so each independent variable was included as distinct interactions with the three alternatives (airplane, intercity bus, and intercity rail) in comparison to automobile, the reference alternative. All independent variables in this model, except for gender, are shown to be statistically significant for one or more alternatives, at the 95 percent confidence level, with expected signs. Variable interactions were explored among independent variables. None were included in the preferred model however, as they did not appear to contribute to the interpretability of the model or individual variables.

Table 6-3: Model 2a - Recent Non-Business Trip MNL[1]

	Airplane		Intercity Bus		Intercity Rail	
Variables	**Coeff.**	**t-val.**	**Coeff.**	**t-val.**	**Coeff.**	**t-val.**
Intercept	-5.687	**-9.18**	-2.464	**-7.34**	-2.815	**-8.71**
Gender (male)[2]	-0.366	-1.02	-0.425	-1.72	-0.158	-0.72
Age 35-54 (yes)[2]	-0.093	-0.15	-0.730	**-2.26**	-0.435	-1.45
Age 55-64 (yes)[2]	0.648	1.09	-0.186	-0.58	-0.118	-0.38
Age 65+ (yes)[2]	1.582	**2.70**	-0.293	-0.81	0.036	0.11
Network distance from zip code centroid to centroid of UC within Metropolitan Area	-0.011	**-2.12**	-0.008	-1.51	0.001	0.14
Distance to destination city	0.009	**10.15**	0.002	**3.10**	0.001	0.80
Owns tablet (yes)[2]	-0.201	-0.60	-0.487	**-2.19**	0.199	0.93
Preference for Auto	-0.051	-0.27	-0.775	**-4.62**	-0.495	**-3.46**
Comfort with Personal Space and Safety on Bus or Train	-0.100	-0.50	0.495	**3.57**	0.326	**2.56**
Auto Dependence	-0.213	-0.91	-0.588	**-4.23**	-0.553	**-4.25**
Log-likelihood	-823.01					
McFadden R2	0.138					

[1] Automobile is reference category
[2] Binary variables presented when value = 1; otherwise value = 0

Sociodemographic variables include gender, age, and owning a tablet. Gender is shown to be statistically significant at the 90 percent confidence level, indicating that males are less likely to choose intercity bus, for this type of travel, compared with automobile. People ages 35-54 years are shown to be less likely to choose intercity bus, than are people younger than 35 years, compared to choosing automobile. This is consistent with the literature (McDonald 2015). People ages 35-54 years are probably more likely to have children, which may make bus travel more challenging than travel by other modes. Traveling with children by automobile is likely to be more flexible than other modes. People younger than 35 years, particularly in this generation, may have less income than those 35-54 years, which also may contribute to the likelihood of choosing bus for this type of travel. People age 65 years and older are shown to be more likely to choose airplane, than are people younger than 35 years, versus choosing automobile. It is possible that people age 65 years and older have more disposable income, compared with those younger than 35 years, which makes it easier to choose commercial air travel for these kinds of trips. For non-business trips, owning a tablet indicates a decreased preference for intercity bus, compared with automobile. Although this is not consistent with the notion of the value of multi-tasking during travel that was raised for business trips, it may be more of an economic indicator for the case of non-business trips. That is, owning a tablet may be an indicator of higher income, compared with owning a smartphone, which now could be more prevalent across income levels. Perhaps owning a tablet is a proxy for higher income that makes one less likely to choose a bus, compared with automobile, for this type of travel.

The only built environment variable included in this model is the network distance from home zip code centroid to the centroid of an urban cluster within a metropolitan area. It indicates that the greater the distance is from a person's home zip code to an urban cluster within a metropolitan area, the less likely one is to choose airplane for this type of non-business travel. Commercial airports are usually located in the vicinity of metropolitan areas. So, the farther one is from a metropolitan area, the farther one is from a commercial service airport, and the less likely one is to choose that mode, compared with choosing automobile.

Three latent attitudinal factors were included in this model. The first, 'Preference for Auto', indicates that people who, although they feel capable of taking a bus or a train, have a strong preference for their car, are less likely to choose bus or train for this type of travel, compared with automobile. The second, 'Comfort with Personal Space and Safety on Bus or Train', shows a positive relationship with choosing intercity bus or intercity rail, compared with automobile for this type of travel. The third, 'Auto Dependence', indicates that people who identify themselves as being more dependent on automobiles compared with other people, are less likely to choose intercity bus or intercity rail for this type of travel, compared with automobile. The results for these attitudinal factors in the model are comparable with the results shown for business trips, and consistent with expectations that personal attitudes play a prominent role in mode choice (Ashok et al. 2002; Daly et al. 2012; Popuri et al. 2011).

Table 6-4 shows results from the best nested logit model built for non-business trips. It has a nesting coefficient between zero and one for the transit nest, indicating correlation between transit modes, which is appropriate (Koppelman and Bhat, 2006). However, the nesting coefficient for the automobile and air travel nest is greater than one, indicating no correlation within the nest. This is inconsistent with random utility maximization principles (Koppelman and Bhat, 2006), suggesting nesting is not needed, so the model is rejected.

Table 6-4: Model 2b - Recent Non-Business Trip NL[1]
Airplane and Auto are nested together; Bus and Rail are nested together

	Nesting Coefficient	t-val.
Auto_Air	4.509	1.85
Transit	0.102	0.12
Log-likelihood	-820.71	
McFadden R2	0.140	

[1] Automobile is reference category; Airplane and Auto are nested together

7. Conclusions

The results of this project reinforce the importance and viability of modeling mode choices for long-distance travel from less populated regions to large metropolitan areas, and the significant roles of trip distance, built environment, personal attitudes, and sociodemographic factors in how people choose mode to make these trips for different purposes. As supported by the survey data used in this research, this kind of long-distance travel, from residences in less populated parts of northern New England to major metropolitan areas in the Northeast, is occurring relatively frequently. These trips are important for quality of life, multimodal planning, and rural economies, supporting the need to understand this kind of travel and account for it in transportation systems planning, design, and programming.

The success of estimating the multinomial logit models indicates that the unique survey dataset used is appropriate for these kinds of long-distance intercity mode choice models, and that similar survey datasets designed, conducted, and augmented for other regions of the country could be appropriate for similar models there. It also indicates that the models are robust enough to make use of this survey dataset, providing useful results, and could be transferable to other regions.

Contributions of this project are the identification of factors and quantification of their influence, on mode choice for intercity travel originating from homes in less populated areas of northern New England, going to large metropolitan destinations in the Northeast. Factors were found to be significant from four categories: sociodemographic, environmental, attitudinal, and trip-specific. These included gender, age, land use, distance to urban metropolitan areas, owning a tablet computer, latent attitudinal factors, and trip distance.

Having a rural designation within one’s home zip code is associated with being less likely to choose intercity rail for these types of long-distance business trips. Previous research indicates increased stated preference for intercity bus and rail originating in rural areas, with increases in fuel costs (Mattson et al. 2010). Looking at investments for passenger rail infrastructure and service, including augmentation of rail facilities and future expansion of intercity rail services, rural areas should be taken into account for targeting growth in potential passenger demand.

Three latent attitudinal variables strongly contributed to the models, showing significance for both business and non-business trips, for this type of travel: preference for automobile; comfort with personal space and safety on bus or train; and automobile dependence. This reinforces the importance of attitudes in decision-making for mode choice (Popuri et al. 2011, Daly et al. 2012, Ashok et al. 2002, Walker and Ben-Akiva 2002). It also indicates the potential value of emphasizing marketing efforts to promote bus or train modes for long-distance intercity travel and to help reduce dependence on the automobile, which continues to dominate mode share for this type of travel.

Another contribution of this research is a better understanding of the differences in mode choices between genders and among age groups, in the context of intercity travel from homes in less populated areas in the study region, going to large metropolitan areas in the Northeast. These differences were looked at both with and without controlling for other factors. More differences were shown among age groups and between genders, for non-business travel than for business travel. Age was only shown to be significant for non-business travel, with people in the middle age ranges being less likely to choose intercity bus, and older people being more likely to choose airplane for non-business trips. This may reflect life stages in terms of flexibility for family travel and discretionary income for retirement, respectively. Gender was only shown to be significant at the 90% confidence level, for both business and non-business trip purposes.

7.1 Limitations

Travel time and cost were not included in the models built during this research. As mentioned at the beginning of Chapter 5, this research work used unconditional, or generalized, multinomial logit models, assuming that mode choice is determined by factors describing the traveler and/or the trip, which some existing research indicates is more appropriate for long-distance travel than time and cost (Anderson and Simkins 2012). However, it is possible that fitting a set of conditional models, that do incorporate time and cost, might prove beneficial for addressing the research question.

The factor analysis was successful in preparing six latent factors related to attitudes towards transportation, three of which contributed significantly to the models built. However, during the course of the factor analysis, it appeared that the survey questionnaire was not developed specifically with the intent for performing factor analysis. This seemed to be the case due to the large number of attitudinal statements that were ultimately not accounted for in the factor analysis. Even with only 49 out of 70 attitudinal statements showing significance at an entry level p-value of 0.25 from the likelihood ratio test during univariable analysis for mode choice, only 24 attitudinal variables were included in the preferred factor model. In the future, designing attitudinal statements differently for similar survey questionnaires could help to better support factor analysis, and potentially generate additional significant latent factors.

One potential factor that influences respondents' mode choice for this type of travel could be seasonality and weather, particularly for the Northeast US. The survey questionnaire did not include information pertaining to time of year or weather conditions for their most recent trip.

The survey option for air travel was “airplane”. It was assumed, during the course of this project research, that choosing “airplane” meant commercial air travel. It is possible some respondents used general aviation, instead of commercial air travel, including air taxi or corporate air travel. This might be likely considering the larger percentage of higher income levels in the survey sample compared with the population.

The survey questionnaire did not collect information about whether or not the destination cities of interest were the final destination of the trip, or just a stopover to another city, either domestic or international. That information might also influence people's mode choice for these kinds of trips.

7.2 Future Research

Future research should continue to improve these types of long-distance mode choice models by incorporating travel time and cost into a set of conditional mode choice models with similar individual-specific variables as those included here, developing more specific attitudinal statements to expand latent factor analysis, and further exploring built environment variables. In addition to further studying mode choice, the other three components of the traditional four step transportation planning process should also be studied more for these types of long-distance intercity trips: trip generation, trip distribution, and route choice. Future survey questionnaire development should also consider including seasonality and weather conditions for trips, capturing potential general aviation travel, and specifying whether the destination city is a final destination, a stop on a trip abroad, or just one leg of a domestic tour. Further improvement to these types of long-distance intercity mode choice models for trips from non-metropolitan areas to large metropolitan centers will continue to promote better planning, engineering, operations, and infrastructure investment decisions for transportation systems in many regions and communities across the United States which have not yet been well studied, possibly reducing levels of service.

Overall Conclusions

Part I indicates that for traveling to NYC by bus or train, an online planning tool as part of the survey was related to positive attitudes about scheduling flexibility and travel time for certain age and education groups. It was also related to negative attitudes about the ability to get and understand schedules for a bus or train to NYC. Further analysis will aim to better quantify the impacts of access to trip planning information, on attitudes about intercity travel by automobile, bus, and passenger rail.

Survey respondents were asked about their likelihood to choose a bus or a train for the imaginary trip to NYC, both at the beginning and end of Part 3, which covered this imaginary trip. Forty-three percent of respondents changed attitudes about choosing a bus or train for a trip to NYC. More people, who did have access to the tool, were becoming more likely to choose a bus or a train, while more people, who did not have access to the tool, were becoming less likely to choose a bus or a train.

The multimodal dataset will be used for future research work, particularly to analyze the accessibility to large metropolitan areas by multiple modes from origins throughout the region. Future work, using this dataset, will include developing an accessibility index that provides a measurement for this type of multimodal accessibility, and exploring how this type of accessibility, to large metropolitan areas, originating from areas outside of large metropolitan areas, varies across the Northeast US. Part of the analysis will explore how multimodal accessibility might vary across the region in ways that aren't explained by distance alone.

Possible research questions, pertaining to the survey and network datasets presented in this report, to be addressed in future research include:

- How does locational and/or individual accessibility to large metropolitan areas vary over space and time?
- If accessibility to large metropolitan areas, from Northern New England, can be measured and mapped, can areas then be identified, with greater accessibility, in truth, than areas with the same measured score? If so, what is it that increases their accessibility?
- How do attitudes among study participants, about traveling by multiple modes, to large metropolitan areas, compare with accessibility levels that are calculated, for their origin zipcodes, using the multimodal network?
- How does accessibility to large metropolitan areas, by multiple travel modes, relate to population density and urban form?
- What is a healthy relationship between the level of accessibility, by multiple travel modes, to population density and urban form?

The research and datasets presented in this report will provide the foundation for future research that will explore multimodal accessibility across the Northeast US, its relationship to population density and urban form, and address some of the possible research questions included here.

Part II results reinforce the importance and viability of modeling mode choices for long-distance travel from less populated regions to large metropolitan areas, and the significant roles of trip distance, built environment, personal attitudes, and sociodemographic factors in how people choose mode to make these trips for different purposes. As supported by the survey data used in this research, this kind of long-distance travel, from residences in less populated parts of northern New England to major metropolitan areas in the Northeast, is occurring relatively frequently. These trips are important for quality of life, multimodal planning, and rural economies, supporting the need to understand this kind of travel and account for it in transportation systems planning, design, and programming.

The success of estimating the multinomial logit models indicates that the unique survey dataset used is appropriate for these kinds of long-distance intercity mode choice models, and that similar survey datasets designed, conducted, and augmented for other regions of the country could be appropriate for similar models there. It also indicates that the models are robust enough to make use of this survey dataset, providing useful results, and could be transferable to other regions.

Contributions of this project are the identification of factors and quantification of their influence, on mode choice for intercity travel originating from homes in less populated areas of northern New England, going to large metropolitan destinations in the Northeast. Factors were found to be significant from four categories: sociodemographic, environmental, attitudinal, and trip-specific. These included gender, age, land use, distance to urban metropolitan areas, owning a tablet computer, latent attitudinal factors, and trip distance.

Having a rural designation within one's home zip code is associated with being less likely to choose intercity rail for these types of long-distance business trips. Previous research indicates increased stated preference for intercity bus and rail originating in rural areas, with increases in fuel costs (Mattson et al. 2010). Looking at investments for passenger rail infrastructure and service, including augmentation of rail facilities and future expansion of intercity rail services, rural areas should be taken into account for targeting growth in potential passenger demand.

Three latent attitudinal variables strongly contributed to the models, showing significance for both business and non-business trips, for this type of travel: preference for automobile; comfort with personal space and safety on bus or train; and automobile dependence. This reinforces the importance of attitudes in decision-making for mode choice (Popuri et al. 2011, Daly et al. 2012, Ashok et al. 2002, Walker and Ben-Akiva 2002). It also indicates the potential value of emphasizing marketing efforts to promote bus or train modes for long-distance intercity travel and to help reduce dependence on the automobile, which continues to dominate mode share for this type of travel.

Another contribution of this research is a better understanding of the differences in mode choices between genders and among age groups, in the context of intercity travel from homes in less populated areas in the study region, going to large metropolitan areas in the Northeast. These differences were looked at both with and without controlling for other factors. More differences were shown among age groups and between genders, for non-business travel than for business travel. Age was only shown to be significant for non-business travel, with people in the middle age ranges being less likely to choose intercity bus, and older people being more likely to choose airplane for non-business trips. This may reflect life stages in terms of flexibility for family travel and discretionary income for retirement, respectively. Gender was only shown to be significant at the 90% confidence level, for both business and non-business trip purposes.

Future research should continue to improve these types of long-distance mode choice models by incorporating travel time and cost into a set of conditional mode choice models with similar individual-specific variables as those included here, developing more specific attitudinal statements to expand latent factor analysis, and further exploring built environment variables. In addition to further studying mode choice, the other three components of the traditional four step transportation planning process should also be studied more for these types of long-distance intercity trips: trip generation, trip distribution, and route choice. Future survey questionnaire development should also consider including seasonality and weather conditions for trips, capturing potential general aviation travel, and specifying whether the destination city is a final destination, a stop on a trip abroad, or just one leg of a domestic tour. Further improvement to these types of long-distance intercity mode choice models for trips from non-metropolitan areas to large metropolitan centers will continue to promote better planning, engineering, operations, and infrastructure investment decisions for transportation systems in many regions and communities across the United States which have not yet been well studied, possibly reducing levels of service.

References

Afifi, A., and Clark, V. (1996). Computer-aided multivariate analysis. CRC Press.

Agresti, A. (1996). An introduction to categorical data analysis. New York: J. J Wiley and Sons.

Ajzen, Icek. (1991) "The theory of planned behavior." Organizational behavior and human decision processes, 50(2), 179-211.

Anderson, M. S., and Simkins, J. M. (2012). "Report D – Revised Comprehensive Report Development of Long Distance Multimodal Passenger Travel Modal Choice Model."

Ashiabor, S., Baik, H., and Trani, A. (2007). "Logit Models for Forecasting Nationwide Intercity Travel Demand in the United States." Transportation Research Record: Journal of the Transportation Research Board, 2007, 1–12.

Ashok, K., Dillon, W. R., and Yuan, S. (2002). "Extending Discrete Choice Models to Incorporate Attitudinal and Other Latent Variables." Journal of Marketing Research, 39(1), 31–46.

Aultman-Hall, L., Harvey, C., Sullivan, J., and LaMondia, J. "Framing Classes of Long Distance Travel." 95th Annual Meeting. Transportation Research Board. Walter E. Washington Convention Center, Mount Vernon Place Northwest, Washington, DC. 12 January 2015. Conference Presentation.

Axhausen, Kay W. "Preparing for the long of it: methodological research for an European survey of long distance travel." conference. National Research Council (United States) Transportation Research Board, 2000.

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand (Vol. 9). MIT press.

Bierce, E., and Kurth, D. "The Use of Three Surveys for Long Distance Travel Estimates in California." 93rd Annual Meeting. Transportation Research Board. January 2014. Washington, DC. Paper No. 14-5674.

Borooah, Vani K. (2002) Logit and probit: Ordered and multinomial models. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-138. Thousand Oaks, CA: Sage.

Brueckner, J. K. (2003). Airline traffic and urban economic development. Urban Studies, 40(8), 1455-1469.

Daly, A., Hess, S., Patruni, B., Potoglou, D., and Rohr, C. (2012). "Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour." TRANSPORTATION, 39(2), 267–297.

Dargay, J. M., and Clark, S. (2012). "The determinants of long distance travel in Great Britain." Transportation Research Part A: Policy and Practice, 46(3), 576–587.

Frändberg, L., and Vilhelmson, B. (2014). “Spatial, Generational and Gendered Trends and Trend-Breaks in Mobility.” Handbook of Sustainable Travel, T. Gärling, D. Ettema, and M. Friman, eds., Springer Netherlands, 15–32.

Fry, R. (2013). “Young Adults After the Recession: Fewer Homes, Fewer Cars, Less Debt.” Pew Research Center’s Social & Demographic Trends Project.

Fuller Jr, E. L., and Hemmerle, W. J. (1966). “Robustness of the maximum-likelihood estimation procedure in factor analysis.” Psychometrika, 31(2), 255–266.

Hjorthol, R. J., Levin, L., and Sirén, A. (2010). “Mobility in different generations of older persons: The development of daily travel in different cohorts in Denmark, Norway and Sweden.” Journal of Transport Geography, Special issue: The mobility of older people, 18(5), 624–633.

Hosmer, D. W., H., Lemeshow, S., and Sturdivant, R. X. (2013). Applied Logistic Regression. John Wiley & Sons.

Kim, J.-O., and Mueller, C. W. (1978). Factor Analysis: Statistical Methods and Practical Issues. SAGE.

Kim, J.-O., and Mueller, C. W. (1978). Introduction to Factor Analysis: What It Is and How To Do It. SAGE.

Kleinbaum, D., Kupper, L., Nizam, A., and Rosenberg, E. (2013). Applied Regression Analysis and Other Multivariable Methods. Cengage Learning.

Koppelman, F. S., and Sethi, V. (2005). “Incorporating variance and covariance heterogeneity in the Generalized Nested Logit model: an application to modeling long distance travel choice behavior.” Transportation Research Part B: Methodological, 39(9), 825–853.

Koppelman, F. S., & Bhat, C. (2006). “A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models.” US Department of Transportation, Federal Transit Administration, 31.

LaMondia, J., Aultman-Hall, L., and Greene, E. (2014). “Long-Distance Work and Leisure Travel Frequencies.” Transportation Research Record: Journal of the Transportation Research Board, 2413, 1–12.

Lee, B. H. Y., Aultman-Hall, L., Coogan, M., and Adler, T. (2015). “Rideshare mode potential in non-metropolitan areas of the northeastern United States.” Journal of Transport and Land Use, 0(0).

Louviere, Jordan J., David A. Hensher, and Joffre D. Swait. (2000). Stated choice methods: analysis and applications. Cambridge University Press.

Mallett, W. J. (2001). "Long-Distance Travel by Low-Income Households" Transportation Research Board. Transportation Research Circular (No. E-C026). Personal Travel: The Long and the Short of It. Conference Proceedings. 1999. Washington, DC.

Mattson, J., Peterson, D., Thomas, W., Ripplinger, D., and Hough, J. (2010). “Assessing Demand for Rural Intercity Transportation in a Changing Environment.” Transportation Research Record, Vol. 2145, pp 108-112. =

McDonald, N. C. (2015). “Are Millennials Really the ‘Go-Nowhere’ Generation?.” Journal of the American Planning Association, 81(2), 90–103.

Miller, Eric. (2004). "The trouble with intercity travel demand models." Transportation Research Record: Journal of the Transportation Research Board 1895: 94-101.

Moeckel, R., Fussell, R., and Donnelly, R. (2013). “Mode choice modeling for long-distance travel.” Transportation Letters, 7(1), 35–46.

Neely, S., Lee, B. H. Y., and Sentoff, K. (2015). “Intercity Travel in Northeastern Rural Regions of the U.S.” University of Vermont Transportation Research Center. Report 13-001

Nwabueze, Chioma, J., Onyeagu, I, S., and Onyedikachi, I. (2009). “Robustness of the maximum likelihood estimation procedure in factor analysis.” African Journal of Mathematics and Computer Science Research, 2(5), 081–087.

Outwater, M., Bradley, M., Ferdous, Pendyala, R., Garikapati, V., Bhat, C.R., Dubey, S., LaMondia, J.J., Hess, S., Daly, A. (2015). “Foundational Knowledge to Support a Long-Distance Passenger Travel Demand Model: Implementation Report” United States Federal Highway Administration. Exploratory Advanced Research Program. DTFH61-10-R-00036

Polzin, S. E., Chu, X., and Godfrey, J. (2014). “The impact of millennials’ travel behavior on future personal vehicle travel.” Energy Strategy Reviews, US energy independence: Present and emerging issues, 5, 59–65.

Pooley, C., and Turnbull, J. (2005). Migration And Mobility In Britain Since The Eighteenth Century. Routledge.

Popuri, Y., Proussaloglou, K., Ayvalik, C., Koppelman, F., and Lee, A. (2011). “Importance of traveler attitudes in the choice of public transportation to work: findings from the Regional Transportation Authority Attitudinal Survey.” Transportation, 38(4), 643–661.

Presser, H. B., and Hermsen, J. M. (1996). “Gender Differences in the Determinants of Work-Related Overnight Travel among Employed Americans.” Work and Occupations, 23(1), 87–115.

Rencher, A. C., & Christensen, W. F. (2012). Methods of Multivariate Analysis, John Wiley & Sons.

Rummel, R. J. (1988). Applied Factor Analysis. Northwestern University Press.

Schimek, P. (1996). Household motor vehicle ownership and use: How much does residential density matter?. Transportation Research Record: Journal of the Transportation Research Board, (1552), 120-125.

Spicer, J. (2005). Making Sense of Multivariate Data Analysis: An Intuitive Approach. SAGE.

Train, Kenneth. (1986). Qualitative choice analysis: Theory, econometrics, and an application to automobile demand. Massachusetts Institute of Technology Press Series in Transportation Studies, Vol. 10. Cambridge, MA

Walker, J., and Ben-Akiva, M. (2002). "Generalized random utility model." Mathematical Social Sciences, Random Utility Theory and Probabilistic measurement theory, 43(3), 303–343.

Yates, R. D., and Goodman, D. J. (2005). Probability and stochastic processes: a friendly introduction for electrical and computer engineers (Vol. 2). John Wiley & Sons

Appendix A: Survey Instrument

Intercity Travel, Information, and Technology Survey Questionnaire

Part 1: Recent intercity travel trips and general travel preferences

Section 1-A: The following questions are about your recent trips.

1. How many times have you visited one of the following cities in the past twelve months? (Exclude trips where the city was not the primary destination and you only passed through it on the way to another destination) It may be helpful to refer to your calendar or daybook to recall your trips from the last twelve months
   [Column = Frequency (drop-down box from 0 to 11, then 12 or more); Row = City]
   - Boston
   - New York City
   - Philadelphia
   - Washington, DC

   [If only one city has frequency > 0], then identify this city as <recent city>, go to 3].

   [If more than one city has frequency > 0], then go to 3].

   [If 0 cities visited, then skip to Section 1-C]

[if visited at least one city above in past twelve months]

2. What mode(s) of transportation have you used for your trip(s) to each city in the past twelve months? Please select all that apply. [Column = Mode; Row = City]
   - Personal auto/car
   - Rental car (including car share) or a borrowed car
   - Intercity bus (e.g., Greyhound, Peter Pan, Megabus)
   - Intercity rail (e.g., Amtrak)
   - Airplane
   - Other

   [Programmer: only show rows for cities that were visited in past 12 months]

[if visited at least one city above in past twelve months]

3. [*If intercity bus or intercity rail selected for ANY city*] How do you usually get information about routes and schedules for bus or rail trips? *Please select all that apply.*
   - Use pamphlets or other printed material
   - Ask a friend or family member
   - Visit the station

- Call the bus or rail company
- Search the internet
- Use smart phone or tablet apps
- Other, please specify:

[Programmer: randomize order of answer options]

[if # of cities visited > 1]

4. Which city did you visit most recently? [Choices are from those cities visited with frequency > 0]
   Answer = <recent city>

Section 1-B: The following questions are about your MOST RECENT trip to <recent city>.

5. [Skip if frequency to # cities visited = 1]
   What mode(s) of transportation did you use for your MOST RECENT trip to <recent city>? *Please select all that apply.*
   - Personal auto/car
   - Rental car (including car share) or a borrowed car
   - Intercity bus (e.g., Greyhound, Peter Pan, Megabus)
   - Intercity rail (e.g., Amtrak)
   - Airplane
   - Other, please specify:

6. What was the purpose of your most recent trip to <recent city>? *Please select all that apply.*
   - Leisure/vacation
   - Visit friends
   - Business
   - Family event
   - Other, please specify:

7. How many people travelled with you on your most recent trip to <recent city>? (Exclude those who did not make at least part of the journey with you)
   [Drop-Down for each age group = Number (drop-down box from 0 to 10, then 11 or more); Row = Age Group]
   - # Adults (18 and over):
   - # Children (under 18):

   [if bus, rail, or plane trip]

8. How did you plan this trip and book your tickets? *Please select all that apply*.

- Went to the airline, bus, or train website
- Went to a travel website (e.g., Expedia.com, Kayak.com)
- Called the airline, bus company, or train line
- Through a travel agency
- A friend or family member booked it for me
- Other, please specify: ______________________

9. [NIGHTS] How many nights did you stay for your most recent trip to < recent city >?
   [Drop-down box from 0 to 6, then 7 or more]

Section 1-C: The following are general travel and communication questions about you and your household.

10. How many registered vehicles (in working order) are available to your household?
    Please include all cars, pickup trucks, minivans, and motorcycles/scooters to which your household has regular access, whether owned, leased, or a company vehicle.
    [Drop-down box from 0 to 9, then 10 or more]

11. Do you have a driver's license?
    - Yes
    - No

12. How many people live in your household? How many of you are licensed drivers? [Two columns: People (including yourself), Drivers (including yourself) - (drop-down boxes from 0 to 9, then 10 or more); Row = Age Group]
    - # Adults (18 and over):
    - # Children (under 18):

13. How do you access the internet? *Please select all that apply.*
    - Internet service at home
    - Internet service at school
    - Internet service at work
    - Public internet service (e.g., at the library, community center)
    - Mobile device with a cellular data plan (e.g., smart phone, internet-enabled tablet)
    - Other, please specify:

Part 2: Travel preferences

In this section, consider the following statements and select how much you agree or disagree on a scale from 1 (completely agree) to 7 (completely disagree).

[programmer: scale formatted as shown below – columns evenly spaced (though feel free to make this prettier and use this format throughout the survey for this scale):

Completely Agree			*Neutral*			*Completely Disagree*
1	*2*	*3*	*4*	*5*	*6*	*7*

14. I feel I am less dependent on cars than my parents are/were.
15. I need to drive my car to get where I need to go.
16. I love the freedom and independence I get from owning one or more cars.
17. It would be hard for me to reduce my driving mileage.
18. For me to be able to leave the driving to someone else(e.g., a bus driver) would be desirable.
19. It would be desirable for my household to be able to have fewer cars.
20. Being able to freely perform tasks, including using a laptop, tablet, or smartphone is an important reason for me to choose bus or train travel.
21. Having reliable WiFi internet access while I travel on a bus or train is important to me.
22. When taking a bus or train, being able to plan my trip and buy tickets online is important to me.
23. It would be important to me to receive email or text message updates about my bus or train trip.
24. I find tablet or smartphone apps for travel and trip planning to be helpful.
25. When the government tries to improve things, it never works.
26. If everyone works together, we could improve the environment and future for the earth.
27. People like me take the bus or the train.
28. I would be willing to pay more when I travel if it would help the environment.
29. I tend to use the fastest form of transportation, regardless of cost.
30. For me, the whole idea of being on a bus or train with other people I do not know seems uncomfortable.
31. I enjoy being out and about and observing people.
32. I don't mind traveling with people I do not know.
33. Having my privacy is important to me when I travel.
34. When I choose a home, I value having adequate space for parking two or more cars.
35. When I choose a neighborhood to live in, I like to be able to walk to a commercial or village center.
36. Living in a multiple family building (e.g., apartment, condo) wouldn't give me enough privacy.
37. I like living in a neighborhood where there is a lot going on.

38. I am confident that if I want to, I can do things that I have never done before.
39. I worry about crime or other disturbing behavior on buses and trains, or while walking in and around the stops/stations.
40. It is important to me to control the radio and the air conditioning in the car.
41. I feel really stressed when driving for a long time in congestion in and around big cities.
42. I prefer to use the most comfortable transportation mode regardless of cost or time.
43. Having a low-stress trip is more important than reaching my destination quickly.
44. I get very annoyed being stuck behind a slow driver.
45. I am usually in a hurry when I make a trip.
46. With my schedule, minimizing time spent traveling is very important to me.
47. I would use the bus or train more often if it were cheaper to ride.
48. Rather than owning a car, I would prefer to borrow, share, or rent a car just for when I need it.

Part 3: An imaginary situation

Imagine that someone has asked you to travel from your home to Manhattan in New York City (NYC) for an important appointment next month and you have decided to go. You will stay one night at a hotel and travel alone. Your host will pay for your hotel costs but not for getting you there; you would be responsible for all costs of gas, parking, or any fares. Assume that, for one reason or another, you have already decided that you will not take any part of the trip by plane.

You now need to choose between taking the entire trip by car (whether yours or not) and taking at least part of the trip by intercity bus or train.

Please imagine the situation described as you answer the questions in the sections that follow.

49. Knowing what you know right now, what mode(s) of transportation do you think are AVAILABLE to you for this trip to NYC? *Please select all that apply.*
   - Personal auto/car
   - Rental car (including car share) or a borrowed car
   - Intercity bus (e.g., Greyhound, Peter Pan, Megabus)
   - Intercity rail (e.g., Amtrak)
   - Other, please specify:

50. How likely are you to choose to take a bus or train for a trip like this to NYC next month?
   - Definitely
   - Very likely
   - Likely
   - Neutral

- Unlikely
- Very unlikely
- Definitely not

51. If you learned there would be no WiFi, and no electrical outlet on the bus or train for this trip, would that make to you less likely to choose a bus or train for this trip?
    - Much less likely
    - Somewhat less likely
    - No change
    - Not applicable to me

TEST GROUP ONLY:

Now we would like you to review a website related to your imaginary trip to NYC. This

website will show you some travel options from your home area to Times Square in NYC by combinations of bus and rail.

When you click on the link below, a second window with this website will open.

You can center the website within the pop-up screen using the up/down arrows on the

right. You can ask to see more rail and bus services by using the down arrow in the center

of your pop-up window.

When you are done reviewing the website, please close the second window and click “next”

to continue.

Please click <here> to review this website.

[Test group only]

Please consider the website you looked at earlier and select the how much you agree or disagree on a scale from 1 (completely agree) to 7 (completely disagree).

52. There are more options than what I expected to travel to NYC by bus and train.
53. After seeing the bus and train options for traveling to NYC, I just don’t think there’s a good way for me to get there by either bus or train.

[Both groups]

54. Having information like this on my smartphone or computer might make it easier for me to understand the kinds of bus and train services available to me.
55. Having so many potential travel options by bus and train is confusing.

Please continue to imagine the NYC trip situation described*,* consider the following statements, and select how much you agree or disagree on a scale from 1 (completely agree) to 7 (completely disagree).

56. When I drive long distances (like from my home area to NYC), I can get tired and stressed.
57. I worry about the difficulty in finding a parking space at a reasonable cost when I get to NYC.
58. I am concerned that the schedule of the bus or train only lets me travel a few times per day, and I need to be flexible.
59. I could deal with the limited schedules offered by a bus or train for this trip from my home to NYC.
60. I like the idea that I might see and meet new people on a bus or train to NYC.
61. I don't like the idea of riding with a lot of people that I don't know on a bus or train.
62. If I took a bus or train to NYC, I might have to be with people whose behavior I find unpleasant.
63. I could be with other people who share my values when I take a bus or train on a trip like this.
64. I think that taking a BUS to NYC would take a lot longer than driving.
65. I think that taking a TRAIN to NYC would take a lot longer than driving.
66. Without thinking about it much, I would guess that the cost of taking the trip by BUS would be less than the cost of the car trip (including gas, tolls, and parking).
67. Without thinking about it much, I would guess that the cost of taking the trip by TRAIN would be less than the cost of the car trip (including gas, tolls, and parking.)
68. It would be really important to me to minimize costs when I plan this trip to NYC next month.
69. I really want to minimize the time I spend on the trip to NYC, even if that means more stress or higher costs.
70. Being able to use my laptop, tablet, or smartphone when traveling makes me more interested in taking a bus or train to NYC.
71. I am the kind of person who would take my own car to NYC.
72. Most people whose opinions I value would approve of my taking this trip by bus or train.
73. My family would think that I should take this kind of trip by car or plane.
74. My colleagues would likely think that it is strange not to go by a car or plane to NYC.
75. When my friends go to NYC, they always take a bus or train.
76. When my family members go to NYC, they always take a bus or train.
77. It might be unsafe to make this trip by bus or train.
78. The experience at the NYC bus or train station would be so unpleasant that I would try to avoid it.
79. It would be easy for me to get the schedules for a bus or train between here and NYC, and I would understand them.
80. I like the idea of taking a bus or train instead of driving for this trip to NYC.
81. I think that the most RATIONAL choice would be to take a bus or train instead of a car.

82. I think that the most PLEASURABLE choice would be to take a bus or train instead of a car.
83. I think that the most STRESSFUL choice would be to take a bus or train instead of a car.
84. All other things being equal, if a bus was cheaper, but less reliable than a train, I would choose to take a bus.
85. I am confident that if I wanted to, I could take a bus or train for such a trip to NYC next month.
86. I would make an effort to choose a bus or train for such a trip to NYC next month.
87. For me to take a bus or train for such a trip to NYC the next month would be impossible.
88. In this imaginary situation, I would plan to take a bus or train for this trip to NYC next month.

[Test group only]

89. I would trust the person who invited me to NYC to recommend how I should travel.
90. I don't know all the things I NEED to do to make this trip work by bus or train.
91. Given what you know about bus and train services to NYC, how likely are you to choose a bus or train for a trip to NYC next month (like the one described in the imaginary situation)?
    - Definitely
    - Very likely
    - Likely
    - Neutral
    - Unlikely
    - Very unlikely
    - Definitely not

[Test Group Only – If likelihood to take train/bus changed from Question 52]

92. We noticed that you are now <more/less> likely to take the train or bus to NYC. Please tell us why you have changed your mind.
    - [open-end]

[Both groups]

93. Thank you for sharing your opinions about the imaginary trip to NYC. In your real life, how seriously would you consider taking a bus or train to NYC?
    - Definitely would consider
    - Very likely would consider
    - Likely would consider
    - Neutral
    - Unlikely to consider
    - Very unlikely to consider
    - Definitely not consider

Part 4: Other information about you and your household.

94. Which of the following do you own? *Please select all that apply.*
    - Desktop computer
    - Laptop
    - Smartphone
    - Tablet (e.g., iPad, Windows 8 Tablet)
    - Standalone GPS Navigation Device (e.g., Garmin, TomTom)
    - None of the above
95. What is your age?
    - 18-24
    - 25-34
    - 35-44
    - 45-54
    - 55-64
    - 65-74
    - 75-84
    - 85 or older
96. What is your gender?
    - Female
    - Male
97. What is your highest completed level of education?
    - Less than high school diploma
    - High school diploma or equivalent
    - Some college, no degree
    - Associate degree
    - Bachelor's degree
    - Graduate or professional degree
98. What is your annual household income? If you are unsure of the answer, please give your best estimate.
    - Under $25,000
    - $25,000 - $49,999
    - $50,000 - $74,999
    - $75,000 - $$99,999
    - $100,000 - $149,999
    - $150,000 - $199,999
    - $200,000 - $249,999
    - $250,000 or more

Appendix B: Survey Data Summary

Part 1: Recent intercity travel trips and general travel preferences

Section 1-A: The following questions are about your recent trips.

How many times have you visited one of the following cities in the past twelve months? (Exclude trips where the city was not the primary destination and you only passed through it on the way to another destination)

Variable	n	missing	unique
id	2560	0	2560
password	2560	0	2560
homezip_1_1	2560	0	676
cityfreq_1_1	2560	0	12
cityfreq_1_2	2560	0	12
cityfreq_1_3	2560	0	11
cityfreq_1_4	2560	0	12

Variable	Number of visits in the past year to…		0	1	2	3	4	5	6	7	8	9	10	12 or more
cityfreq_1_1	*Boston*	Frequency	703	552	361	221	167	110	112	24	43	3	48	216
		%	27	22	14	9	7	4	4	1	2	0	2	8
cityfreq_1_2	*New York*	Frequency	1768	484	148	58	39	15	14	6	9	1	7	11
		%	69	19	6	2	2	1	1	0	0	0	0	0
cityfreq_1_3	*Philadelphia*	Frequency	2291	181	47	14	11	8	3	1	2	0	1	1
		%	89	7	2	1	0	0	0	0	0	0	0	0
cityfreq_1_4	*Washington, D.C.*	Frequency	2102	328	81	21	13	5	3	1	1	1	1	3
		%	82	13	3	1	1	0	0	0	0	0	0	0

What mode(s) of transportation have you used for your trip(s) to each city in the past twelve months? Please select all that apply.

						Not Selected		Selected	
Variable	Destination	Mode	n	missing	unique	Frequency	Percent	Frequency	Percent
mode_1_1	*Boston*	*Personal Auto/Car*	1857	703	2	168	9%	1689	91%
mode_1_2		*Rental or borrowed car*	1857	703	2	1795	97%	62	3%
mode_1_3		*Intercity bus*	1857	703	2	1671	90%	186	10%
mode_1_4		*Intercity rail*	1857	703	2	1655	89%	202	11%
mode_1_5		*Airplane*	1857	703	2	1802	97%	55	3%
mode_1_6		*Other*	1857	703	2	1805	97%	52	3%
mode_2_1	*New York*	*Personal Auto/Car*	792	1768	2	341	43%	451	57%
mode_2_2		*Rental or borrowed car*	792	1768	2	745	94%	47	6%
mode_2_3		*Intercity bus*	792	1768	2	672	85%	120	15%
mode_2_4		*Intercity rail*	792	1768	2	588	74%	204	26%
mode_2_5		*Airplane*	792	1768	2	663	84%	129	16%
mode_2_6		*Other*	792	1768	2	762	96%	30	4%
mode_3_1	*Philadelphia*	*Personal Auto/Car*	269	2291	2	137	51%	132	49%
mode_3_2		*Rental or borrowed car*	269	2291	2	235	87%	34	13%
mode_3_3		*Intercity bus*	269	2291	2	258	96%	11	4%
mode_3_4		*Intercity rail*	269	2291	2	246	91%	23	9%
mode_3_5		*Airplane*	269	2291	2	166	62%	103	38%
mode_3_6		*Other*	269	2291	2	262	97%	7	3%
mode_4_1	*Washington, D.C.*	*Personal Auto/Car*	458	2102	2	277	60%	181	40%
mode_4_2		*Rental or borrowed car*	458	2102	2	414	90%	44	10%
mode_4_3		*Intercity bus*	458	2102	2	440	96%	18	4%
mode_4_4		*Intercity rail*	458	2102	2	408	89%	50	11%
mode_4_5		*Airplane*	458	2102	2	220	48%	238	52%
mode_4_6		*Other*	458	2102	2	449	98%	9	2%

[If intercity bus or intercity rail selected for ANY city] How do you usually get information about routes and schedules for bus or rail trips? Please select all that apply.

					Not Selected		Selected	
Variable	Source of route/schedule information	n	missing	unique	Frequency	Percent	Frequency	Percent
information_1_1	*Use pamphlets or other printed material*	618	1942	2	531	86%	87	14%
information_1_2	*Ask a friend or family member*	618	1942	2	555	90%	63	10%
information_1_3	*Visit the station*	618	1942	2	545	88%	73	12%
information_1_4	*Call the bus or rail company*	618	1942	2	563	91%	55	9%
information_1_5	*Search the internet*	618	1942	2	68	11%	550	89%
information_1_6	*Use smart phone or tablet apps*	618	1942	2	451	73%	167	27%
information_1_7	*Other, please specify*	618	1942	2	612	99%	6	1%
information_1_7_x	*Other, Specified*	2560	0	10				

[if # of cities visited > 1] Which city did you visit most recently?

Variable	n	missing	unique	City	Frequency	Percent
city_1_1	1996	564	4	Boston	1567	79%
				New York City	258	13%
				Philadelphia	46	2%
				Washington DC	125	6%

[Skip if frequency to # cities visited = 1] What mode(s) of transportation did you use for your MOST RECENT trip to <recent city>? Please select all that apply.

Variable	Mode(s) used on most recent trip	n	missing	unique	Not Selected		Selected	
					Frequency	Percent	Frequency	Percent
modes_1_1	*Personal auto/car*	1996	564	2	383	19%	1613	81%
modes_1_2	*Rental car or borrowed car*	1996	564	2	1937	97%	59	3%
modes_1_3	*Intercity bus*	1996	564	2	1865	93%	131	7%
modes_1_4	*Intercity rail*	1996	564	2	1802	90%	194	10%
modes_1_5	*Airplane*	1996	564	2	1881	94%	115	6%
modes_1_6	*Other, please specify*	1996	564	2	1957	98%	39	2%
modes_1_6_x	*Other, specified*	2560	0	44				

What was the purpose of your most recent trip to <recent city>? Please select all that apply.

Variable	Purpose of most recent trip	n	missing	unique	Not Selected Frequency	Not Selected Percent	Selected Frequency	Selected Percent
purpose_1_1	*Leisure/vacation*	1996	564	2	797	40%	1199	60%
purpose_1_2	*Visit friends*	1996	564	2	1609	81%	387	19%
purpose_1_3	*Business*	1996	564	2	1726	86%	270	14%
purpose_1_4	*Family event*	1996	564	2	1685	84%	311	16%
purpose_1_5	*Other, please specify*	1996	564	2	1904	95%	92	5%
purpose_1_5_x	*Other, specified*	2560	0	179				

How many people travelled with you on your most recent trip to <recent city>? (Exclude those who did not make at least part of the journey with you)

Variable	Number of...	n	missing	unique		0	1	2	3	4	5	6	7	8	10	11 or more
occ_1_1	***Other adults (18 and over) on most recent trip***	1996	564	11	**Frequency**	322	754	636	154	79	18	11	4	3	4	11
					%	16	38	32	8	4	1	1	0	0	0	1
occ_1_2	***Children (under 18) on most recent trip***	1996	564	10	**Frequency**	1624	195	116	38	9	3	1	0	1	1	8
					%	81	10	6	2	0	0	0	0	0	0	0

[if bus, rail, or plane trip] How did you plan this trip and book your tickets? Please select all that apply.

Variable	Planning of most recent trip:	n	missing	unique	Not Selected Frequency	Not Selected Percent	Selected Frequency	Selected Percent
plan_1_1	*Went to the airline, bus, or train website*	436	2124	2	166	38%	270	62%
plan_1_2	*Went to a travel website (e.g., Expedia.com, Kayak.com)*	436	2124	2	383	88%	53	12%
plan_1_3	*Called the airline, bus company, or train line*	436	2124	2	420	96%	16	4%
plan_1_4	*Through a travel agency*	436	2124	2	415	95%	21	5%
plan_1_5	*A friend or family member booked it for me*	436	2124	2	406	93%	30	7%
plan_1_6	*Other, please specify*	436	2124	2	396	91%	40	9%
plan_1_6_x	*OTHER Specified*	2560	0	80				

How many nights did you stay for your most recent trip to < recent city >?

Variable	n	missing	unique		Number of Nights on Recent Trip: 0	1	2	3	4	5	6	7 or more
nights_1_1	1996	564	8	**Frequency**	1016	424	305	125	64	29	10	23
				%	51	21	15	6	3	1	1	1

How many registered vehicles (in working order) are available to your household? Please include all cars, pickup trucks, minivans, and motorcycles/scooters to which your household has regular access, whether owned, leased, or a company vehicle.

Variable	n	missing	unique		Number of Vehicles in Household 0	1	2	3	4	5	6	7	10 or more
vehicles_1_1	2560	0	9	**Frequency**	30	585	1244	495	150	30	18	7	1
				%	1	23	49	19	6	1	1	0	0

Do you have a driver's license?

Variable	n	missing	unique		Frequency	Percent
license_1_1	2560	0	2	**Yes**	2522	99%
				No	38	1%

How many people live in your household? How many of you are licensed drivers?

Variable	Description	n	missing	unique		0	1	2	3	4	5	6	7	8
household_1_1	*adults in HH*	2560	0	7	**Frequency**	0	424	1643	334	130	23	5	1	0
	(18 and over)				**%**	0	17	64	13	5	1	0	0	0
household_1_2	*adults in HH:*	2560	0	7	**Frequency**	23	475	1647	298	99	16	2	0	0
	Licensed drivers				**%**	1	19	64	12	4	1	0	0	0
household_2_1	*children in HH*	2560	0	8	**Frequency**	1986	283	208	66	12	2	2	0	1
	(under 18)				**%**	78	11	8	3	0	0	0	0	0
household_2_2	*children in HH:*	2560	0	5	**Frequency**	2448	95	13	3	1	0	0	0	0
	Licensed drivers				**%**	96	4	1	0	0	0	0	0	0

How do you access the internet? Please select all that apply.

					Not Selected		Selected	
Variable	Internet access	n	missing	unique	Frequency	Percent	Frequency	Percent
internet_1_1	*Internet service at home*	2560	0	2	59	2%	2501	98%
internet_1_2	*Internet service at school*	2560	0	2	2420	95%	140	5%
internet_1_3	*Internet service at work*	2560	0	2	1305	51%	1255	49%
internet_1_4	*Public internet service*	2560	0	2	2199	86%	361	14%
internet_1_5	*Mobile device with a cellular data plan*	2560	0	2	1157	45%	1403	55%
internet_1_6	*Other, please specify*	2560	0	2	2546	99%	14	1%
internet_1_6_x	*Other, specified*	2560	0	16				

Part 2: Travel preferences

Variable	n	missing	unique	Preference		Completely Agree			Neutral			Completely Disagree
						1	2	3	4	5	6	7
travelpreferences_1_1	2560	0	7	I feel I am less dependent on cars than my parents are/were.	**Frequency**	90	126	172	502	444	527	699
					%	4	5	7	20	17	21	27
travelpreferences_2_1	2560	0	7	I need to drive my car to get where I need to go.	**Frequency**	1022	623	367	279	104	82	83
					%	40	24	14	11	4	3	3
travelpreferences_3_1	2560	0	7	I love the freedom and independence I get from owning one or more cars.	**Frequency**	1138	653	348	256	61	54	50
					%	44	26	14	10	2	2	2
travelpreferences_4_1	2560	0	7	It would be hard for me to reduce my driving mileage.	**Frequency**	567	670	428	454	260	102	79
					%	22	26	17	18	10	4	3
travelpreferences_5_1	2560	0	7	For me to be able to leave the driving to someone else(e.g., a bus driver) would be desirable.	**Frequency**	241	392	514	664	288	246	215
					%	9	15	20	26	11	10	8
travelpreferences_6_1	2560	0	7	It would be desirable for my household to be able to have fewer cars.	**Frequency**	122	175	217	643	308	395	700
					%	5	7	8	25	12	15	27
travelpreferences_7_1	2560	0	7	Being able to freely perform tasks, including using a laptop, tablet, or smartphone is an important reason for me to choose bus or train travel.	**Frequency**	271	346	431	678	247	268	319
					%	11	14	17	26	10	10	12
travelpreferences_8_1	2560	0	7	Having reliable WiFi internet access while I travel on a bus or train is important to me.	**Frequency**	515	521	485	494	170	169	206
					%	20	20	19	19	7	7	8
travelpreferences_9_1	2560	0	7	When taking a bus or train, being able to plan my trip and buy tickets online is important to me.	**Frequency**	788	646	438	446	105	70	67
					%	31	25	17	17	4	3	3
travelpreferences_10_1	2560	0	7	It would be important to me to receive email or text message updates about my bus or train trip.	**Frequency**	443	534	533	557	160	139	194
					%	17	21	21	22	6	5	8
travelpreferences_11_1	2560	0	7	I find tablet or smartphone apps for travel and trip planning to be helpful.	**Frequency**	446	496	443	637	166	135	237
					%	17	19	17	25	6	5	9
travelpreferences_12_1	2560	0	7	When the government tries to improve things, it never works.	**Frequency**	377	358	357	772	291	240	165
					%	15	14	14	30	11	9	6
travelpreferences_13_1	2560	0	7	If everyone works together, we could improve the environment and future for the earth.	**Frequency**	1020	663	421	284	72	52	48
					%	40	26	16	11	3	2	2
travelpreferences_14_1	2560	0	7	People like me take the bus or the train.	**Frequency**	120	183	287	768	344	375	483
					%	5	7	11	30	13	15	19
travelpreferences_15_1	2560	0	7	I would be willing to pay more when I travel if it would help the environment.	**Frequency**	119	266	543	890	335	213	194
					%	5	10	21	35	13	8	8
travelpreferences_16_1	2560	0	7	I tend to use the fastest form of transportation, regardless of cost.	**Frequency**	111	246	379	649	487	377	311
					%	4	10	15	25	19	15	12
travelpreferences_17_1	2560	0	7	For me, the whole idea of being on a bus or train with other people I do not know seems uncomfortable.	**Frequency**	107	190	302	593	433	493	442
					%	4	7	12	23	17	19	17
travelpreferences_18_1	2560	0	7	I enjoy being out and about and observing people.	**Frequency**	625	788	590	359	104	65	29
					%	24	31	23	14	4	3	1
travelpreferences_19_1	2560	0	7	I don't mind traveling with people I do not know.	**Frequency**	292	439	503	595	296	252	183
					%	11	17	20	23	12	10	7
travelpreferences_20_1	2560	0	7	Having my privacy is important to me when I travel.	**Frequency**	570	682	585	484	135	76	28
					%	22	27	23	19	5	3	1
travelpreferences_21_1	2560	0	7	When I choose a home, I value having adequate space for parking two or more cars.	**Frequency**	954	663	368	287	106	89	93
					%	37	26	14	11	4	3	4
travelpreferences_22_1	2560	0	7	When I choose a neighborhood to live in, I like to be able to walk to a commercial or village center.	**Frequency**	298	331	456	770	280	233	192
					%	12	13	18	30	11	9	8
travelpreferences_23_1	2560	0	7	Living in a multiple family building (e.g., apartment, condo) wouldn't give me enough privacy.	**Frequency**	759	456	337	387	218	190	213
					%	30	18	13	15	9	7	8
travelpreferences_24_1	2560	0	7	I like living in a neighborhood where there is a lot going on.	**Frequency**	185	275	421	721	405	314	239
					%	7	11	16	28	16	12	9
travelpreferences_25_1	2560	0	7	I am confident that if I want to, I can do things that I have never done before.	**Frequency**	859	796	494	282	67	45	17
					%	34	31	19	11	3	2	1
travelpreferences_26_1	2560	0	7	I worry about crime or other disturbing behavior on buses and trains, or while walking in and around the stops/stations.	**Frequency**	209	351	615	571	344	297	173
					%	8	14	24	22	13	12	7
travelpreferences_27_1	2560	0	7	It is important to me to control the radio and the air conditioning in the car.	**Frequency**	464	516	582	637	166	115	80
					%	18	20	23	25	6	4	3
travelpreferences_28_1	2560	0	7	I feel really stressed when driving for a long time in congestion in and around big cities.	**Frequency**	546	543	559	342	230	193	147
					%	21	21	22	13	9	8	6
travelpreferences_29_1	2560	0	7	I prefer to use the most comfortable transportation mode regardless of cost or time.	**Frequency**	158	270	486	769	452	254	171
					%	6	11	19	30	18	10	7
travelpreferences_30_1	2560	0	7	Having a low-stress trip is more important than reaching my destination quickly.	**Frequency**	414	605	611	579	221	95	35
					%	16	24	24	23	9	4	1
travelpreferences_31_1	2560	0	7	I get very annoyed being stuck behind a slow driver.	**Frequency**	553	595	704	413	173	76	46
					%	22	23	28	16	7	3	2
travelpreferences_32_1	2560	0	7	I am usually in a hurry when I make a trip.	**Frequency**	108	228	488	768	457	313	198
					%	4	9	19	30	18	12	8
travelpreferences_33_1	2560	0	7	With my schedule, minimizing time spent traveling is very important to me.	**Frequency**	310	453	547	730	227	168	125
					%	12	18	21	29	9	7	5
travelpreferences_34_1	2560	0	7	I would use the bus or train more often if it were cheaper to ride.	**Frequency**	297	329	401	747	279	228	279
					%	12	13	16	29	11	9	11
travelpreferences_35_1	2560	0	7	Rather than owning a car, I would prefer to borrow, share, or rent a car just for when I need it.	**Frequency**	59	68	116	195	253	457	1412
					%	2	3	5	8	10	18	55

Part 3: An imaginary situation

Knowing what you know right now, what mode(s) of transportation do you think are AVAILABLE to you for this trip to NYC? Please select all that apply.

Variable	Imaginary trip - Modes available:	n	missing	unique	Not Selected Count	Not Selected Percent	Selected Count	Selected Percent
modesavailable_1_1	*Personal auto/car*	2560	0	2	461	18%	2099	82%
modesavailable_1_2	*Rental car or borrowed car*	2560	0	2	973	38%	1587	62%
modesavailable_1_3	*Intercity bus*	2560	0	2	719	28%	1841	72%
modesavailable_1_4	*Intercity rail*	2560	0	2	590	23%	1970	77%
modesavailable_1_5	*Other, please specify*	2560	0	2	2530	99%	30	1%
modesavailable_1_5_x	*Other, specified*	2560	0	51				

How likely are you to choose to take a bus or train for a trip like this to NYC next month?

Variable	n	missing	unique		Definitely	Very likely	Likely	Neutral	Unlikely	Very unlikely	Definitely not
busortrain_1_1	2560	0	7	**Frequency**	340	688	528	328	309	206	161
				%	13	27	21	13	12	8	6

If you learned there would be no WiFi, and no electrical outlet on the bus or train for this trip, would that make to you less likely to choose a bus or train for this trip?

Variable	n	missing	unique		Much less likely	Somewhat less likely	No change	Not applicable to me
wifi_1_1	2560	0	4	**Frequency**	337	632	1429	162
				Percent	13%	25%	56%	6%

Control Group and Test Group

Variable	n	missing	unique		Control - Did not see Intercity Travel Website	Test - Saw Intercity Travel Website
group	2560	0	2	**Frequency**	1278	1282
				Percent	50%	50%

Variable	n	missing	unique	Statement		Completely Agree 1	2	3	Neutral 4	5	6	Completely Disagree 7
websitestatements_1_1	1282	1278	7	There are more options than what I expected to travel to NYC by bus and train.	**Frequency**	357	289	244	234	64	46	48
					%	28	23	19	18	5	4	4
websitestatements_2_1	1282	1278	7	After seeing the bus and train options for traveling to NYC, I just don't think there's a good way for me to get there by either bus or train.	**Frequency**	66	66	81	236	205	240	388
					%	5	5	6	18	16	19	30
websitestatements_3_1	1282	1278	7	Having information like this on my smartphone or computer might make it easier for me to understand the kinds of bus and train services available to me.	**Frequency**	349	337	246	216	47	42	45
					%	27	26	19	17	4	3	4
websitestatements_4_1	1282	1278	7	Having so many potential travel options by bus and train is confusing.	**Frequency**	104	127	211	245	191	201	203
					%	8	10	16	19	15	16	16

Variable	Statement	n	missing	unique		1 Completely Agree	2	3	4 Neutral	5	6	7 Completely Disagree
tripstatements_1_1	When I drive long distances (like from my home area to NYC), I can get tired and stressed.	2560	0	7	**Frequency**	389	438	597	439	304	216	177
					%	15	17	23	17	12	8	7
tripstatements_2_1	I worry about the difficulty in finding a parking space at a reasonable cost when I get to NYC.	2560	0	7	**Frequency**	1229	618	337	208	67	43	58
					%	48	24	13	8	3	2	2
tripstatements_3_1	I am concerned that the schedule of the bus or train only lets me travel a few times per day, and I need to be flexible.	2560	0	7	**Frequency**	174	368	576	640	377	252	173
					%	7	14	22	25	15	10	7
tripstatements_4_1	I could deal with the limited schedules offered by a bus or train for this trip from my home to NYC.	2560	0	7	**Frequency**	481	605	651	455	199	89	80
					%	19	24	25	18	8	3	3
tripstatements_5_1	I like the idea that I might see and meet new people on a bus or train to NYC.	2560	0	7	**Frequency**	145	219	402	917	357	278	242
					%	6	9	16	36	14	11	9
tripstatements_6_1	I don't like the idea of riding with a lot of people that I don't know on a bus or train.	2560	0	7	**Frequency**	129	158	336	670	440	453	374
					%	5	6	13	26	17	18	15
tripstatements_7_1	If I took a bus or train to NYC, I might have to be with people whose behavior I find unpleasant.	2560	0	7	**Frequency**	239	343	740	694	254	181	109
					%	9	13	29	27	10	7	4
tripstatements_8_1	I could be with other people who share my values when I take a bus or train on a trip like this.	2560	0	7	**Frequency**	169	258	431	1225	212	133	132
					%	7	10	17	48	8	5	5
tripstatements_9_1	I think that taking a BUS to NYC would take a lot longer than driving.	2560	0	7	**Frequency**	326	459	528	492	366	215	174
					%	13	18	21	19	14	8	7
tripstatements_10_1	I think that taking a TRAIN to NYC would take a lot longer than driving.	2560	0	7	**Frequency**	227	330	442	543	436	335	247
					%	9	13	17	21	17	13	10
tripstatements_11_1	Without thinking about it much, I would guess that the cost of taking the trip by BUS would be less than the cost of the car trip (including gas, tolls, and parking).	2560	0	7	**Frequency**	522	680	617	444	173	75	49
					%	20	27	24	17	7	3	2
tripstatements_12_1	Without thinking about it much, I would guess that the cost of taking the trip by TRAIN would be less than the cost of the car trip (including gas, tolls, and parking.)	2560	0	7	**Frequency**	333	522	559	538	316	163	129
					%	13	20	22	21	12	6	5
tripstatements_13_1	It would be really important to me to minimize costs when I plan this trip to NYC next month.	2560	0	7	**Frequency**	492	540	638	546	173	97	74
					%	19	21	25	21	7	4	3
tripstatements_14_1	I really want to minimize the time I spend on the trip to NYC, even if that means more stress or higher costs.	2560	0	7	**Frequency**	109	193	353	778	553	316	258
					%	4	8	14	30	22	12	10
tripstatements_15_1	Being able to use my laptop, tablet, or smartphone when traveling makes me more interested in taking a bus or train to NYC.	2560	0	7	**Frequency**	438	534	545	514	162	128	239
					%	17	21	21	20	6	5	9
tripstatements_16_1	I am the kind of person who would take my own car to NYC.	2560	0	7	**Frequency**	384	312	367	452	317	275	453
					%	15	12	14	18	12	11	18
tripstatements_17_1	Most people whose opinions I value would approve of my taking this trip by bus or train.	2560	0	7	**Frequency**	492	517	514	793	113	62	69
					%	19	20	20	31	4	2	3
tripstatements_18_1	My family would think that I should take this kind of trip by car or plane.	2560	0	7	**Frequency**	191	263	281	831	313	307	374
					%	7	10	11	32	12	12	15
tripstatements_19_1	My colleagues would likely think that it is strange not to go by a car or plane to NYC.	2560	0	7	**Frequency**	123	179	258	842	368	373	417
					%	5	7	10	33	14	15	16
tripstatements_20_1	When my friends go to NYC, they always take a bus or train.	2560	0	7	**Frequency**	137	231	336	1003	304	261	288
					%	5	9	13	39	12	10	11
tripstatements_21_1	When my family members go to NYC, they always take a bus or train.	2560	0	7	**Frequency**	107	191	277	887	306	313	479
					%	4	7	11	35	12	12	19
tripstatements_22_1	It might be unsafe to make this trip by bus or train.	2560	0	7	**Frequency**	69	118	298	622	457	520	476
					%	3	5	12	24	18	20	19
tripstatements_23_1	The experience at the NYC bus or train station would be so unpleasant that I would try to avoid it.	2560	0	7	**Frequency**	88	113	271	638	531	486	433
					%	3	4	11	25	21	19	17
tripstatements_24_1	It would be easy for me to get the schedules for a bus or train between here and NYC, and I would understand them.	2560	0	7	**Frequency**	854	636	521	336	117	56	40
					%	33	25	20	13	5	2	2
tripstatements_25_1	I don't know all the things I NEED to do to make this trip work by bus or train.	2560	0	7	**Frequency**	154	276	415	584	344	367	420
					%	6	11	16	23	13	14	16
tripstatements_26_1	I like the idea of taking a bus or train instead of driving for this trip to NYC.	2560	0	7	**Frequency**	756	551	468	370	186	110	119
					%	30	22	18	14	7	4	5
tripstatements_27_1	I think that the most RATIONAL choice would be to take a bus or train instead of a car.	2560	0	7	**Frequency**	629	579	474	443	184	117	134
					%	25	23	19	17	7	5	5
tripstatements_28_1	I think that the most PLEASURABLE choice would be to take a bus or train instead of a car.	2560	0	7	**Frequency**	539	530	480	491	231	145	144
					%	21	21	19	19	9	6	6
tripstatements_29_1	I think that the most STRESSFUL choice would be to take a bus or train instead of a car.	2560	0	7	**Frequency**	132	126	235	464	471	517	615
					%	5	5	9	18	18	20	24
tripstatements_30_1	All other things being equal, if a bus was cheaper, but less reliable than a train, I would choose to take a bus.	2560	0	7	**Frequency**	105	188	359	596	589	375	348
					%	4	7	14	23	23	15	14
tripstatements_31_1	I am confident that if I wanted to, I could take a bus or train for such a trip to NYC next month.	2560	0	7	**Frequency**	1089	662	428	238	60	35	48
					%	43	26	17	9	2	1	2
tripstatements_32_1	I would make an effort to choose a bus or train for such a trip to NYC next month.	2560	0	7	**Frequency**	604	571	519	446	180	100	140
					%	24	22	20	17	7	4	5
tripstatements_33_1	For me to take a bus or train for such a trip to NYC the next month would be impossible.	2560	0	7	**Frequency**	109	100	129	459	414	493	856
					%	4	4	5	18	16	19	33
tripstatements_34_1	In this imaginary situation, I would plan to take a bus or train for this trip to NYC next month.	2560	0	7	**Frequency**	706	572	459	387	163	113	160
					%	28	22	18	15	6	4	6
tripstatements_35_1	I would trust the person who invited me to NYC to recommend how I should travel.	2560	0	7	**Frequency**	237	433	633	770	198	133	156
					%	9	17	25	30	8	5	6

Part 4: Other information about you and your household.

Given what you know about bus and train services to NYC, how likely are you to choose a bus or train for a trip to NYC next month (like the one described in the imaginary situation)?

We noticed that you are now <more/less> likely to take the train or bus to NYC. Please tell us why you have changed your mind.

Thank you for sharing your opinions about the imaginary trip to NYC. In your real life, how seriously would you consider taking a bus or train to NYC?

Variable	n	missing	unique		Definitely	Very likely	Likely	Neutral	Unlikely	Very unlikely	Definitely not
choosebusortrain_1_1	2560	0	7	**Frequency**	370	740	495	361	258	188	148
				%	14	29	19	14	10	7	6
whylikelihoodchanged_1_1	2560	0	533								
likelybustrain_1_1	2560	0	7	**Frequency**	638	590	654	229	194	150	105
				%	25	23	26	9	8	6	4

Which of the following do you own? Please select all that apply.

					Not Selected		Selected	
Variable	**Owns:**	**n**	**missing**	**unique**	**Frequency**	**Percent**	**Frequency**	**Percent**
own_1_1	*desktop computer*	2560	0	2	963	38%	1597	62%
own_1_2	*laptop*	2560	0	2	439	17%	2121	83%
own_1_3	*smartphone*	2560	0	2	868	34%	1692	66%
own_1_4	*tablet*	2560	0	2	1150	45%	1410	55%
own_1_5	*standalone GPS navigation device*	2560	0	2	1276	50%	1284	50%
own_1_6	*none of the listed devices*	2560	0	2	2552	100%	8	0%

Variable	n	missing	unique	Categories	Frequency	%
age_1_1	2560	0	8	18-24	65	3
				25-34	359	14
				35-44	376	15
				45-54	563	22
				55-64	695	27
				65-74	431	17
				75-84	64	2
				85 or older	7	0
gender_1_1	2560	0	2	Female	1491	58%
				Male	1069	42%
education_1_1	2560	0	6	Less than high school diploma	15	1%
				High school diploma or equivalent	250	10%
				Some college no degree	455	18%
				Associate degree	241	9%
				Bachelor's degree	863	34%
				Graduate or professional degree	736	29%
income_1_1	2534	26	8	Under $25,000	155	6%
				$25,000 - $49,999	422	17%
				$50,000 - $74,999	557	22%
				$75,000 - $99,999	517	20%
				$100,000 - $149,999	579	23%
				$150,000 - $199,999	195	8%
				$200,000 - $249,999	58	2%
				$250,000 or more	51	2%
state	2560	0	4	Massachusetts	937	37%
				New Hampshire	727	28%
				Maine	521	20%
				Vermont	375	15%