THE UNIVERSITY OF TEXAS AT AUSTIN
CENTER FOR TRANSPORTATION RESEARCH

TECHNICAL REPORT 0-7081-1
TxDOT PROJECT NUMBER 0-7081

Understanding the Impact of Autonomous Vehicles on Long-Distance Passenger and Freight Travel in Texas: Final Report

Kara Kockelman, Ph.D.
Yantao Huang, Ph.D.
Fatemeh Fakhrmoosavi, Ph.D.
Kenneth Perrine
Priyanka Paithankar
Jason Hawkins, Ph.D.
Natalia Zuniga-Garcia, Ph.D
Maithreyi Vellimana

August 2022
Published August 2023

https://library.ctr.utexas.edu/ctr-publications/0-7081-1.pdf

Technical Report Documentation Page

1. Report No. FHWA/TX-23/0-7081-1	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Understanding the Impact of Autonomous Vehicles on Long-Distance Passenger and Freight Travel in Texas: Final Report		5. Report Date Submitted: August 2022 6. Performing Organization Code
7. Author(s) Kara Kockelman, Ph.D. http://orcid.org/0000-0003-4763-1304; Yantao Huang, Ph.D.; Fatemeh Fakhrmoosavi, Ph.D.; Kenneth Perrine; Priyanka Paithankar; Jason Hawkins, Ph.D.; Natalia Zuniga-Garcia, Ph.D.; Maithreyi Vellimana		8. Performing Organization Report No. 0-7081-1
9. Performing Organization Name and Address Center for Transportation Research The University of Texas at Austin 3925 W. Braker Lane, 4th Floor Austin, TX 78759		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 0-7081
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Division 125 E. 11th Street Austin, TX 78701		13. Type of Report and Period Covered Technical Report September 2020 – August 2022 14. Sponsoring Agency Code

15. Supplementary Notes

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

16. Abstract

In efforts to predict the long-distance travel impacts (for passengers and freight) of self-driving cars and trucks across Texas and the US, researchers estimated models for long-distance domestic passenger and freight trips before and after the introduction of autonomous vehicles (AVs) and applied the passenger models to a 10% synthetic US population (12.1M households and 28.1M individuals across 73,056 census tracts). To generate disaggregated passenger trips, travel demand models—including trip frequency, season, purpose, party size, mode choice, and destination choice models—and a vehicle ownership model were estimated. Different datasets, including a 2021 long-distance passenger AV survey designed in this study, the 2016/17 National Household Travel Survey, and EPA Smart Location and FHWA rJourney datasets, were used for model estimation. Assuming a $3,500 technology cost premium (e.g., in year 2040), total person-miles traveled per capita for existing long-distance trips are estimated to rise 35% (from 280 to 379 miles per month).

For freight-travel impacts, a four-step travel demand modeling process was used for trip generation and distribution, mode choice, and traffic assignment across human-driven trucks, automated trucks (ATrucks), rail, and intermodal rail. When ATruck shipping costs drop to half that of human-driven trucks, truck mode share is predicted to increase by 4.2% (from 57.0% to 61.2%), with a 6.0% increase in ton-miles transported by truck. Such cost reductions are expected to increase Texas trucks' mode splits by over 10% on commodities like food, paper, and primary metal.

17. Key Words Long-Distance Travel, Autonomous Vehicles, Travel Demand, Mode Choice, Destination Choice, Freight.	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Alexandria, Virginia 22312; www.ntis.gov.

19. Security Classif. (of report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	264	

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

THE UNIVERSITY OF TEXAS AT AUSTIN
CENTER FOR TRANSPORTATION RESEARCH

Understanding the Impact of Autonomous Vehicles on Long-Distance Passenger and Freight Travel in Texas: Final Report

Kara Kockelman, Ph.D.
Kenneth A. Perrine
Yantao Huang, Ph.D.
Fatemeh Fakhrmoosavi, Ph.D.
Priyanka Paithankar
Jason Hawkins, Ph.D.
Natalia Zuniga-Garcia, Ph.D.
Maithreyi Vellimana

CTR Technical Report:	0-7081-1
Report Date:	Submitted: August 2022
Project:	0-7081
Project Title:	Understanding the Impact of Autonomous Vehicles on Long-Distance Travel Mode and Destination Choice in Texas
Sponsoring Agency:	Texas Department of Transportation
Performing Agency:	Center for Transportation Research at The University of Texas at Austin

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

Center for Transportation Research
The University of Texas at Austin
3925 W. Braker Lane, 4th floor
Austin, TX 78759

http://ctr.utexas.edu/

Disclaimers

Author's Disclaimer: The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Federal Highway Administration or the Texas Department of Transportation (TxDOT). This report does not constitute a standard, specification, or regulation.

Patent Disclaimer: There was no invention or discovery conceived or first actually reduced to practice in the course of or under this contract, including any art, method, process, machine manufacture, design or composition of matter, or any new useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States of America or any foreign country.

Engineering Disclaimer

NOT INTENDED FOR CONSTRUCTION, BIDDING, OR PERMIT PURPOSES.

Project Engineer: Kara Kockelman
Professional Engineer License State and Number: Texas No. 93443
P.E. Designation: Research Supervisor

Acknowledgments

The authors express appreciation to key supporters within TxDOT, including project manager Martin Dassi and project monitoring committee members Geena Maskey, James Kuhr, Ryan Granger, Ash Duong, and Sondra Johnson. The authors also thank Sarah McGavick from the Center for Transportation Research for her editorial support.

Products

0-7081-P1: Executive Guide: Investigating the Impacts of AVs on Long-Distance Passenger and Freight Travel

0-7081-P2: Workshop: Self-Driving Vehicles' Impacts on Long-Distance Passenger and Freight Travel

Table of Contents

List of Tables

List of Figures

Executive Summary

As fully automated or "autonomous" vehicles (AVs) become increasingly available in the coming years, their travel, trade, emissions, cost, and other implications need to be anticipated. Prior studies predict AVs dominating US passenger travel between 100 and 500 miles (one-way) and freight travel of over 300 ton-miles. With network vehicle-miles traveled (VMT) predicted to rise by over 25% (due to many Texas air travelers shifting to shared AVs, others extending their current ground-trip distances, and still others making more trips), this project gathered new data to simulate changes in freight and passenger flows across Texas and the US for all competing modes in the future. Integrating related trends like AV technology price change, the research team anticipated the impacts of AVs on passenger mode and destination choice for long-distance trips over 75 miles as well as the impacts of automated trucks (ATrucks) on freight travel.

The team designed and distributed a long-distance passenger-travel survey with almost 70 questions (as described in Chapter 2, Chapter 3, and Appendix A) tackling aspects of long-distance travel, AVs, and shared autonomous vehicle (SAV) use, as well as the COVID-19 pandemic's effects on long-distance travel. The survey included a mix of revealed and stated preference questions for reasonably recent long-distance trips and future scenarios. The survey responses from 1,004 individuals (45% Texans and 55% other US residents) indicated that Texans were about 40% more willing to travel long-distance using AVs than other US respondents in scenarios where the use of AVs imposes a 0–50% increase in travel time. Texans reacted similarly to other US respondents in scenarios where long-distance AV trips impose an added cost (60% were "unlikely" to travel in an AV that incurred a 50% cost increase, and 37% were "unlikely" to choose an AV when it would increase costs by 25%). More than 40% of Texans were "absolutely" or "more likely" to use AVs for a long-distance trip than before the pandemic if it didn't impact costs. If AV travel decreased long-distance travel costs by half, the share of Texans who would be at least "more likely" to travel in an AV went up to almost 60%. For 60% of Texans, using an AV would not change their destination choice, but about 20% would decide to visit a place that's farther away. Regarding increasing travel party size, 32% of Texans would do so on a long-distance AV trip, while only 22% of others surveyed would. People who would prefer an AV for long-distance travel most related this choice to added safety, followed by reliability. Safety was also the priority for respondents who opted not to use AVs for long-distance travel, and they tended to consider faulty software to be a potential issue. Texans were slightly more willing to use an SAV in a pandemic scenario, compared to respondents across the US.

The team used different datasets (Chapter 4), including the long-distance AV survey, 2016/17 National Household Travel Survey (NHTS), EPA Smart Location dataset, Air Passenger Origin and Destination Survey (DB1B), the rJourney dataset from the Federal Highway Administration (FHWA), and a 2017 AV fleet survey for passenger model estimation. The fourth and fifth versions of the Bureau of Transportation Statistics' and Federal Highway Administration's Freight Analysis Framework (FAF) and TxDOT's Statewide Analysis Model (SAM) were used for freight travel model estimation and application. To simulate US long-distance passenger travel, the research team used PopGen 2.0 software to synthesize household and person data for 28.1 million persons (in 12.1 million households) across 2,351 Public Use Microdata Areas (PUMAs) to mimic the nation's population distribution across 50 states and the District of Columbia consistent with census datasets (using the nation's 73,056 census tracts). The team used these persons and households to estimate the set of travel demand models. These seven travel demand model or equation types are used in the following sequence: household vehicle ownership, trip frequency, travel season, trip purpose, party size, destination choice, and mode choice, in order to generate specific trips for each person and travel party within the US (no international trips were modeled). The research team's key estimations from these passenger mode applications are as follows:

- The team estimated 0.85 vehicles per capita in 2019, which is consistent with the US census data's vehicles per capita of 0.83 in 2020. 61% of US households are predicted to own at least one AV after their introduction assuming a $3,500 AV technology cost in year 2040.

- 2.003 long-distance trips per month per capita were estimated for the current 10% synthetic population, which matches the NHTS data. Long-distance trips are assumed to rise 15% after AVs are in market.

- Mode splits for long-distance, domestic trips prior to AV access were estimated as 64.10% by private automobile, 30.42% by rental car, and 5.49% by air. After AVs become available for purchase (with a premium cost of $3,500) and SAVs are available with $0.70/mile operation cost, mode splits shift to 31.67% by conventional human-driven vehicle, 23.02% by conventional rental car, 23.54% by AV, 18.24% by SAVs, and 3.53% by air.

- Assuming a $3,500 AV technology cost premium in 2040 in today's dollars, total person-miles traveled (PMT) per capita in long-distance trips is estimated to rise 35% (from 280 to 379 miles per month). For the same AV technology cost premium scenario, vehicle-miles traveled (VMT) in long-distance trips increases from 121 to 152 miles per capita per month as many travelers shift from air to cars and shorter trips.

In terms of freight-travel impacts, this project applied demand models with and without ATrucks (Chapter 6 for US and Chapter 7 for Texas). A four-step travel demand modeling process (with feedback loops, for congestion's effects on mode and destination choices) was used for Texas freight trip generation and distribution, mode choice, and traffic assignment. It was applied to both freight and passenger travel (allowing for AVs and SAVs) to better reflect traffic congestion and travel times. The model simulates just one time of day to recognize that many long-distance trips span many times of day (involving both peak and off-peak traffic levels), and thus different congestion settings. A base case scenario without AV, SAV, and ATruck modes was run first, for comparison to the scenarios that include automation technology, to reflect the cost- and time-saving impacts on passenger and trade flows. For the freight mode choice model, human-driven trucks (HTrucks) and ATruck alternatives are nested under the truck mode after AV introduction to the market, and in competition with carload rail (CL) and intermodal rail (IM). Various parameter settings were tested, and key observations are as follows:

- According to the freight prediction model for the Texas region in year 2045, 1.7 billion tons of goods (or 60% of total tons moved, about 1.3 trillion ton-miles) will be transported across the state by HTrucks and ATrucks (based on assumptions that ATrucks cost 1.5 times as much as HTrucks but save HTrucks' dwelling time), while the rest will be transported by CL (23.5%) and IM (16.5%).

- All modeled industries would witness an increase in the mode share of trucks and decline in CL and IM due to the use of ATrucks. This happened even for the commodities that are oriented toward the use of CL or IM. The increase in truck ton-miles is quite stable across trips of all distances in Texas, at around 5%. Across all commodities transported in ton-miles, trucks experienced a 7.8% increase in tonnage, while CL's and IM's tonnages dropped by 12.6% and 2.3%, respectively.

- Coal truck ton-miles are predicted to rise by 86%, as this commodity shifts away from CL modes (which are expected to still dominate coal transport after ATruck implementation).

- In the long term, if ATruck costs drop to half of HTruck costs, heavy trucking's mode share in tons is predicted to rise by 4.2% (from 57.0% to 61.2%), with a 7.3% increase in tons transported by truck.

This project's primary product is an executive guide, titled *Investigating the Impacts of Automated Vehicles on Long-Distance Passenger and Freight Travel,* for use by practitioners and other interested parties. Chapter 1 of this guide synthesizes relevant prior research and current industry practices. Chapter 2 details the long-distance travel survey's design, and Chapter 3 highlights summary statistics and key results of that survey. Chapter 4 introduces other datasets central to model estimation and application for future-year predictions of passenger and freight travels. Chapter 5 describes estimation and application of long-distance travel demand models for passengers across the US with and without AV mode options, and Chapter 6 describes the models for ATruck impacts on freight choices (across 20 commodity sectors). Chapter 7 applies the Chapter 6 long-distance travel demand models for freight across Texas, with and without ATruck options. Chapter 8 explains the value of this research and provides a cost-benefit analysis. Guide appendices contain other supporting documents for investigating the impacts of AVs and ATrucks on long-distance travel, including the long-distance travel survey instrument (Appendix A), Python scripts for the weighting process of the survey and summary statistics (Appendix B and C), US census regions and divisions (Appendix D), US and Texas top freight flow commodity rankings (Appendix E), and international long-distance trip models (Appendix F).

Literature Synthesis

Before the 1995 American Travel Survey (ATS) data (for trips over 100 miles one-way) became available, travel demand researchers focused mostly on urban mobility and local congestion issues (Aultman-Hall, 2018). Due to rising need and interest, several long-distance (or inter-city) travel surveys emerged in the US, including questions in the 2001, 2009, and 2016/17 National Household Travel Surveys (NHTS) and state-focused long-distance surveys in Ohio, Michigan, California, Colorado, Utah, Vermont, and Alabama. The FHWA's (2015) rJourney long-distance travel demand model incorporated data from the 1995 ATS, the 2001 NHTS, and long-distances surveys in California, Colorado, and Ohio. European survey projects include Axhausen and Youssefzadeh's (1999) Methods for European Surveys of Travel Behaviour (MEST), Frei et al.'s (2010) Knowledge-base for Intermodal Passenger Travel in Europe" (KITE), and Zumkeller et al.'s (2010) Intermodal Linking of Passenger Transport Modes Considering User Needs (INVERMO) project.

Long-distance trips are a key component of the nation's traffic volumes, congestion levels, emissions, crashes, and pavement damage. Long-distance trips are usually defined as one-way trips over 50 miles, but that definition does vary by context, ranging from 75-mile (as in this report) to 100-mile trips, and out-of-town travel or overnight travel (Aultman-Hall et al., 2018; LaMondia et al., 2016a; Bacon and LaMondia, 2016). According to 2016/17 NHTS trip records, 43.3% of US person-miles traveled (PMT) come from one-day one-way trips over 50 miles. These are just 2.5% of all person trips being made in the US but almost half of all PMT (McGuckin, 2018). Although the 2001 NHTS was the last national travel survey in the US to ask about long-distance trips in detail (Hu and Reuscher, 2004), the 2016/17 NHTS (which focused on one-way travel during one survey day) estimates over 1,500 trillion miles of long-distance PMT annually (McGuckin, 2018).

1.1. Literature Review

This section reviews various aspects of the literature related to long-distance travel, including the factors impacting long-distance trip-making, new predictions of inter- and intra-regional travel patterns due to autonomous vehicles (AVs) and fully autonomous trucks (ATrucks) with environmental implications, long-distance tourism, and long-distance travel predictions during the COVID-19 pandemic.

1.1.1. Long-Distance Travel Factors

Given the distances and costs involved, long-distance travel decisions are heavily influenced by travel cost and time considerations, and can include choices of whether and where to stay overnight en route. Long-distance personal travel is more often as compared to daily trip-making (e.g., for school, work, medical appointments, and shopping). And long-distance passenger-vehicle trips tend to involve higher occupancies than intra-urban trips (over 2 persons per long-distance trip, rather than just 1.1 for work trips and 1.3 for other intra-urban trips). Schedule coordination (between members of the same travel party) is also important, and larger vehicles tend to be used compared to those used for intra-urban trip-making (LaMondia et al., 2016b).

Key variables impacting long-distance passenger trips include frequency, distance, mode, destination, household income, traveler age, education level, presence of children in the household, and more (Van Wee et al., 2006; Sandow and Westin, 2010; Collia et al., 2003; Holz-Rau et al., 2014; LaMondia et al., 2016a; Cho, 2013). Specific events and objectives—like professional conferences, vacations, weddings, funerals, sports tournaments, and music concerts (Yang et al., 2016; Burke and Woolcock, 2013; McKercher et al., 2008; Aguilera, 2008)—regularly motivate long-distance travel for many different types of people.

A key factor impacting long-distance business trip choices is employer reimbursement (Schaeffer, 2009; Cai et al., 2011). For those receiving reimbursement, higher-speed modes (like air travel with a ride from a transportation network company at either end) may still capture a high portion of mode share in many markets, especially for business trips over 750 miles (one-way). Frequent long-distance business travelers tend to be highly educated, with higher income, status, and position in their organization (Gustafson, 2012), so their mode choices may be less affected by cost. Travel time, seat-scheduling flexibility, and other factors may take center stage for that subset of long-distance travelers (Unger et al., 2016). Tourism and business purposes are key designations for long-distance passenger travel, with tourism and leisure trips happening more often during holiday periods (Große et al., 2019).

Statewide and nationwide trip models have sought to evaluate the impacts of long-distance travel (Erhardt et al., 2007; Rohr et al., 2013; Bernardin et al., 2017; LaMondia et al., 2016b; Perrine et al., 2020). To anticipate congestion impacts, understanding long-distance-trip timing is very important, with many travelers (and freight carriers) purposefully avoiding the a.m. and p.m. peaks around and through urban areas. This has been examined in some depth in past studies

(Aultman-Hall et al., 2015; Sullivan et al., 2016; Scheiner, 2010; Reichert and Holz-Rau, 2015).

Li et al. (2022) conducted a 73-question online survey (with a focus on the Austin area) in 2019 that captured 2,327 long-distance trips (over 100 miles each way) made by 929 respondents over the prior 12 months. Predictive models for long-distance business trips per adult per year, numbers of overnights, travel times, and other travel attributes were developed; the results show that those with higher education tend to travel more often—for both business and non-business purposes—everything else constant. Older and/or lower-income people are likely to spend more nights at business trip destinations, as well as locations that are farther away from their origins. Persons who travel long-distance more frequently are more likely to spend less time in transit or en-route, and those who travel more often for business tend to spend more nights away, especially if the destination is more than 300 miles from their home.

1.1.2. Long-Distance Trips by AVs

The short- and long-distance travel impacts of AVs have been studied in recent years, including predictions of increased national vehicle miles traveled (VMT) by heavy trucks and passenger vehicles. Trip generation increase is an interesting place to start, since 19% of Americans with disabilities leave their homes infrequently and are unlikely to take long-distance trips (BTS, 2003). Harper et al. (2016) estimated a 14% increase in US VMT due to travel by currently non-driving, travel-restricted, and elderly Americans thanks to AVs. Meyer and Deix (2014) noted that if AVs allow disabled persons to make the same length and number of car trips as non-disabled Americans, their VMT will rise by over 50%. Lee and Kockelman's (2019) holistic look at all vectors of added passenger travel predicts a 25% rise in VMT as a result of AV introduction, and Huang et al.'s (2019) look at the Texas Triangle predicts a rise of over 40% in VMT for that megaregion long-term (i.e., by year 2045) due to AVs, with the number of links exceeding daily capacity more than doubling.

AVs reduce the burden of travel for drivers and may improve the quality of travel for passengers, who can now focus on more meaningful interactions with those previously focused on driving. Business travelers will be freed to work en-route to their destination, as they would on a plane, but with no airport access, egress, or wait times. Families and friends traveling together may be able to have quality interaction time while traveling with great flexibility in departure time and at a reduced cost, as compared to trains and airlines. Thanks to easier "driving," a driver's value of travel time (VOTT), or his/her willingness to pay (WTP) to save travel time, is expected to fall 20–50% or more. LaMondia et al. (2016a) explored long-distance mode choices originating in Michigan and forecasted that over 25% of airline trips under 500 miles will shift to AVs. Such changes will have important impacts on airlines, infrastructure planning and future land use (especially on and around long-distance-transportation facilities), highway congestion, and the travel industry more generally. Gurumurthy and Kockelman (2020) designed, disseminated, and then analyzed a nationwide survey on AVs' impacts on Americans' passenger travel choices and found that AV-sharing and dynamic ride-sharing should rise over time, for a variety of reasons, with shared AVs (SAVs) particularly popular for long-distance business travel.

To analyze the impacts of AVs in the United States, Perrine et al. (2020) added a new AV mode to a subset of the rJourney mode and destination choice models. With a base scenario assuming AV operating costs to be 20% higher than those of conventional vehicles, AVs reduced US airline revenues from domestic travel by a dramatic 53%. Availability of SAVs and AVs also shifted destination choices, for an overall 6.7% decline in US PMT from existing long-distance trip-generation rates. Such research needs much further development and can be supplemented with newer Texas- and long-distance-focused surveys, incorporating more complete details on Texas airport offerings, airline responses, and a thoughtful prediction of market shares over time (rather than simply a "before" vs. long-term "after" scenario comparison). Kim et al. (2020) surveyed more than 3,000 Georgians regarding their expectations of 16 potential changes brought by AVs. Results show that more than half of the respondents expressed enthusiasm for changing their activity patterns due to AVs, in terms of conducting more leisure and long-distance travel, as well as traveling to farther destinations.

1.1.3. Automated Trucking

Most existing AV studies focus on intra-urban trip-making, but few (Huang et al., 2019, Huang and Kockelman, 2019) anticipate changes in freight transport. Fully automated, self-driving trucks or "ATrucks" are those that can leave truck terminals and travel to a destination without human intervention (Viscelli, 2018). ATrucks may be equipped with other automated functions, like drop-offs and pick-ups, but most experts expect an attendant on board, doing other types of work, sleeping as needed, and ensuring successful deliveries and pickups while protecting the truck asset and its cargo en route (Yankelevich, et al., 2018). Vehicle attendants' ability to multi-task will allow for extended use of commercial trucks (e.g., every day, closer to 24 hours a day) and greater labor productivity, resulting in lower per-mile and per-ton-mile freight delivery costs.

Trucks carry over 2 trillion ton-miles of freight around the US each year, which is almost 40% of the nation's total (BTS, 2017). Investment in and use of ATrucks will affect not only national and regional economies (Clements and Kockelman, 2017) but trade patterns, production levels, and goods pricing. Commercial trucks consume about 20% of the nation's transportation fuel, and self-driving technologies are predicted to reduce those diesel fuel bills by 4 to 7% (Barth et al., 2004; Shladover et al., 2006).

ATrucks can reduce some environmental impacts, lower crash rates, and increase efficiency in warehousing operations, line-haul transportation, and last-mile deliveries. Platooned convoys should enable following truck drivers to avoid certain restrictions on service hours, enabling longer driving distances. Uranga (2017) predicts greater use of ATrucks before passenger vehicle automation, thanks to the more obvious economic benefits of self-driving trucks (which start with higher price tags as investment is less of a cost burden for trucking companies relative to personal travelers).

Huang and Kockelman (2019) anticipated changes in US highway and rail trade patterns in the long-term, following widespread ATruck availability. They used a random-utility-based multiregional input-output (RUBMRIO) model, driven by foreign export demands, to simulate changes in freight flows among 3,109 US counties and 117 export zones via a nested-logit model for shipment or input origin and mode, including the shipper's choice between ATrucks and conventional or human-driven trucks (HTrucks). Mode and shipment-origin choices (by freight carrier) were investigated to see how freight flow patterns would shift based on cost and operations changes due to vehicle automation. An ATruck's total ownership and operation cost (normalized per mile of driving) was assumed to be 25% lower than that of a conventional truck, thanks to some safety benefits (per mile traveled), fewer rest stops (thus more productive use of the vehicle and the operator, if one is on board, who can stay with the expensive vehicle and help with pickups and deliveries), and much better use of the on-board operator, who can do work that adds value while en route (e.g., taking the place of office staff, thereby reducing the shipping firm's payroll). Such savings help offset the higher acquisition cost of a self-driving truck (relative to a conventional heavy-duty truck).

Different VOTT and cost scenarios were explored, to provide a sense of variation in the uncertain future of ground-based trade flows. Using the current US Freight Analysis Framework (FAF) data for travel times and costs—and assuming that ATrucks lower trucking costs by 25% (per ton-mile delivered)—truck flow values in ton-miles were predicted to rise 11% due to automation's lowering of trucking costs, while rail flow values fall 4.8%. Rail flows were predicted to rise 6.6% for trip distances between 1,000 and 1,500 miles, with truck volumes rising for all other distance bands. All major cities were predicted to see lower rail flows (inbound and outbound), with San Jose, CA, and Washington, DC, experiencing more than 70% reductions in outbound rail flows. Truck flows were also predicted to lose many interactions between the northwestern US and Florida and northeastern states, while experiencing greater interactions along the western US (between California and Oregon), and also between the Great Lakes region (including Michigan and Illinois) and California. Rail flows were estimated to rise only in and around New Mexico, while dropping noticeably elsewhere (e.g., in Texas and from San Francisco and Arizona to the Great Lakes and northeastern areas, respectively).

1.1.4. AVs' Effects on Intra-Regional Travel Patterns

To date, there have been more than 20 simulations of AVs' and SAVs' effects on within-region travel. For example, Childress et al. (2015) used a Seattle, Washington, activity-based travel model (including short-term travel choices and long-term work-location and auto-ownership choices) to anticipate AVs' impacts on regional travel (assuming higher roadway capacities, lowered VOTT, reduced parking costs, and increased car-sharing). They estimated that higher-income households are more likely to choose the AV mode, as costly technology and VOTT reductions for higher-VOTT travelers are likely to be more significant. When SAVs cost $1.65 per mile (almost the same as current ride-sharing taxi services, like Lyft and Uber), drive-alone trips were estimated to fall by one-third and transit shares to rise by 140%, since modeled households let go of conventional vehicles and bought AVs, or shifted to SAVs as well as other travel options.

Zhao and Kockelman (2018) extended Capital Area Metropolitan Planning Organization's (CAMPO) demand model for the Austin region to include AV and SAV mode options and predicted a 20% rise in regional VMT over a 10-year horizon, assuming AVs and SAVs were widely available. Without shorter headways and common use of ride-sharing (between strangers), congestion would worsen. Liu et al. (2017) simulated conventional vehicles and SAVs in the Austin network using different SAV fare assumptions (e.g., $2, $1, and $0.50 per mile) with agent-based MATSim software. Assuming that the average SAV can serve 17 to 20 person trips per day, higher SAV fleet sizes deliver greater vehicle replacement rates, with one SAV replacing 5.6 to 7.7 conventionally owned and operated household vehicles. Martinez and Crist (2015) replaced Lisbon, Portugal's buses and cars with mid-size and small SAVs and estimated a 1:10 replacement rate, along with a 6% increase in VMT. Importantly, reduced parking needs freed up sizable land areas throughout the region.

1.1.5. Environmental Implications

Transportation accounted for 28.5% of US greenhouse gas (GHG) emissions in 2016, among which 41.6% was from passenger cars, 22.9% from freight trucks, 9% from air travel, and 2% from rail (EPA, 2018). Van Goeverden et al. (2016) estimated that trips longer than 100 km (62 miles) account for 45% of VMT and 50% of GHG emissions due to person travel in Holland using 2011 data, and passenger rail had one-third the emissions per passenger-km compared to automobile for personal trips. The Federal Aviation Administration (2015) reported that relative to air travel, transit buses are less efficient in fuel consumption per passenger mile due to lower passenger vehicle occupancy, and US rail travel energy efficiency was only slightly higher.

While more driving generally means more emissions from the transport sector, passenger AVs are expected to be highly or fully electric, due to all the computing and sensor demands on board. Wadud et al. (2016) sought to combine automation's effects on VMT and GHG emissions but found striking variability in future-year outcomes, since AVs can generate so much new travel, electrification of vehicle drivetrains is not guaranteed, and changing power-grid feedstocks deliver very different emissions impacts. Fagnant and Kockelman (2014) micro-simulated right-sized SAV fleet operations for a 10-mile-squared "town" and compared net emissions to those from the standard US household vehicle mix. They estimated that SAV systems may raise VMT 10% or more but save a great deal of carbon monoxide and VOCs, thanks to keeping engines warm and catalytic converters hot (with very few "cold starts" for a hard-working SAV fleet). When they added dynamic ride-sharing (between strangers), they were able to get total VMT to fall, rather than rise (Fagnant et al., 2015). Lee and Kockelman's (2019) recent work that randomizes all the possible contributions to VMT and energy use suggests that the overall trend is downward on energy and GHGs, thanks to most AVs being electric. But the VMT and congestion effects remain stark, and in the wrong direction. Even once the US DOT and/or state DOTs mandate one-second headways on highways (and eventually local streets), congestion associated with at-grade intersections and excessive VMT increases will still delay travelers.

1.1.6. Long-Distance Tourism

Distance plays an important role in tourists' travel behaviors, and the relationships between distance and tourist behavior have been extensively explored. It is believed that distance negatively affects tourists' destination choices and tourists are more likely to take short-distance rather than long-distance trips (Lee et al., 2014). However, tourists' preference for short-haul trips relates to many other factors as well, such as time cost, financial cost, and safety concerns (McKercher, 2008; Xue and Zhang, 2020).

Travel distance is commonly used to divide tourists into two groups: short-distance (short-haul) and long-distance (long-haul), since tourist behaviors change abruptly once long-distance travel is established as an objective (Bao and McKercher, 2008). The definition of long-distance tourism is not decisive, and many scholars use different versions. For example, Boerjan (1995) considered long-haul travel by airplane as trips with a travel duration of more than 5 hours or travel distance of more than 3,000 miles to destination. Lehto et al. (2002) defined long-haul travel as "travel more than four nights or more by plane outside of the international area." As tourists' preferences and the technologies facilitating travel change, the definition of long-haul travel is also changing.

Many studies have been conducted to compare tourist behaviors between long-distance and short-distance travel, in terms of motivation, preferred activities, length of stay, travel intensity, and consumption. Long-distance travelers tend to engage in multi-destination travel and attempt to satisfy multiple trip purposes (McKercher and Cros, 2003; Tideswell and Faulkner, 1999). The long-distance traveler has a higher length of stay and travel expenditure (Jackman et al., 2020; Nicolau et al., 2018). Also, long-distance travelers care more about quality and product features and tend to spend more on travel (Boerjan, 1995; Lo and Lam, 2004).

1.2. 2016/17 NHTS Long-Distance Data Summary

The 2016/17 NHTS data were collected with travel dates starting on April 19, 2016, and ending on April 25, 2017 (FHWA, 2017). A respondent's designated 24-hour travel day started at 4:00 a.m. (local time) of the assigned travel day and ended at 3:59 a.m. of the following day. Weights were utilized to produce well-balanced population-level estimates, including weights of household, trip, person, and vehicle (FHWA, 2017). Below are some explorations of the 2016/17 NHTS data, and comparisons of travel behaviors between the respondents from entire US and the state of Texas. For most of the analysis, results are shown from both the US and Texas, as well as for three types of trips: trips across all distances (all-distance trips), long-distance trips greater than 50 miles one-way (LD50+ trips), and long-distance trips greater than 100 miles one-way (LD100+ trips).

1.2.1. Long-Distance Trips vs. All Trips

The differences between long-distance trips and all-distance trips are compared in this section. Figure 1 shows the percentage of person trips taken each day of the week. The US and Texas show a similar pattern. As expected, for trips across all distances, each day of the week has similar person trips, but they are slightly more likely on weekdays. Long-distance trips (LD50+ trips and LD100+ trips) are more likely to occur on weekend days.

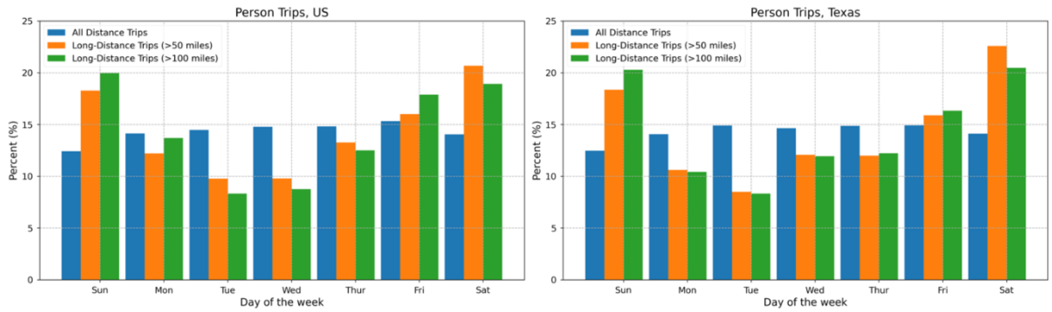


Figure 1. Percentage of one-way person trips by day of the week in both the US (left) and Texas (right)

The all-distance and long-distance person-trip counts by different trip purposes are presented by day of the week in Figure 2 and Figure 3). Since the LD50+ and LD100+ trips have a similar pattern, the figures compare only the all-distance trips and LD50+ trips. Several findings can be observed from the figures below:

- Compared with all-distance trips, LD50+ trip purposes are made up of larger shares of visit friends or relatives, work-related business, and other social or recreational trips. This is expected because the destinations of trips with these purposes are more likely to be over 50 miles away.

- The US and Texas show similar weekly patterns for both all-distance trips and LD50+ trips. Compared with long-distance trips, all-distance trips include more shopping and other social or recreational trips on weekends, more dental or medical trips on weekdays, and more school or church trips on Saturday.

- LD50+ trips have a higher share of visit friends or relatives and "other" trips than all-distance trips do, for each day of the week. Similar to all-distance trips, there are more shopping, visit friends or relatives, and other social or recreational trips on weekends, but more dental or medical trips on weekdays.

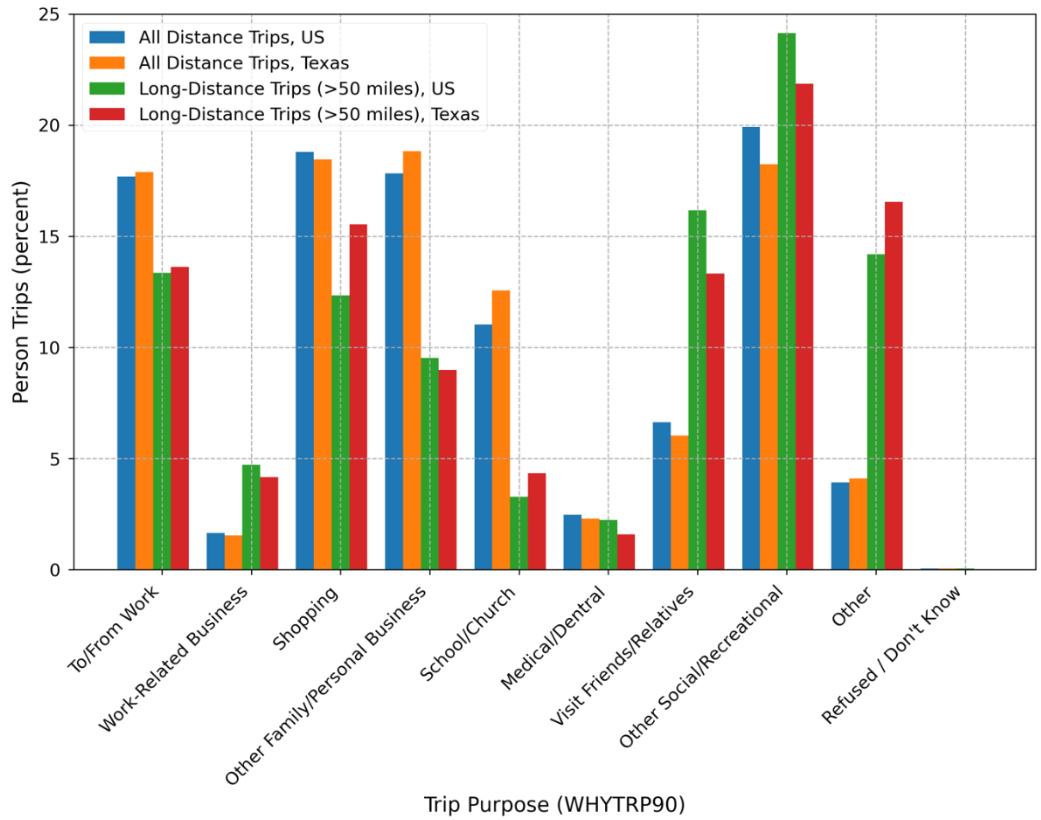


Figure 2. Percentage of one-way person trips by trip purpose (WHYTR90) in the US and Texas

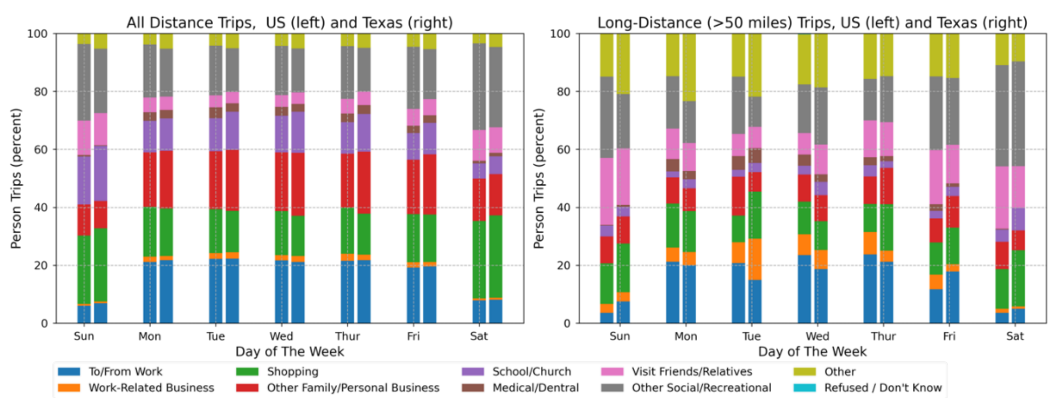


Figure 3. Percentage of one-way person trips by trip purpose (WHYTR90) and day of the week in the US and Texas

1.2.2. Person Trips by Mode

Figure 4 shows the share of person trips by transportation mode for both all-distance and long-distance (>50 miles) trips in the US and Texas. For all-distance trips, the five dominant modes are car, SUV, van, pickup truck, and walking. Texas has a larger portion of pickup truck and SUV trips compared to the US for both all-distance trips and long-distance trips. As expected, the five dominant modes for long-distance trips are car, SUV, van, pickup truck, and airplane.

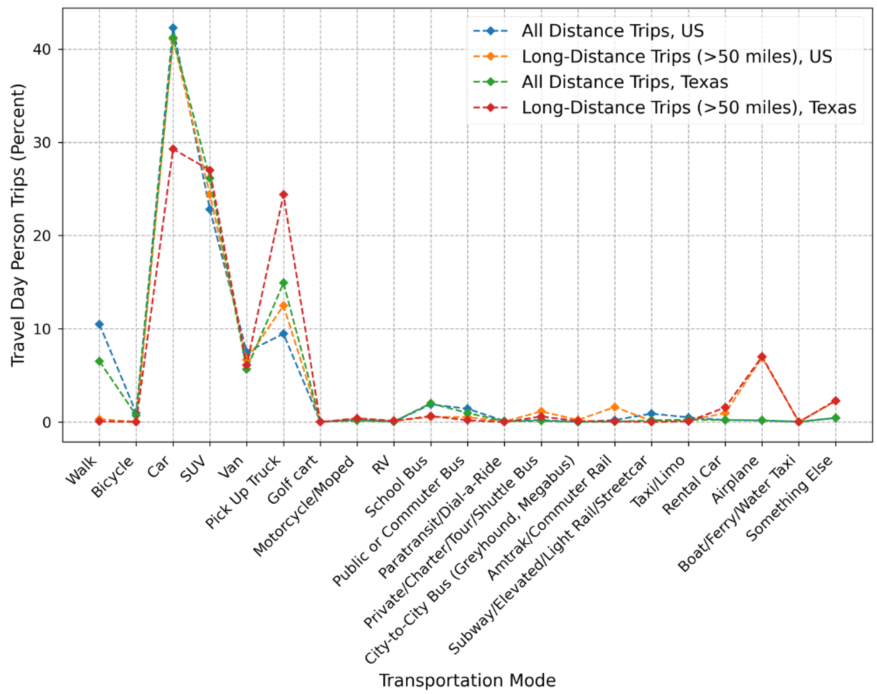


Figure 4. Percentage of one-way person trips by different transportation modes in the US and Texas

Mode share by travel distance is further presented. For simplicity, vehicle trips are split into single-occupancy vehicle (SOV) and high-occupancy vehicle (HOV) driving trips and all transit trips are categorized together. As observed in Figure 5 and Figure 6, the US and Texas have a similar pattern. Automobile (SOV + HOV) is the dominant mode (84% in the US vs. 89% in Texas) for travel of less than 400 miles, while air becomes the dominant mode (72% in the US vs. 64% in Texas) for trips longer than 400 miles, which is similar to the results obtained by Moeckel et al. (2015). Most bike and walk trips are less than 25 miles. Texas has a higher share of vehicle mode choice for trips over 600 miles than the US overall (19% for the US and 29% for Texas). Some transit trips for both Texas and US are also observed to be longer than 100 miles, indicating some people choose transit, like rail or shuttle bus, for long-distance travel.

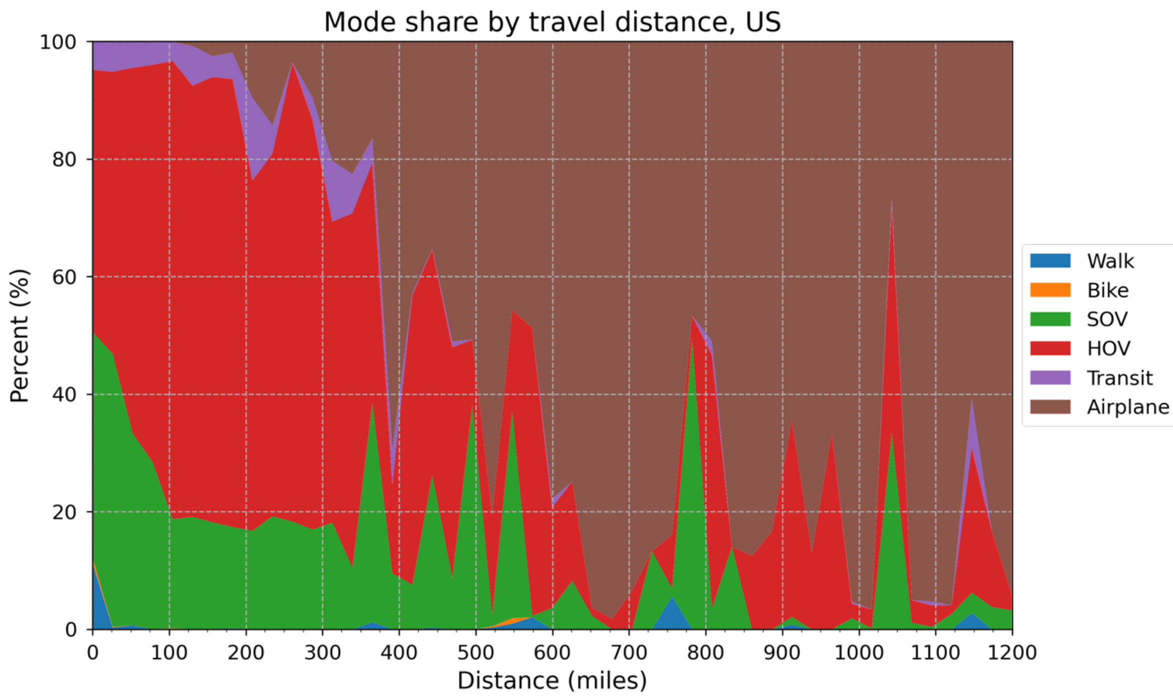


Figure 5. The mode share by travel distance for US person trips

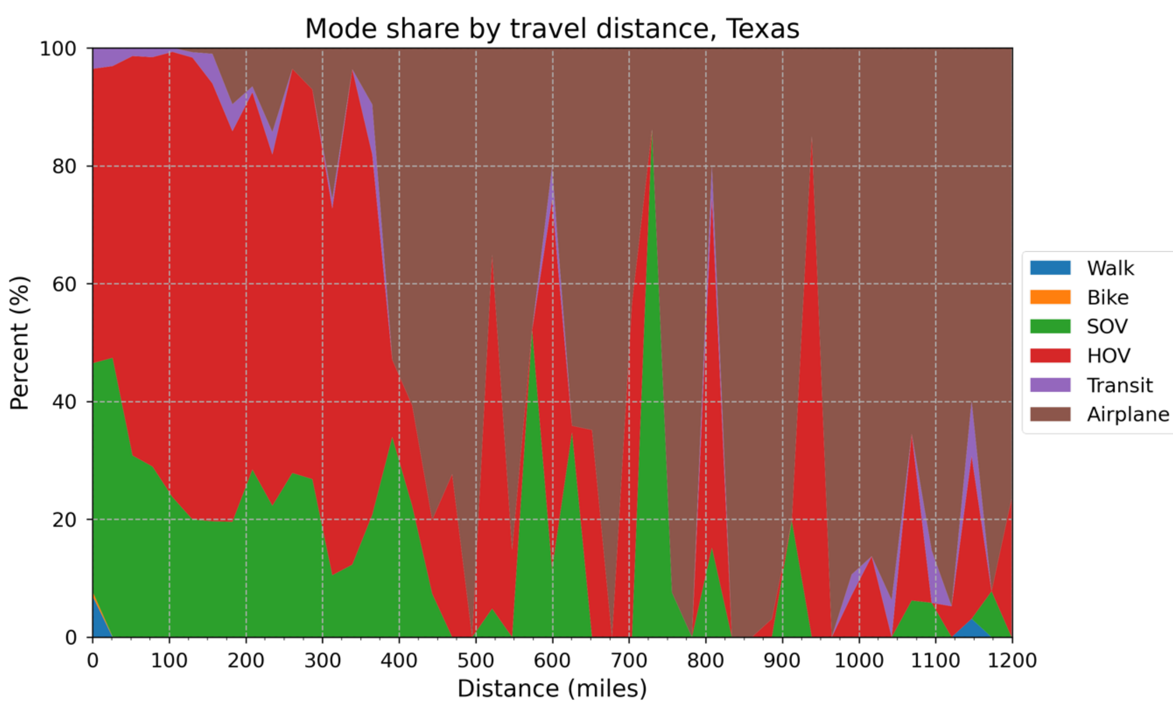


Figure 6. The mode share by travel distance for Texas person trips

1.2.3. Person-Miles Traveled (PMT)

This section presents PMT by trip purpose. PMT is calculated by summing weighted trip miles. The total PMT for one-way trips is 3,951.2 billion for US all-distance trips (352.1 billion in Texas), among which 1,707.7 billion, about 43% (158.0 billion, about 45%, in Texas), are LD 50+ trips, and 1,356.6 billion, about 34% (130.3 billion, about 37%, in Texas), are LD 100+ trips. Texas has higher shares of LD50+ and LD100+ trips, compared with the US average. Figure 7 indicates that Friday has the greatest share of PMT for all-distance trips in the US, while for Texas it's Saturday. For long-distance trips, Sunday has the biggest share of PMT both for the US and Texas.

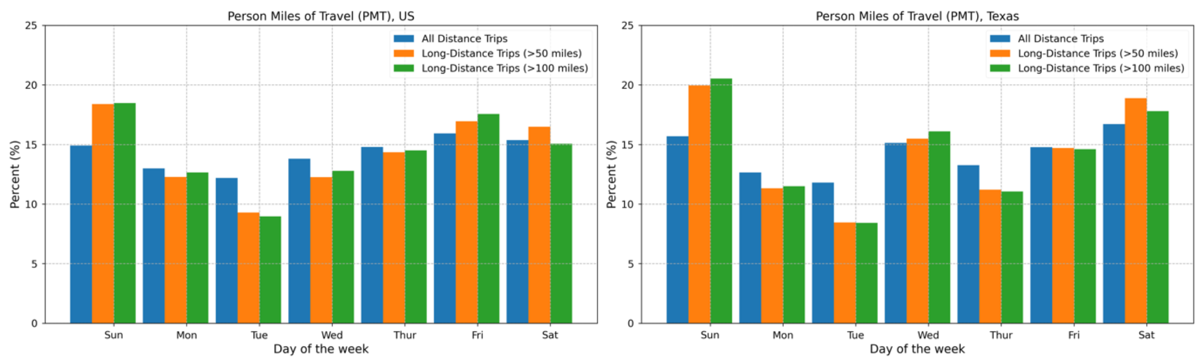


Figure 7. The PMT share by day of the week in both the US and Texas

PMT by trip purpose (in two different categories in NHTS: "TRIPPURP" and "WHYTRP90") was explored, and the results are shown in Figure 8. The US and Texas show a similar pattern for both trip purpose categorization methods.

- For the trip purpose method of WHYTRP90, to- or from-work trips make up the biggest share of PMT for all-distance trips, followed by "other" trips, while for long-distance trips, "other" trips are the biggest share of PMT, followed by other social or recreational trips. Long-distance trips (>50 miles) have larger PMT shares of "other," visit friends, and work-related business trips compared to their all-distance trip shares.

- For the trip purpose category of TRIPPURP, the non-home-based (NHB) makes up the biggest share of PMT for both all-distance trips and long-distance trips (LD50+ and LD100+). For long-distance trips, the percentages of PMT on home-based trips (HBO, HBSHOP, HBO) are lower than they are for all-distance trips.

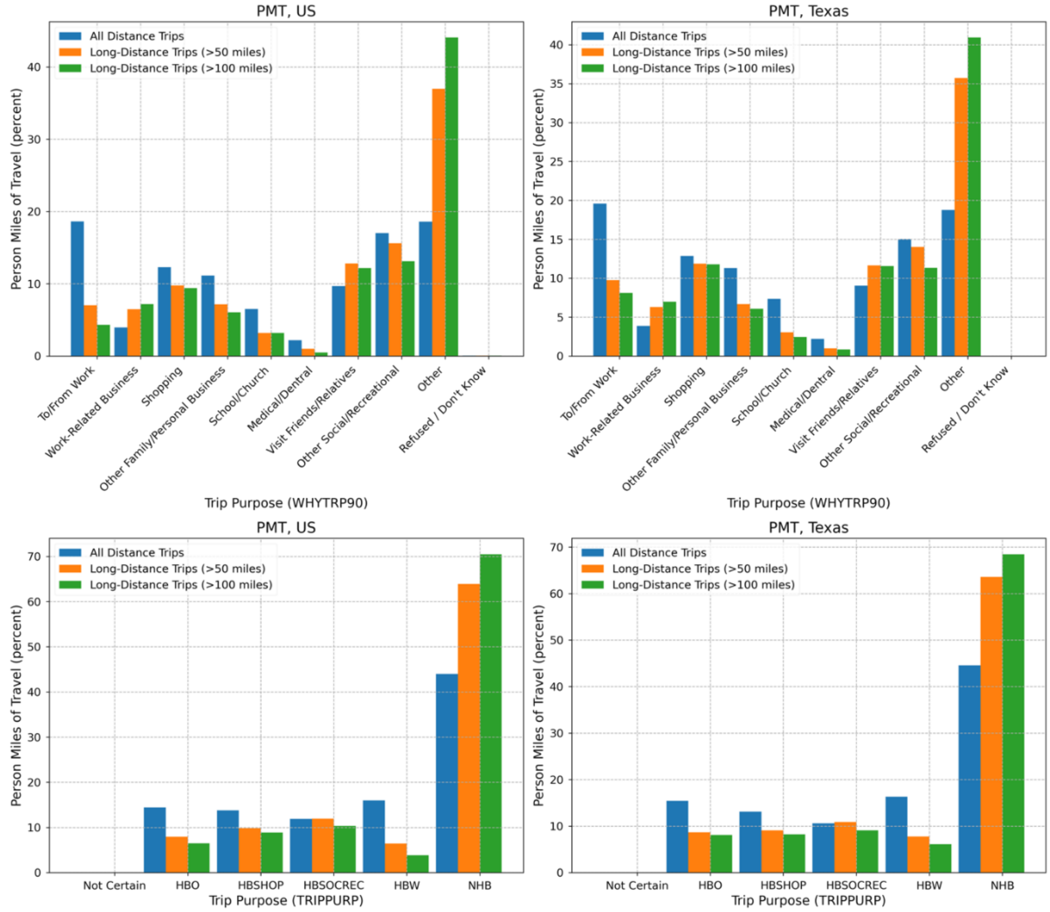


Figure 8. PMT share by WHYTRP90 (top) and TRIPPURP (bottom) in the US and Texas

The PMT by mode was also explored, using the same definition of travel mode used in 1.2.2. Figure 9 shows that, for all-distance trips, Texas and the US present similar patterns, with automobile (SOV+HOV) trips making up the majority of total PMT. Also, for trips across all distances and long distances, Texas's PMT are made up of a higher proportion of automobile (SOV+HOV) and a lower proportion of air modes than the US overall, indicating that on average, Texas prefers road trips more than the US does as a whole. Compared with mode share for all-distance trips, the share of PMT by air increases for long-distance trips (LD50+) both in the US and Texas, but the share of PMT by SOV decreases dramatically, suggesting that SOV trips tend to be short-distance trips.

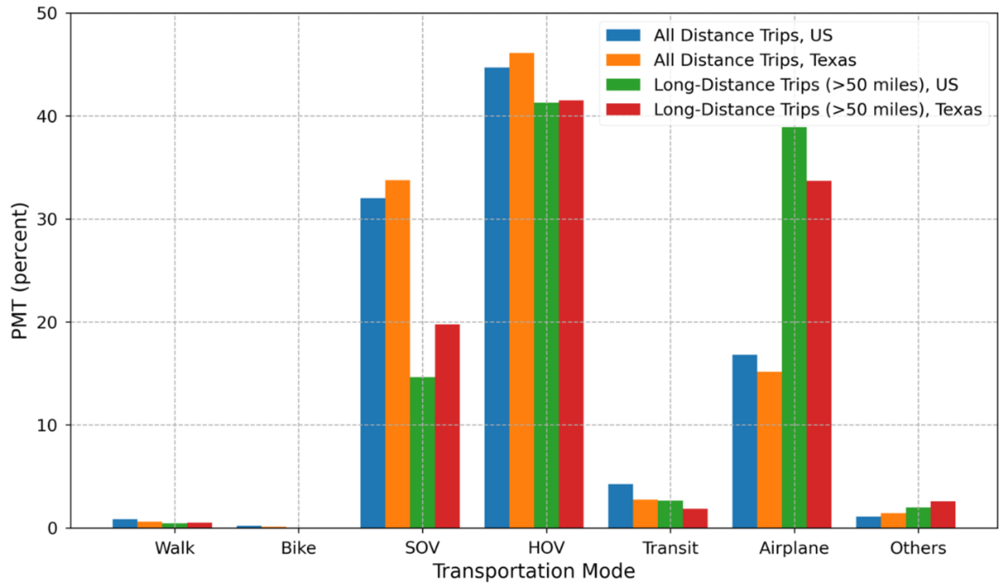


Figure 9. The PMT share by mode for both US and Texas

1.2.4. Travel Party Size

This section analyzes the number of persons on the trip (party size) by date (Figure 10 and Figure 11). The party size is the weighted average number of persons on a trip, considering duplicate trips (household members on the same trip). Figure 10 and Figure 11 show the travel party size for all-distance trips and long-distance trips by week and by month, respectively. Party size varies more for modes like transit or airplane, so the party size of personal car travel (SOV+HOV) was explored by week and by month. As observed, long-distance (>50 miles) trips tend to have more people in a party than all-distance trips: the average party sizes for long-distance trips are 1.38 for the US and 1.66 for Texas, and the average party sizes for all-distance trips are 1.21 for the US and 1.42 for Texas. For all-distance trips, the trend of party size for the US and for Texas by timeline looks similar, while for long-distance trips, Texas has a larger party size in June, July, and January than the broader US.

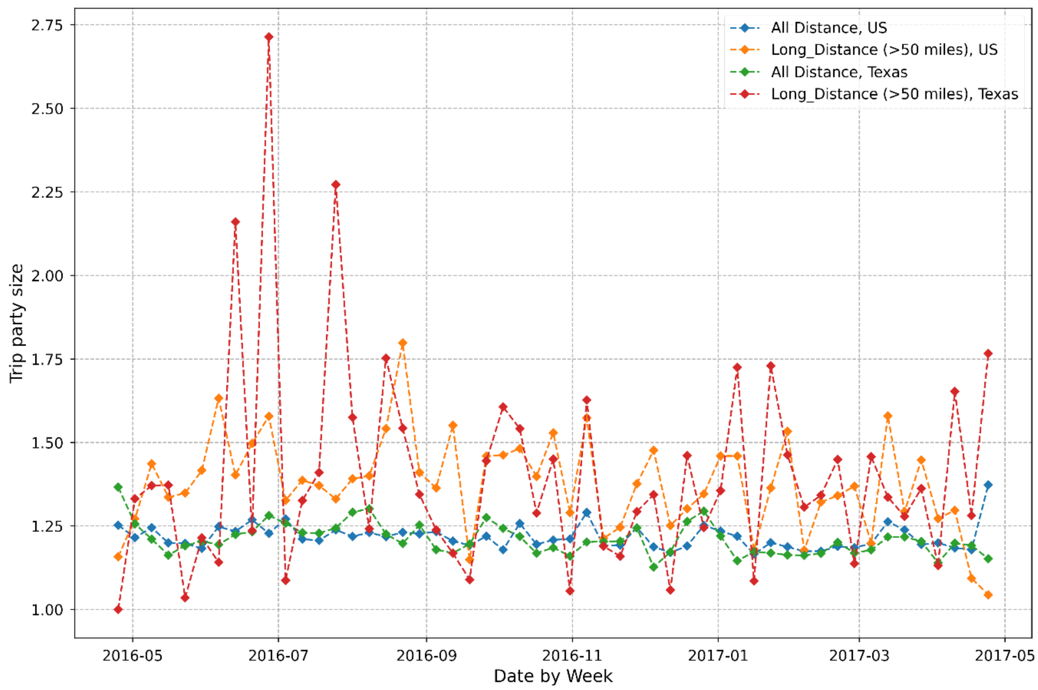


Figure 10. The travel party size of automobiles (SOV+HOV) by week in the US and Texas

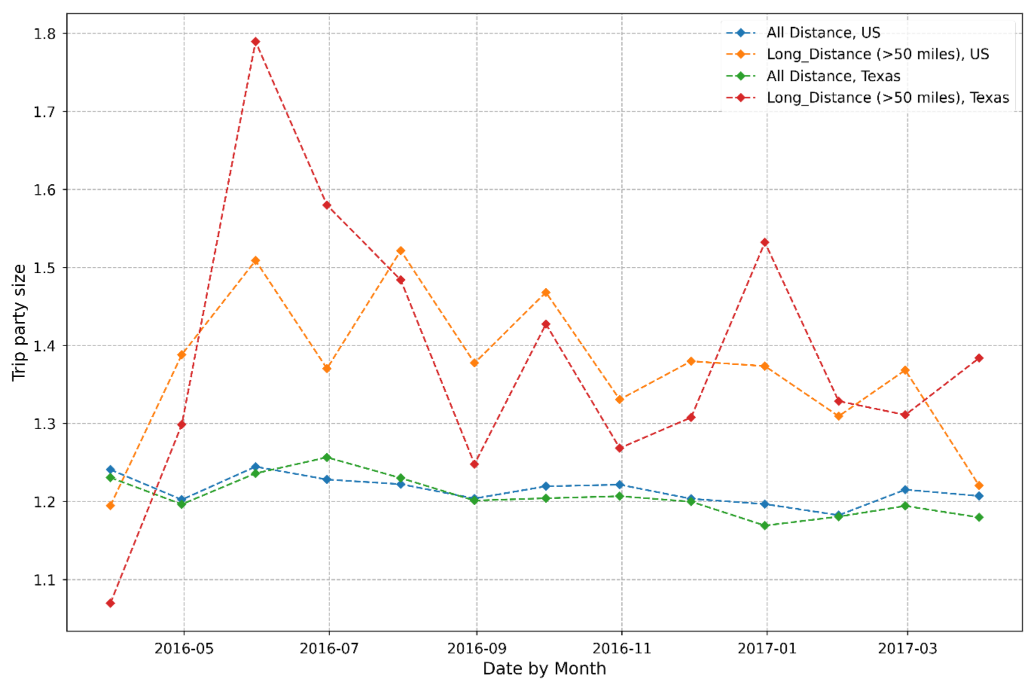


Figure 11. The travel party size of automobiles by month in the US and Texas

Table 1 compares the average vehicle occupancies (based on miles traveled, using PMT/VMT) for four types of travel methods in the US and Texas.

Table 1. Average vehicle occupancy

PMV/VMT	Texas	US
Car	1.89	2.15
SUV	2.28	2.73
Van	2.77	3.93
Pickup truck	1.85	2.11
Rental car	2.27	2.51

In contrast to commonly seen statistics of average vehicle occupancy weighted by trip counts, this table shows the average number of members in a travel party by trip distance. For most travel methods, two or more people are traveling together, with the exceptions of Texans in cars and pickup trucks. Texans drive around relatively "solo" compared to all Americans, as the average occupancies for every mode are lower in Texas than the US overall. This owes to plentiful parking and low gas prices, increasing Texans' rates of owning vehicles for individual use.

1.2.5. Loop Trips vs. One-Way Trips

The 2016/17 NHTS data collected round trips because the survey was designed to be completed online without an interviewer, in contrast to earlier surveys in the series, where an interviewer could help respondents artificially split the loop in half. Round trips were considered those where the respondent listed the same start and end locations, often with the destination unspecified. The total number of trips reported in the survey and using expansion factors is 371.2 billion (30.4 billion in Texas), of which 1.9% (1.7% in Texas) are loop trips and 98.1% (98.3% in Texas) are one-way trips. Table 2 summarizes the trip length and trip duration of both valid loops and one-way trips, showing similar patterns in the US and Texas. The mean and median trip durations for loop trips are longer than for one-way trips. Most of the loop trips are around the home, so the mean and median trip lengths in miles of loop trips are shorter compared to one-way trips.

Table 2. Summary of trip distance and trip duration in US and Texas

	US				**Texas**			
	Trip Duration (min)		**Trip Distance (miles)**		**Trip Duration (min)**		**Trip Distance (miles)**	
Trip Type	Loop	One-Way	Loop	One-Way	Loop	One-Way	Loop	One-Way
Sample Size	19,587	901,526	19,587	901,526	3,296	174,777	3,296	174,777
Mean	37.6	21.2	2.3	11.2	39.1	21.5	2.4	11.8
Median	30.0	15.0	1.0	3.5	30.0	15.0	1.0	3.7
Std Dev	60.4	31.7	10.0	67.7	68.1	31.7	11.1	69.2

	US				Texas			
	Trip Duration (min)		Trip Distance (miles)		Trip Duration (min)		Trip Distance (miles)	
Min	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
Max	1200.0	1140.0	440.0	9621.1	1160.0	1050.0	300.0	6515.9

Loop trips represent only a small portion of overall trips and require further efforts to be manually separated for analysis (Buehler et al., 2020), so only one-way trips were used for the data analysis in this section.

1.3. COVID-19's Effects on US Households and Passenger Travel

The COVID-19 pandemic has impacted the transportation system and travel behaviors, especially long-distance travel. Almost all modes and all purposes of trips witnessed a worldwide drop from February to May 2020 (Beck and Hensher, 2020; Molloy et al., 2020). More people tend to work from home during the pandemic (Wang, 2020; Geman, 2020), and telecommuting is expected to continue long after the pandemic, based on a survey during late March 2020, when half of the employed adult respondents were working from home (Guyot and Sawhill, 2020).

1.3.1. Travel Trends During COVID-19

Since people hesitate to use shared modes during COVID-19, public transit has been extensively affected (Beck and Hensher, 2020; Transit, 2020; Wang et al., 2020). According to the Bureau of Transportation Statistics (2020), there was a 30–50% reduction in trips less than 25 miles between March and September 2020 in the US, compared to 2019. Trips between 25 and 100 miles dropped at the beginning of March and reached their lowest point in early April, a roughly 47% reduction compared to the same time in 2019, but the trend began to recover and remained stable (about 20% reduction vs. 2019) after May. Interestingly, compared to a stable trend in 2019, trips longer than 100 miles in 2020 first witnessed a small peak pre-pandemic but dropped to their lowest point, about a 30% reduction, in April. Then, however, the trend began to bounce back to normal, reaching 2019 levels at the end of May, and exceeding them by 64% at the end of August 2019. The same reduction trend was witnessed for US commercial flights, both domestic and international, which experienced a nearly 70% drop in mid-April, but then domestic flights started to revive from their earlier drops with a slight increase seen for international flights. International passenger vehicle trips, which presumably are long-distance trips, also greatly decreased during the first year of the COVID-19 pandemic.

A similar trend of trip counts and mode share has been observed in many other countries. In Switzerland, there was widespread suppression of travel demand across almost all modes during the pandemic (Molloy et al., 2020). In Australia, commuting trips fell from an average of seven per week down to three, and half of respondents canceled their planned air travel (Beck and Hensher, 2020). In India, about 41% of commuters stopped traveling during the transition to the lockdown phase, and 5.3% of commuters shifted from public to private modes (Pawar, 2020). However, biking has become more popular during the pandemic, with trips increasing by 22% in places such as Greater Manchester in the UK (Rannard, 2020), inducing some short-distance trips. The same trend has been observed in Switzerland (Molloy et al., 2020), where biking trip lengths increased up to a 180% in early April.

Overall, COVID-19 has disrupted long-distance travel globally. In the US, drops have been observed in flight trips (both international and domestic) and other public modes, as well as international passenger vehicle trips, but domestic passenger vehicle trips are recovering, indicating greater long-distance travel, which is more likely for recreation purpose. At the same time, the use of bikes has increased, leading to more shorter trips in active modes.

1.3.2. Travel Trend Predictions

Surveys have been conducted to understand travel behavior before, during, and after the pandemic. The "COVID-19 and the Future Survey" (covidfuture.org) is a joint project of Arizona State University and the University of Illinois at Chicago with support from the National Science Foundation. Researchers in the project team focus on identifying the impacts that are likely to endure after the pandemic is over. The survey consists of questions in different categories for "before and now and future" scenarios of COVID-19, in terms of employment, working and studying, shopping and dining, daily transport, attitudes, demographics, and social networks. States with denser populations yielded more responses (about 20 metropolitan areas across the US), and Arizona is oversampled. The results are weighted to match the regional and national population, in terms of age, gender, education, Hispanic population, vehicle, income, and children.

By mid-November, the survey showed that 20% of respondents had been tested, with 1.6% positive, and 12% believed that they had had COVID-19. Among all respondents, 33% had faced a decreased income, but a few people observed an increase in income. COVID-19 also caused more home relocation; 14% of respondents had moved in the seven months since April 2020, matching the annual percentage of home moves in the US in all of 2019.

Survey results revealed trends related to travel, such as working from home, e-shopping, people's attitudes toward travel, and mode choice. An estimated 35% of respondents experienced increased productivity at home, mostly by recapturing time previously spent commuting, followed by other reasons like flexible hours and comfortable spaces at home. Based on the survey results, working at home would reduce transit use by half as well as driving. This has already been observed in many cities like New York and Washington, D.C., which feature reliable mass public transportation systems (Sadik-Khan and Solomonow, 2020). In terms of e-shopping, about 20% of respondents use online grocery shopping at least once per week. And among those who plan to use online grocery shopping a few times per month after the pandemic, 12.6% never used e-shopping before the pandemic. Most respondents agree that one should stay at home until COVID-19 subsides, and they perceive high risk in the use of public transportation and airplanes. This indicates that concerns regarding sharing space with other people are the main reason for not using air travel post-pandemic, but some respondents indicated that they would make more trips because they have been pent up at home for too long and desire to make up for travel that was canceled. Survey results also indicated that the new familiarity with conducting business meetings online would be the top reason for reduced business air trips in the future. As the pandemic influenced future transportation patterns, it will impact the way AVs transform future mobility. Survey results indicate that some people will tend to keep telecommuting after the pandemic (Mokhtarian and Grossman, 2020). It will be interesting to see how the increase in teleworking and the implementation of AVs would alter the way people make long-distance trips.

Survey Design and Application

AVs and SAVs will become increasingly available over the coming years. For this reason, the implications of these services on travel, trade, emissions, travel cost, and other factors need to be anticipated across Texas. This chapter addresses the design of a long-distance passenger-travel survey to predict impacts on mode and destination choices, travel preferences, VMT changes, and near- and long-term policies and cost scenarios across emerging technologies. The survey questions as well as the analyses of responses will be explained in more detail in the next chapters.

2.1. Survey Design

The long-distance passenger-travel survey was designed and coded into Qualtrics software. The final set of questions are shown in Appendix A. The survey consists of 70 questions (20–25 minutes) tackling aspects of long-distance travel, AV and SAV usage, and effects of the COVID-19 pandemic. The survey includes a mix of revealed and stated preference questions for current or recent trips and futuristic scenarios. A set of adaptative questions is included to evaluate travel time changes and users' willingness to pay to ride in AVs. Questions related to the effects of COVID-19 are also included, and different scenarios are tested for a future COVID-19-like virus to understand the impacts of possible future pandemics. The survey is divided into seven sections; each one collects different types of details. The following subsections describe the survey sections and their informational goals.

2.1.1. Introduction and Long-Distance Trip Definitions

The first section of the survey introduces the study and gives contact information for the research team. It also provides definitions of relevant concepts such as self-driving vehicles, one-way and round trips, and long-distance travel before respondents are shown the questions. For this study, long-distance travel is defined as a one-way trip that takes more than 75 miles from the origin to the destination (or round trips that involve more than 150 miles of travel, in total).

2.1.2. Long-Distance Trip Frequency and Trip Purpose

The trip frequency and trip purpose section asks how many non-business and business long-distance trips a respondent took in the calendar years 2019 and 2020. The separation by calendar year will facilitate users' memory and provide the average annual frequency of long-distance trips before the impacts of the

COVID-19 pandemic. This section also includes questions about mode choice and trip purpose for two ranges of distances: (1) between 75 miles and 500 miles, and (2) more than 500 miles. Research in the field indicates that users are more likely to travel by airplane when the distance is greater than 500 miles. Therefore, the survey differentiates these two types of long-distance trips. Finally, this section includes questions about trip purpose and trip frequency changes due to the COVID-19 pandemic. It contains questions that focus on travel plans in the following periods:

- Before the COVID-19 pandemic
- During the COVID-19 pandemic
- After COVID-19 is no longer a threat

2.1.3. Self-Driving and Shared Vehicle Technology Definition and Preferences

This section includes a further definition of self-driving automation. It introduces the concept of an SAV fleet, of vehicles that can be shared among people in a city or region. Figures of the service are shown, and a link is provided to access further information on the topic. Figure 12 shows an example of the definition of self-driving automation in the formats that the user will access for both computer and mobile versions. This section tests the understanding of the user, adding follow-up questions about the definitions. It also includes questions about preferences for using a self-driving car or a shared self-driving car and willingness to share a trip with other users in the same ride.

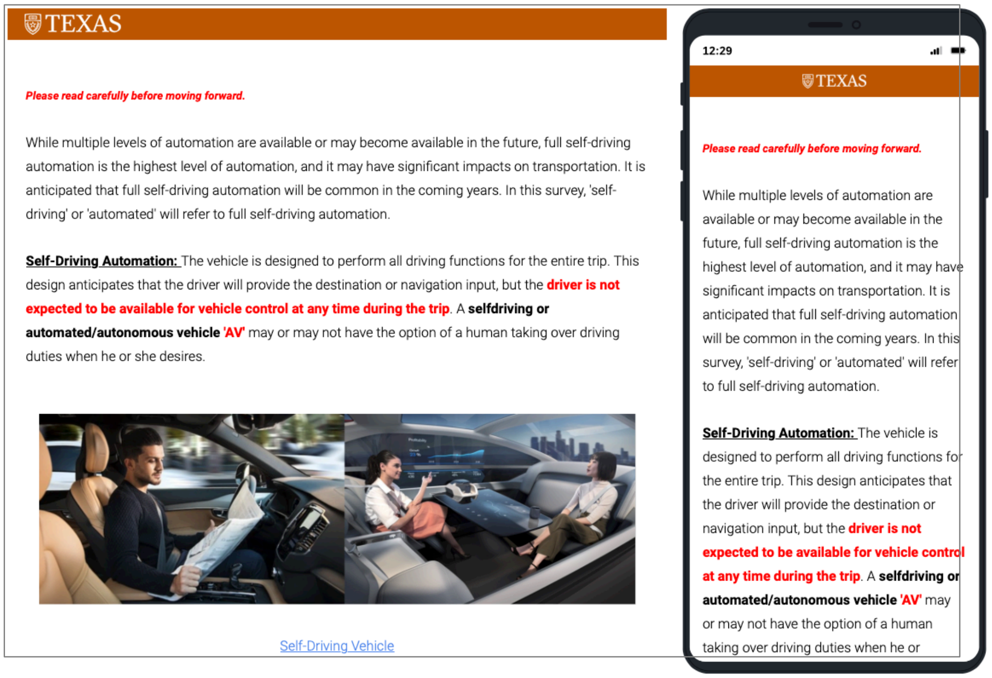


Figure 12. Self-driving automation definition, computer and mobile versions

2.1.4. Long-Distance Trip Revealed and Stated Preferences

This section includes a mix of revealed and stated preference questions. The survey asks for the description of a long-distance trip of more than 75 miles from the origin to the destination, made during a pre-COVID-19 period (12 months before March 2020). The respondent provides details of the trip, such as origin, destination, mode, duration, travel party size, and cost. While answering those questions, the respondent is asked what would have changed if he/she could have access to a self-driving vehicle for that same trip. This approach helps the respondent envision him- or herself in that situation and provide more realistic responses. The questions include the likelihood of changing their travel mode to the self-driving car under different cost and time assumptions and their willingness to pay for these situations. The survey also asks for possible travel party size, destination, and duration changes if the respondent had access to self-driving cars.

2.1.5. Long-Distance Future Scenario Questions

This section provides the user a futuristic scenario where self-driving cars are widely used for different services such as mail delivery and ride-sourcing (such as Uber and Lyft). Figure 13 shows the description presented to the respondent. Furthermore, self-driving vehicles are assumed to be affordable for the survey respondent. Based on this scenario, the survey asks questions about changes in frequency, duration, distance, destination, and departure time of possible long-distance trips using self-driving cars.

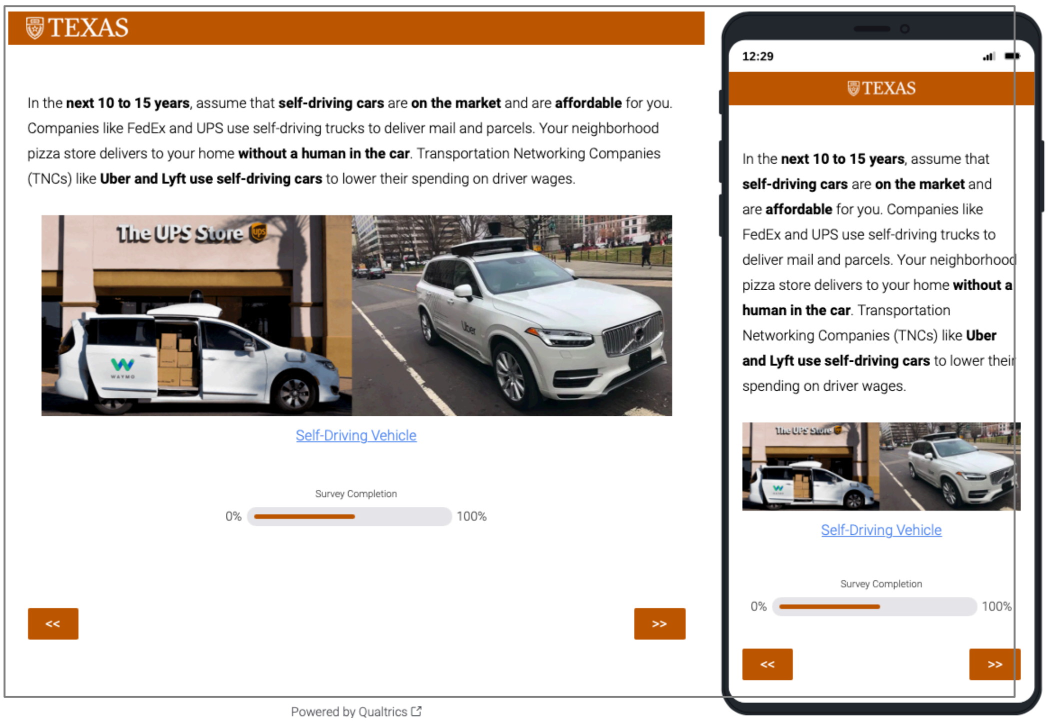


Figure 13. Description of the futuristic scenario, computer and mobile versions

Furthermore, this section includes mode choice preferences for long-distance trips in situations where the respondent is in a pandemic with a COVID-19-like virus under two different conditions:

- Situation 1. Consider a scenario where a COVID-19-like virus vaccine is universal and more than 70% of the population has received it with a clinical success rate higher than 85%.
- Situation 2. Consider a scenario where a COVID-19-like virus vaccine is universal and more than 50% of the population has received it with a clinical success rate higher than 85%.

2.1.6. Demographics

The demographic section collects user information such as sex and age range and household information such as income and number of vehicles. It also includes COVID-19-related questions such as if the user is considered to be at high risk of developing severe symptoms, which would influence their travel behavior during the pandemic period and other answers related to the topic.

2.1.7. End of Survey

The final section provides space for adding additional details the user would like to share. It also provides the option for the respondent to share his or her email address for further communication.

2.2. Survey Application

The survey was distributed online to a total sample of 1,004 US residents, using the large US panel of respondents that Dynata (previously named ResearchNow and SSI) maintains. To ensure the sample would catch people traveling through or into and out of Texas, it included 45% Texas residents and 55% residents of other continental US states (i.e., excluding Hawaii, Alaska, and Puerto Rico).

Survey Data Analysis

The influencing factors impacting long-distance trips; predictions of the impacts of AVs, SAVs, and ATrucks on long-distance trip-making; and impacts on tourism and overall travel behavior were explored through a thorough review of literature, data, and survey responses in Chapter 1. These facets of travel behavior are incredibly important in predicting network use, travel demand, and industry costs and expenses by mode and will be highly impacted by the introduction of automation technology. The most critical observations from the intensive literature review involve accepted patterns in long-distance travel in both the US and Texas. Long-distance travel is heavily impacted by frequency, distance, mode, destination, household income, traveler age, education level, and presence of children in the household; the opportunity for employee reimbursement is a factor in the case of long-distance travel for business.

Automation will allow for large cost savings as drivers will be able to spend their "driving" time in more productive ways, which will lead to an altered value of travel time and higher willingness to pay for an automated mode that allows for this heightened utility. The survey conducted during this project allows examination of how AV travelers may use their time on the road and how their travel choices may impact other modes of travel. The literature review found that the most common purpose for long-distance trips is visiting friends and relatives and that long-distance tourists are more likely to have higher numbers of destinations and trip purposes, greater expenditures, and trips of longer durations. It is also observed in both Texas and across the US that the primary mode for long-distance travel up to 400 miles is an automobile; beyond that distance air travel becomes the dominant mode. The survey addressed how the introduction of AVs impacts this distinction in mode choice. Finally, the literature surrounding the impacts of COVID-19 indicated that air travel greatly decreased during the height of the pandemic but has since returned to near-normal levels, with the number of longer work trips not bouncing back as quickly as the number of shorter trips and an overall heightened concern about using shared modes of transportation. The survey aims to address the discussed topics in-depth and provide respondents with various options on how they would address long-distance travel with the introduction of AVs and in a post-pandemic world.

As referenced in-depth in previous section, the survey developed is one of the first of its kind that focuses directly on long-distance travel with the use of AVs. The survey also accrues data on multiple variables that are excluded from other surveys and data sources on this topic, such as costs of accommodation, airlines' business expenses, and fuel, as well as stops and total trip duration—factors that cannot be captured in typical 24-hour travel behavior surveys. The survey also contributes to new research focusing on how long-distance travel has been impacted by the COVID-19 pandemic and how the pandemic may impact people's travel choices in the long term.

3.1. Survey Results

Following demographics determined by the US Census, the targets in Table 3 were set to collect a sample of responses that was representative of the nation and the state of Texas. A total sample of 1,004 responses was obtained after the filtering and cleaning process described in the next section. The final pool includes 451 (45%) Texans and 553 (55%) respondents from the rest of the nation. Figure 14 shows the location of the respondents across the nation.

Table 3. Demographic distribution of cleaned data

	Texas			US		
	Sample:	451		Sample:	553	
	Target	**Survey**	**%**	**Target**	**Survey**	**%**
Gender						
Male	249	189	76%	246	243	99%
Female	252	262	104%	254	310	122%
Age						
18 to 24 years	65	52	80%	60	51	85%
25 to 34 years	98	69	70%	90	98	109%
35 to 44 years	92	92	100%	82	126	154%
45 to 54 years	83	50	60%	80	103	129%
55 to 64 years	76	81	107%	83	107	129%
65 years and over	87	107	123%	106	68	64%
Census Region (Appendix D)						
Northeast				94	118	126%
Midwest				114	125	110%
West				131	137	105%
South				161	173	107%
Education						
High school or less	108	75	69%	190	97	51%
Some college/assoc/tech degree	152	154	101%	130	183	141%
Bachelor's degree	183	130	71%	113	168	149%
Master's degree or higher	67	82	122%	67	100	149%

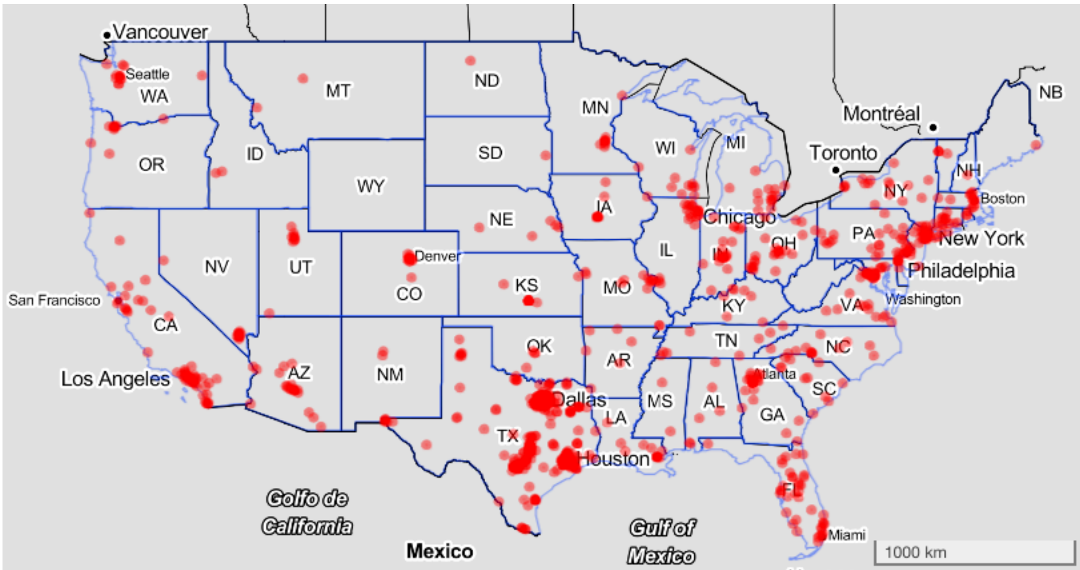


Figure 14. Respondents across the continental US

3.1.1. Response Cleaning

To preserve the accuracy of and reduce bias in the survey results, the responses were heavily monitored to ensure that only complete responses were examined. The cleaning tasks essentially include visual inspection of data format and value range and consistency checking to help identify incomplete records and invalid or inconsistent field entries. Responses were kept based on completion within a reasonable amount of time (longer than 13 minutes), no inconsistencies in responses (e.g., inconsistent zip code and state of residence), and legible and reasonable responses. As responses were eliminated by the research team, the sample demographic goals were readjusted. Poor response IDs were sent back to Dynata to be removed and for the survey to be retargeted toward groups with impacted response rates during the data collection process.

3.1.2. Weighting

As Table 3 and Figure 14 demonstrate, about half the sample comes from Texas, allowing both a detailed representation of the Texas region and a comparison with the rest of the US. This approach helps to comprehensively depict AV-related long-distance travel preferences across the US. The collected data were further weighted using the iterative proportional fitting (IPF) method to match the most recent five years of data from the American Community Survey (ACS) (Roth et al., 2017). The weighting targets incorporated the demographic distribution of age, region, and gender. The southern US area was separated into two parts: Texas and the rest of the southern US, such that Texas as well the US as a whole both match the ACS demographic distribution using the same weight set. The weights are used to generate the summary statistics; the following results are all weighted results.

3.1.3. Introduction and Definitions

Section 1 serves as an introductory portion of the survey to ensure that survey participants are actively reading and retaining pertinent information that will improve responses throughout the rest of the survey. The introductory section includes a basic introduction to the survey purpose, the scope of self-driving vehicles posited in survey scenarios, and the definitions of different long-distance trip types. No results are determined from this section.

3.1.4. Long-Distance Trip Frequency and Purpose

In Section 2 of the survey, respondents revealed how their trip-making behavior changed between the years 2019 and 2020, regarding the type of trip, mode of transportation, and frequency of travel, with additional consideration of how the COVID-19 pandemic impacted these responses. Table 4 shows the demographic distribution of the respondents who made long-distance non-business and business trips during 2019 and 2020, for the Texas region and the whole US. The year 2019 was not impacted by the COVID-19 pandemic, unlike the year 2020. Business and non-business trips suffered equally from the impacts of COVID-19, so the share of each does not change. In both years, more non-business trips were made compared to business trips, with a ratio of 3:2 in Texas and 2:1 in the US.

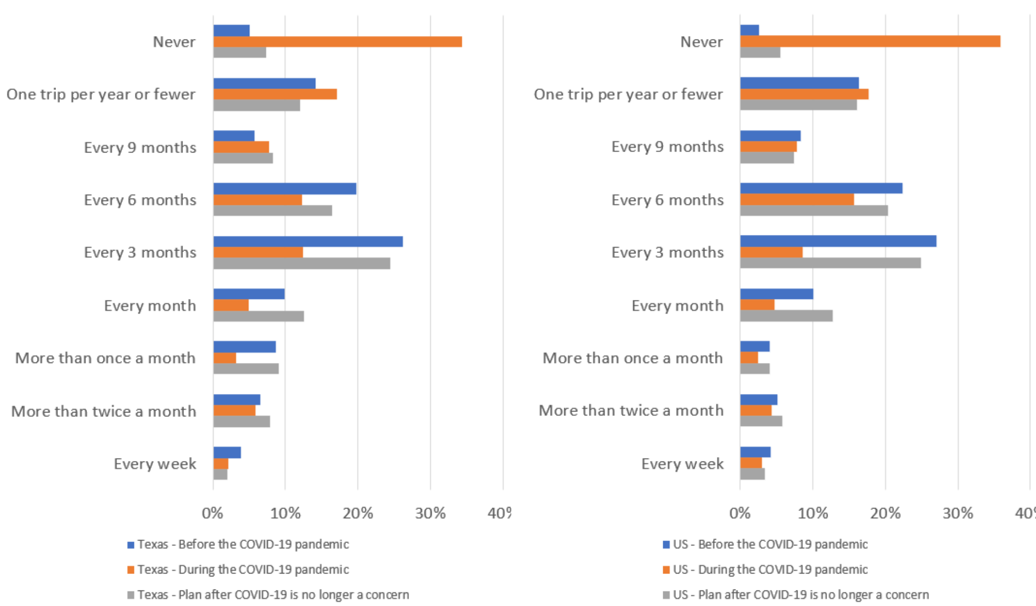
Table 4. Demographic distribution of long-distance trip occurrence in 2019 and 2020

	Texas				US			
Trip purpose	Non-Business		Business		Non-Business		Business	
Calendar year	2019	2020	2019	2020	2019	2020	2019	2020
	61%	60%	39%	40%	66%	67%	34%	33%
Gender								
Female	37%	35%	34%	25%	52%	51%	34%	38%
Male	63%	65%	66%	75%	48%	49%	66%	62%
Age								
18 to 24 years	12%	14%	22%	33%	5%	6%	5%	7%
25 to 34 years	26%	29%	27%	22%	26%	27%	52%	55%
35 to 44 years	25%	26%	21%	18%	15%	17%	12%	13%
45 to 54 years	17%	18%	17%	21%	15%	18%	19%	20%
55 to 64 years	10%	8%	9%	2%	19%	20%	9%	4%
65 or more years	11%	6%	3%	4%	20%	11%	3%	1%

	Texas		US			
Trip purpose	Non-Business	Business	Non-Business		Business	
Census Region (Appendix D)						
Northeast			28%	26%	14%	15%
Midwest			12%	11%	5%	4%
South*			31%	35%	34%	41%
West			30%	29%	47%	41%
Texas			9%	11%	11%	15%

*Note: Southern US here includes Texas.

An assessment of the survey results by demographics revealed that, compared to women, men made more long-distance trips in Texas for both non-business and business trips in 2019 and 2020. In particular, Texas men's share of business long-distance trips relative to women's increased in 2020 compared to 2019, a shift likely the result of the pandemic. At the national level, this difference was not apparent for non-business trips: about the same share of men and women made non-business trips, a proportion unaffected by the pandemic. Interestingly, when moving the lens from Texas to the US, more women made long-distance business trips than men during the pandemic. In terms of the age distribution, adults younger than 24 years old in Texas made more long-distance trips compared to the US overall, but adults older than 55 in Texas made fewer long-distance trips than the US average. In the US, travel of young adults were less affected by the pandemic while the population older than 55 was more affected. About a quarter of the long-distance trip-makers from the southern US were Texans. Figure 15 charts the pandemic's effects on long-distance trip frequency.

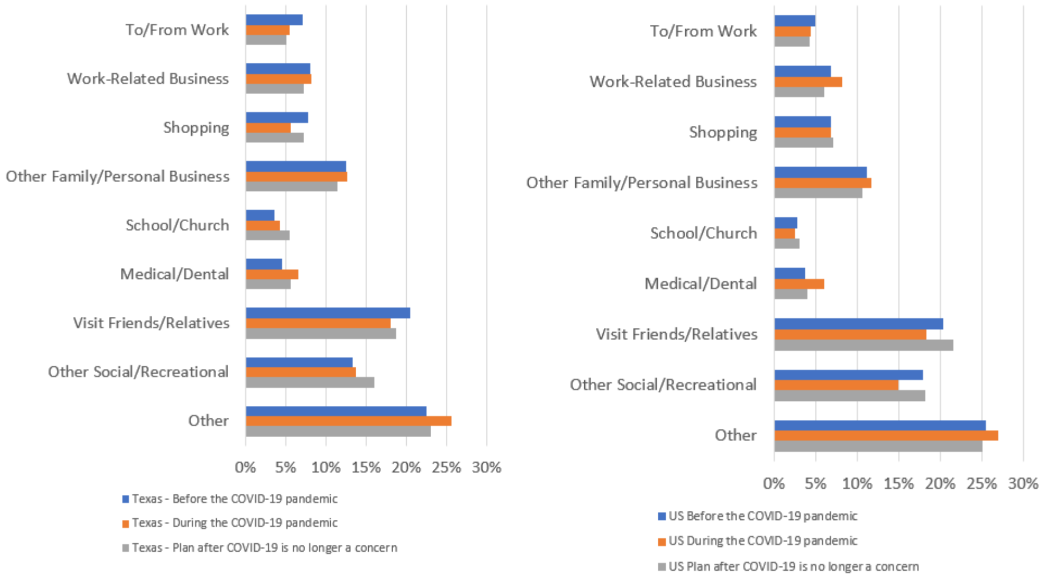


a) Texas region b) The US

Figure 15. Long-distance trip frequency under the impacts of COVID-19

COVID-19 had a clear impact on trip-making, as seen from the increase in the population that did not make a single long-distance trip during the pandemic, as of the survey. Almost one-third of the US population did not make long-distance trips during the pandemic, the same proportion in the Texas region. Before the pandemic, a frequency of one long-distance trip per season was the most common, followed by one long-distance trip every half year. During the pandemic, a huge reduction was observed for those who used to take two to six long-distance trips per year. The situation is expected to be mitigated after the pandemic for both the US and the Texas region, and more people will make 12 to 24 long-distance trips, even compared to before the pandemic. However, fewer people will commute long-distance weekly, and more people will prefer to stay in their home region all the time.

Figure 16 shows the trip purposes of all long-distance trips that people made before and after the pandemic, as well as their long-distance travel plan when the pandemic is no longer a concern. The trip purpose information is collected only for those who made long-distance trips during the pandemic. Since this figure does not involve the trip frequency associated with the purposes, the work-related trip shares are underrepresented. Therefore, the insight from this figure mainly lies in the change in trip purposes under the pandemic. During the pandemic, Texans' long-distance work trips and shopping trips decreased, while trips for school and church, medical and dental, and recreational purposes increased. COVID-19 policies varied between states during the pandemic, and Texans' long-distance travel has shown a different pattern compared to the average statistics in the US. At the national level, long-distance trips for school and church, visiting friends and relatives, and other social and recreational purposes all fall, but people's plans show that the trend will recover to levels observed before the pandemic, with an even higher trip rate.



a) Texas region b) US

Figure 16. Long-distance trip purpose under the impacts of COVID-19

Respondents also indicated their primary travel mode for long-distance trips between 75 and 500 miles and over 500 miles. Personal car is the main mode choice for long-distance trips between 75 and 500 miles, especially for non-business purposes. Business trips are more often time-constrained and typically subsidized by employers, so airplane mode is used more. Compared to US travelers, Texans relied more on personal vehicles for both business and non-business trips, resulting in their larger mode share. Since Texas is a coastal state, boats and ships were also used more, generally, compared to regions of the US with inland states. Figure 17 charts the long-distance trip mode share. Overall, COVID-19 had a clear impact on trip-making, as the number and frequency of trips decreased, and people were more likely to use personal modes of transportation rather than shared modes like airplanes and trains.

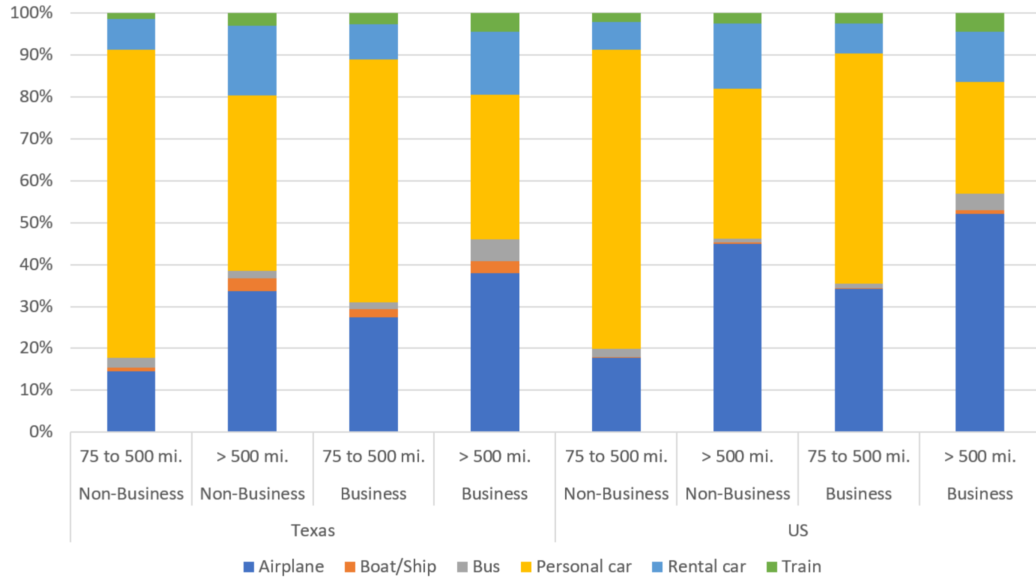


Figure 17. Long-distance trip mode share

3.1.5. Self-Driving and Shared Vehicle Technology Definition and Activity Preferences

This section of the survey further reminds respondents of the automation technology definition and the potential for ride-sharing services in the form of an SAV fleet. Texans' attitudes towards pursuing different activities in a self-driving car are also investigated (Figure 18). The survey found that passengers of a self-driving car were most likely to spend their time watching the landscape, listening to music, and eating or drinking. This is consistent with Lenz's (2016) findings, which indicated that users would most likely use the time to enjoy the landscape and talk to other passengers, and be least likely to work, as opposed to Das et al.'s (2017) study, which found that users would most likely use the time to perform tasks related to their main job. Some studies also argue that, based on how we see transit users engaging in non-work activities to simply pass time, we may not see AV users devoting their in-vehicle time to work (Singleton, 2019). Since the Texas region and the US show a very similar pattern in terms of activity preference, the US pattern is not shown here.

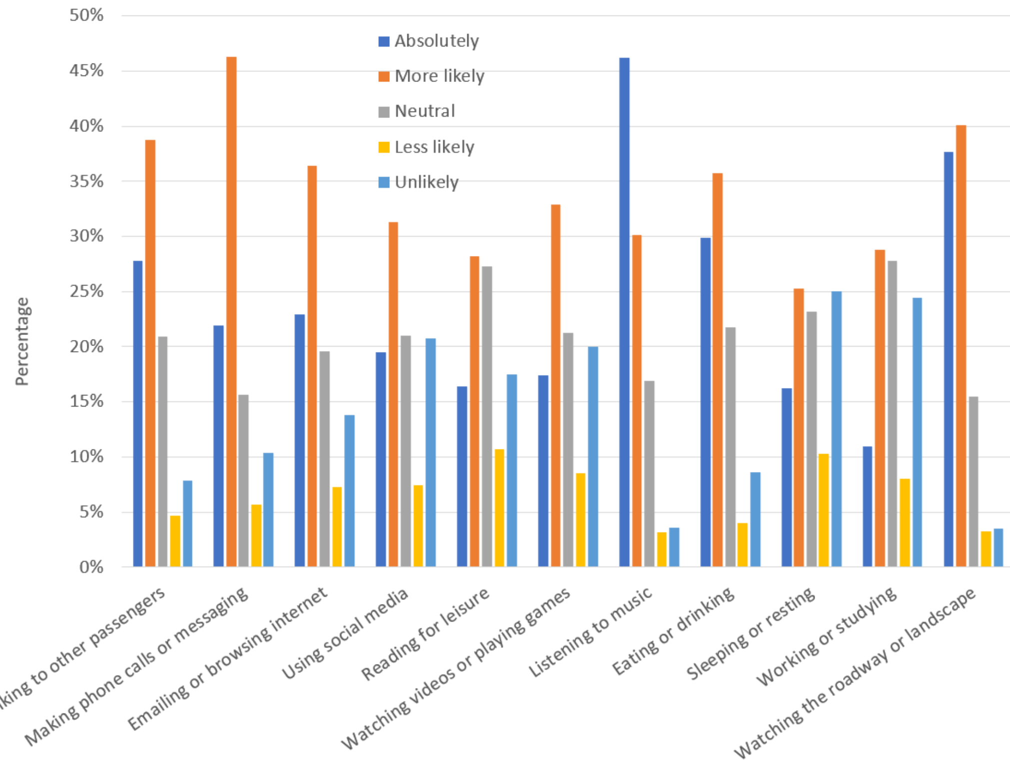


Figure 18. Texans' likelihood to perform different activities in a self-driving car

3.1.6. Long-Distance Trip Revealed and Stated Preferences

The next section of the survey asked respondents to answer a series of questions relating to a specific long-distance trip taken before the pandemic. This portion revealed crucial information about trip duration, trip chaining, and time and financial expenses for all modes taken to complete the trip. All of these details are critical to our modeling and prediction of future travel and its impact on the market share of various modes. Respondents were also asked to consider a hypothetical scenario where this trip is made with an AV. By offering various options of costs and time savings, the survey revealed respondents' perceptions of how AVs would change their trip in terms of the amount paid, duration of travel, duration of stay, and party size.

Figure 19 shows the departure date of the long-distance trip taken by the respondent before the pandemic. Most of the long-distance trip-makers traveled in 2019 and early 2020, but some of the trips described occurred in 2018. Most of the travel occurred in summer (around July). The trip destinations are shown in Figure 20. New York, California, Texas, and Florida were the top four destinations, perhaps due to some of their attractive coastal locations; Florida was the most popular overall.

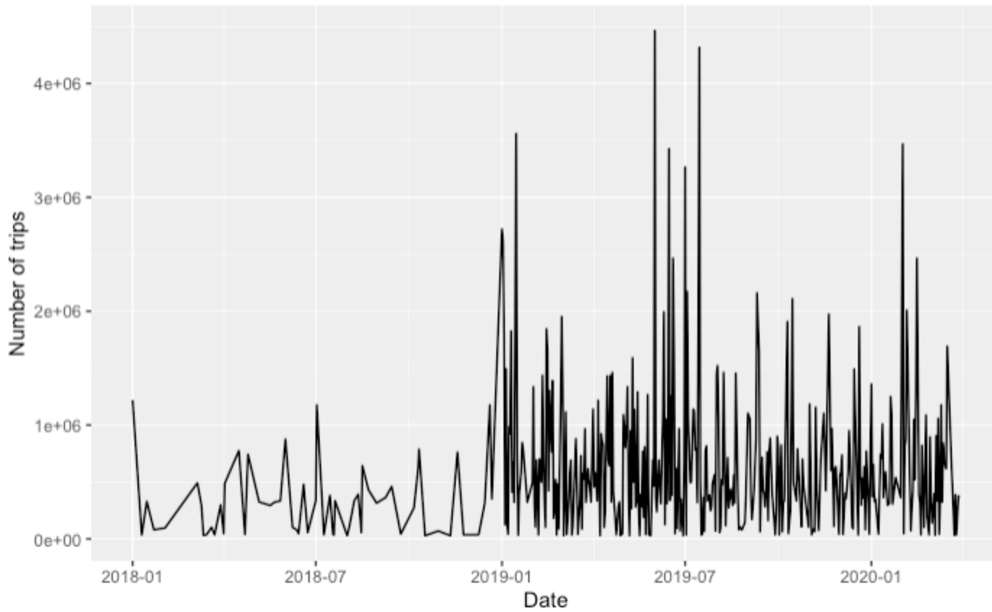


Figure 19. Long-distance trip departure date

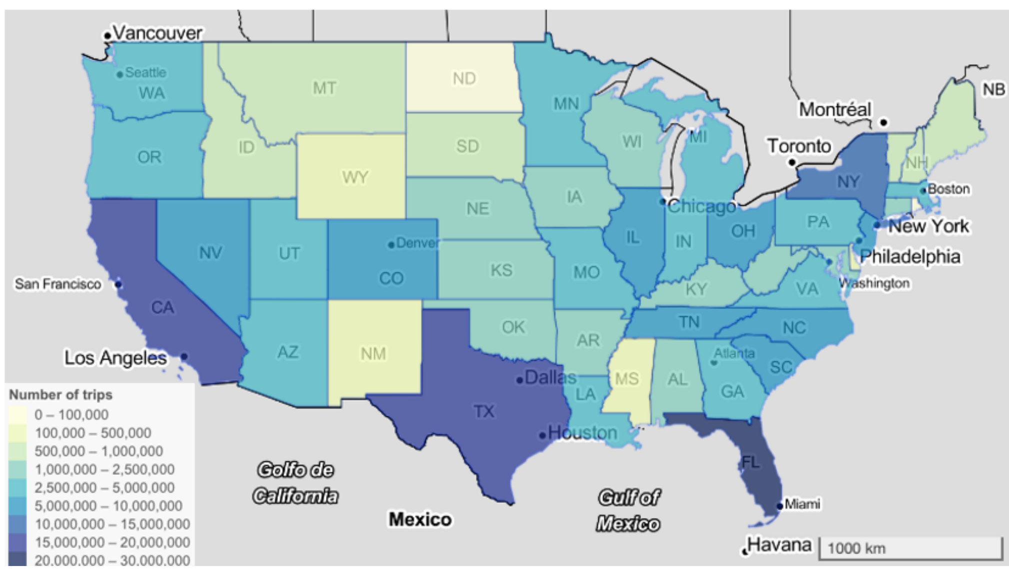


Figure 20. Long-distance trip destinations

In Texans' responses, about 73.5% of the long-distance trips were round trips, while chain trips accounted for 15.7%, followed by one-way trips (10.7%). This is similar to the broader US, which had a few more round trips (76.3%) but fewer chain trips (13.1%). Among these long-distance trips in the US, 47.5% were shorter than 500 miles, 26.1% were between 500 and 1,000 miles, and 26.3% were longer than 1,000 miles. Compared to the US overall, Texans made more long-distance trips between 75 and 500 miles (50.5%), with fewer trips longer than 500 miles, possibly due to more commuting trips among the four major Texas metro areas: Dallas Fort-Worth, Houston, San Antonio, and Austin.

For the specific long-distance trip that respondents were asked to recall, they were more likely to think of a non-business trip, particularly trips visiting friends and families, as well as vacations (Figure 21).

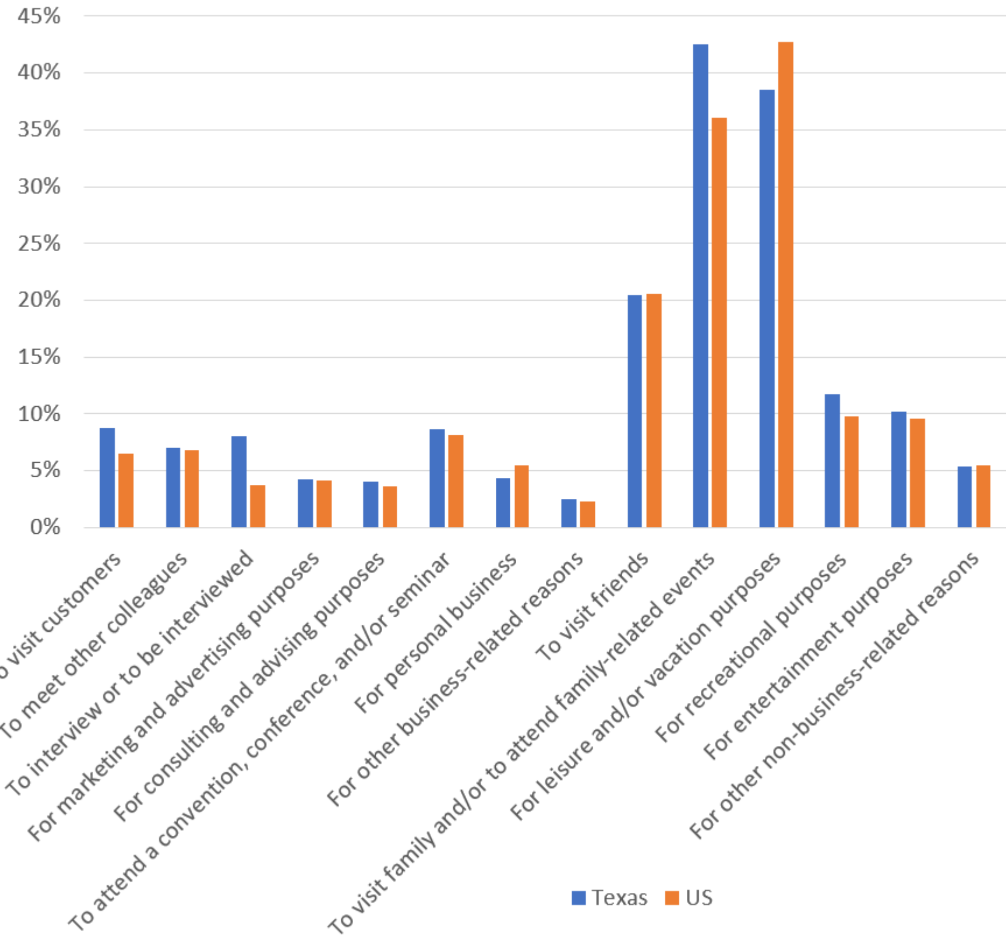


Figure 21. Trip purposes across Texas and US

As Figure 2 indicated, non-business trips are more frequent than business trips, but the sample may bias toward the long-distance trips that can be easily recalled. This survey is different from the NHTS in that respondents need to offer the trip details for a specific day. The mode choice pattern for this specific long-distance trip follows the general pattern that was obtained for 2019 and 2020. Texans favored personal cars, rental cars, and boats for long-distance trips more than the rest of the US (Figure 22). However, air travel takes a larger mode share in the US compared to in Texas, because the US is much larger than Texas, allowing more origin-destination (OD) pairs that are far away and thus favoring airplanes.

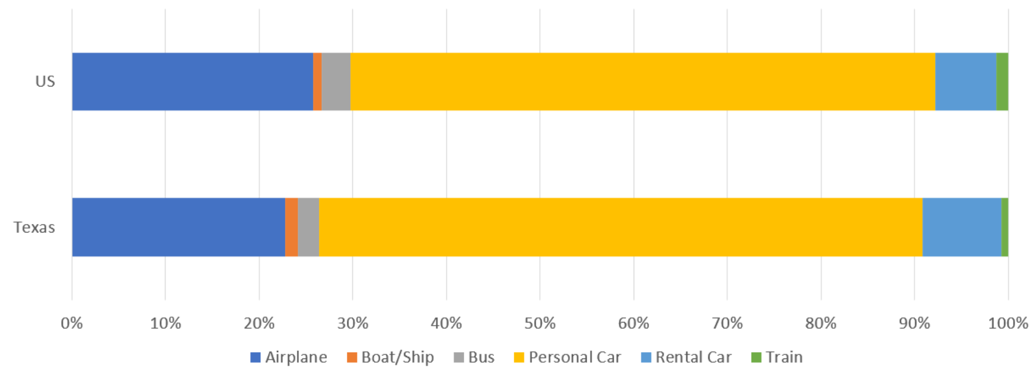


Figure 22. Long-distance trip mode choice

For those who took an airplane as the primary travel mode for this specific trip, the travel times of different legs of the trip were collected. The average travel time within Texas was about 6.7 hours, which was about 1.2 hours shorter than the average long-distance travel time across the US (a figure that encompasses door-to-door time elapsed, not merely the in-air component). As Table 5 indicates, time onboard planes accounted for just over half of the total travel, since a great deal of time is spent accessing, waiting at, and egressing the airport.

Table 5. Average time spent on different legs for air travel

	Texas	US
Time scheduling the trip to the airport (e.g., reserving a van or calling Uber/Lyft, renting a car)	0.36 hours	0.38 hours
Time traveling to the airport (driving or being driven by someone else)	0.60	0.65
Time parking at the airport	0.18	0.13
Time spent going through airport security	0.37	0.38
Time waiting at the airport	0.96	0.87
Airplane onboard time	3.38	4.66
Time scheduling the trip from the airport (e.g., reserving a van, calling Uber/Lyft, renting a car)	0.17	0.22
Time traveling from the airport to your destination (driving or being driven by someone else)	0.66	0.57
Time parking at your destination	0.10	0.08
Total	6.67	7.86

Table 6 and Figure 23 show the respondents' willingness to use AVs for long-distance trips under different travel time and travel cost assumptions. Table 6 shows four different travel time assumptions to ascertain the willingness to use AVs, and a longer travel time question was asked only if the respondent replied "Yes" or "Maybe" to a shorter travel time scenario. Looking at the "Yes" answers reveals that Texans were about 40% more willing to travel long-distance with AVs than the US respondents. Texans are even willing to spend more time traveling with AVs; if the travel time involved increased by 50%, 15% of Texans would or may choose an AV as opposed to only 9% of US respondents. Therefore, Texans seem to perceive more benefits from freeing their hands for long-distance trips.

Table 6. Willingness to use AVs for long-distance trips by the change in travel time

Travel time assumption	Texas				US			
	No change	10% increase	25% increase	50% increase	No change	10% increase	25% increase	50% increase
Yes	38%	20%	12%	8%	28%	13%	8%	5%
Maybe	26%	18%	12%	7%	29%	16%	9%	4%
No	27%	22%	10%	7%	33%	24%	8%	5%
I do not know	9%	5%	3%	2%	10%	5%	4%	2%

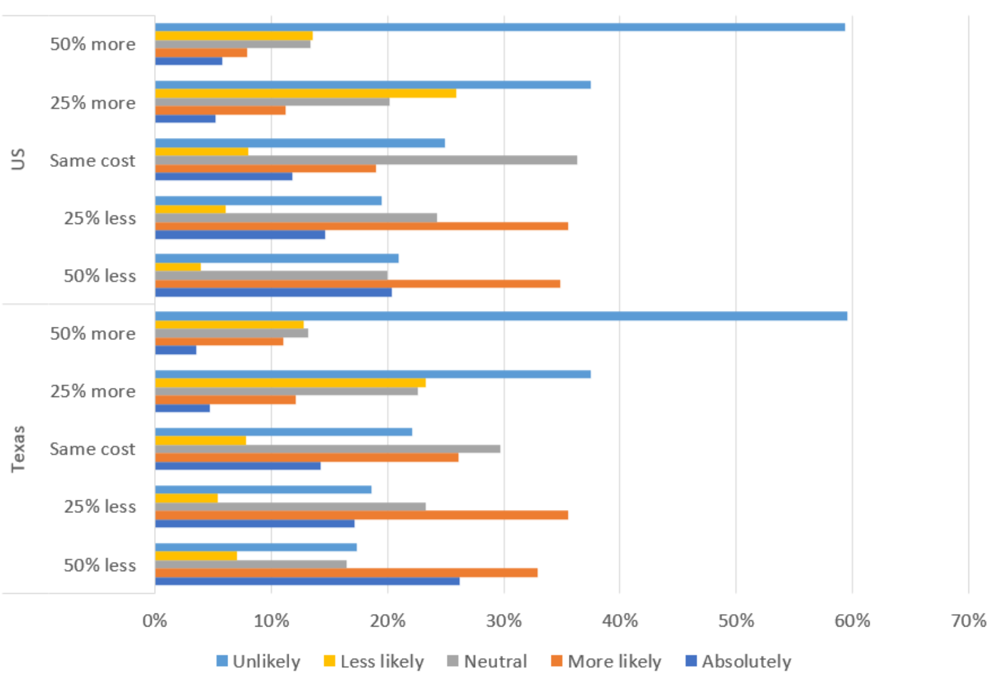


Figure 23. Willingness to use AVs for long-distance trips by the change in travel cost

In terms of the cost variations, Texans reacted similarly to US respondents when the long-distance trip cost either 25% or 50% more. For example, if the AV long-distance trip costs 50% more compared to traveling with a human-driven vehicle (HV), about 60% of Texans and the US population were “unlikely” to travel in an AV. However, Texans showed more willingness to use AVs for long-distance travel when the cost remains the same or becomes lower. More than 40% of Texans were “absolutely” or “more likely” to use AVs under the same cost scenario. The share went up to almost 60% when AVs dropped to half the cost.

Figure 24 and Figure 25 show respondents’ destination choice and willingness to include more stops during long-distance travel. This factor was included because the introduction of AVs may create changes in how travelers structure their long-distance travel. For example, long-distance trip-makers may want to make more stops (such as for leisure or family visits) because AVs lighten the driving burden and also minimize the time otherwise needed for drivers to take a break. AVs may also make it possible for long-distance trip-makers to travel to destinations farther away than they would consider when traveling using a human-driven vehicle. And since AVs can drive overnight, people may just stay in the AV to avoid another overnight stop at hotels, which would reduce the number of involuntary stops along the way. The results show that almost half of the respondents expressed a willingness to make more stops during the long-distance trip if an AV were used. In contrast to Texas respondents, the US population was more likely to make the same number of stops along the way, or even have fewer stops. However, destination choice is robust, because almost 70% of the population would not change their destination with an AV. Similarly, 60% of Texans would not change their destination and about 20% would change to visiting a destination farther away.

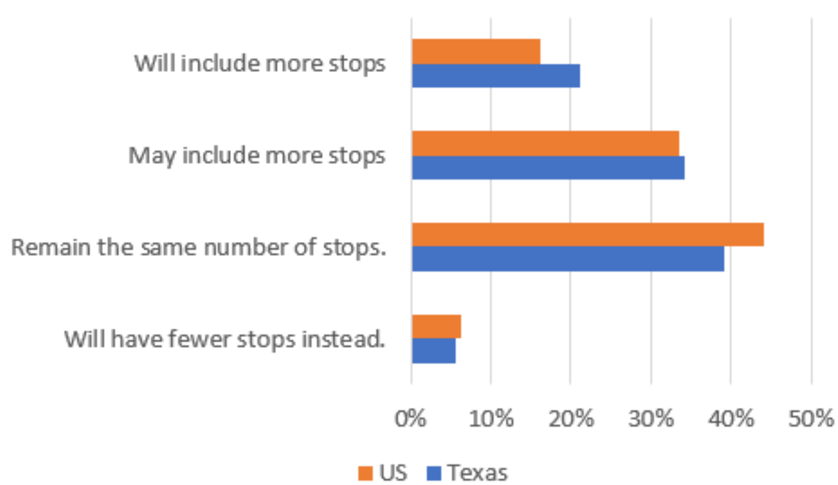


Figure 24. Willingness to include more stops in long-distance trips with AVs

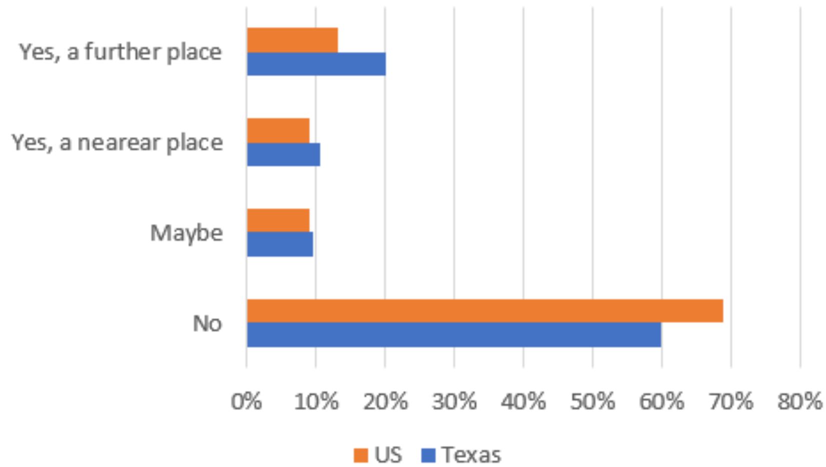


Figure 25. Change in destination choice for long-distance trips with AVs

Trip duration was also investigated in terms of whether respondents would like to extend or shorten their stay, due to the flexibility that AVs can offer. Although many people would like to include additional stops along the way, at least 60% (in both the US and Texas) do not want to shorten or extend their stay, perhaps due to the time constraint on the vacation or other reasons. However, Texans are more flexible in their schedule; almost 40% are willing to extend their stay, probably due to the convenience that an AV can provide (Figure 26).

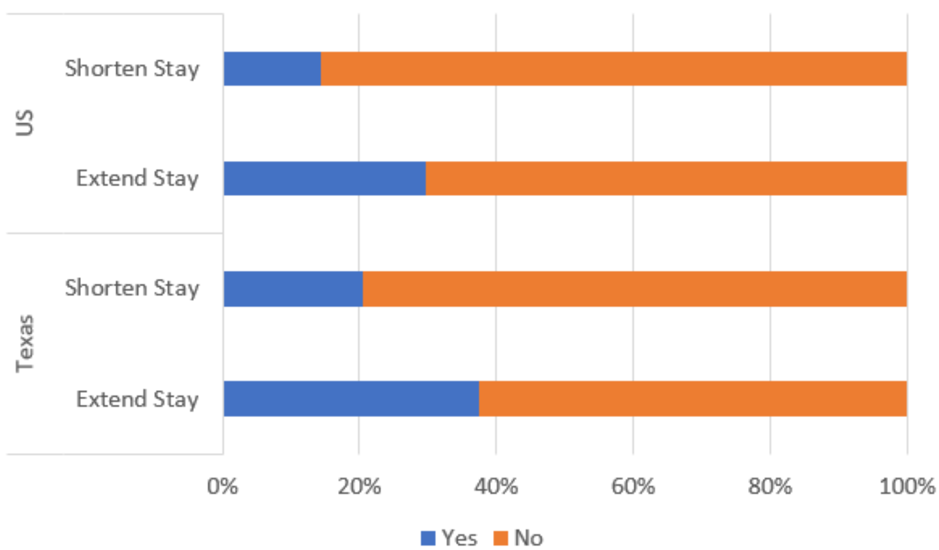


Figure 26. Willingness to change stay duration in long-distance trips with AVs

Figure 27 further shows the travel duration distribution of respondents' long-distance trips. The "0 day" point on the axis means it is an overnight trip. The majority of long-distance trips lasted less than one week (83% in Texas and 81% in the US). In Texas, more long-distance trips lasted 5 days or less, and fewer trips lasted more than 6 days. Few long-distance trips between two weeks and one month were observed. Of trips listed as taking over one month, which make up roughly 5% of trips, this may include one-way trips due to home relocation as well as temporary jobs.

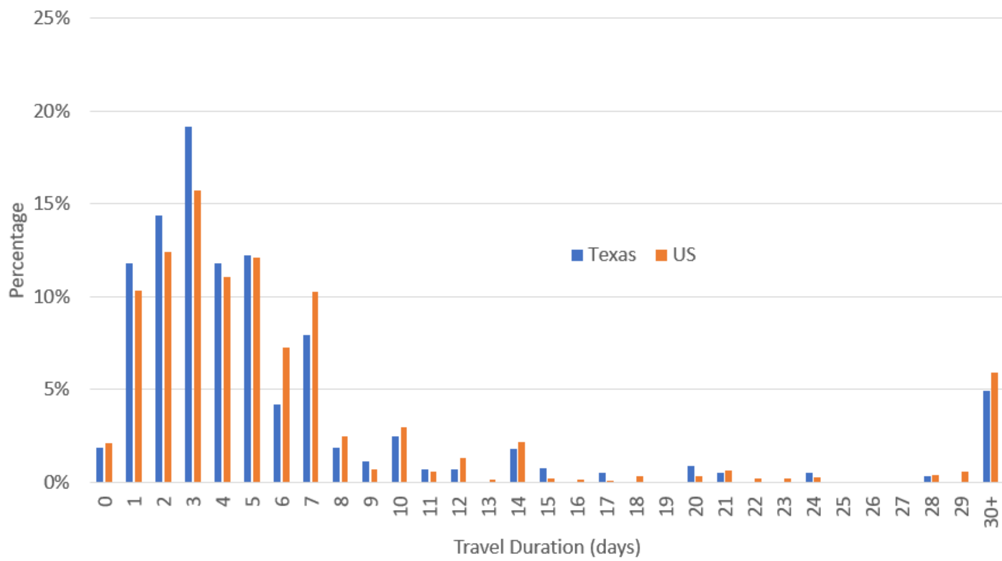


Figure 27. Long-distance travel duration

Travel party size is another key feature of long-distance travel, but one rarely captured by prior surveys. Table 7 shows that both the US and Texas populations were more likely to travel with family members and friends for non-business trips. The average party size was larger in Texas than in the US in general, with the average companion number including 30% more family members and a roughly doubled party size of friends and colleagues or associates. For long-distance travel both in the US and Texas, the most common traveling party size was two, followed by traveling alone. Figure 28 indicates that Texas had a smaller share of party sizes of one and two compared to the rest of the US, with more long-distance trips of three or more people traveling together.

Table 7. Travel party size by traveler type

	Texas			US		
	Family members	Friends	Colleagues and/or associates	Family members	Friends	Colleagues and/or associates
Travel without	29%	79%	90%	33%	83%	95%
Travel with	71%	21%	10%	67%	17%	5%
Average # to travel with	2.03	0.75	0.62	1.53	0.39	0.28

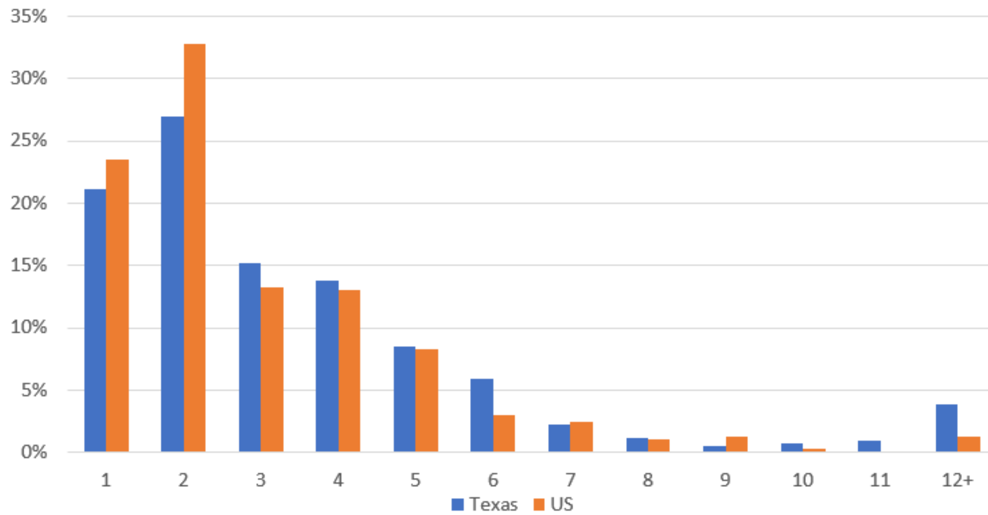


Figure 28. Travel party size distribution

Since the majority of the long-distance trips were non-business family trips, children were often involved. Results show that more than 40% of families traveled with at least one child. Texas had a larger party size of children, compared to the general case in the US (Figure 29). With automation technology, households may also bring more children since AVs help in better attending to them. About 16% of both Texans and the US population (including those who do not have children in the household) would be more likely to travel with an AV. Shifting the focus from children to anyone they would like to travel with, 32% of Texans would travel with more people in an AV for long-distance trips, while only 22% of the US population would increase their travel party size.

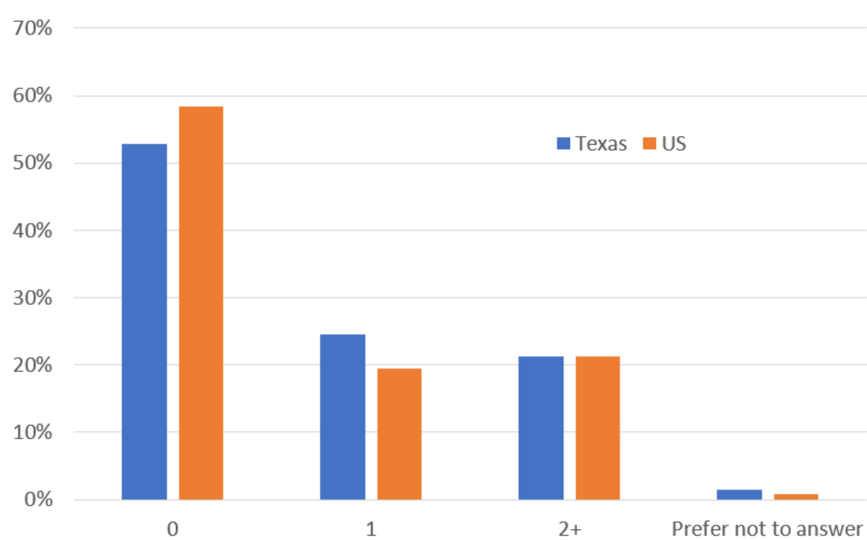


Figure 29. Travel party size of children

The survey also asked about reasons for respondents' preference for and against taking long-distance trips by AVs (Figure 30 and Figure 31). People who would like to use an AV for long-distance travel enjoyed the safety most, followed by the reliability. The convenience offered by AVs came next, which was valued more than their ability to self-park. However, safety was also the main reason that people opted *not* to use AVs for long-distance travel, citing concerns about the potential for faulty software. Interestingly, enjoying the act of driving was also another key point for those not wanting to travel in an AV, even though driving for long periods of time may be tiresome and tedious.

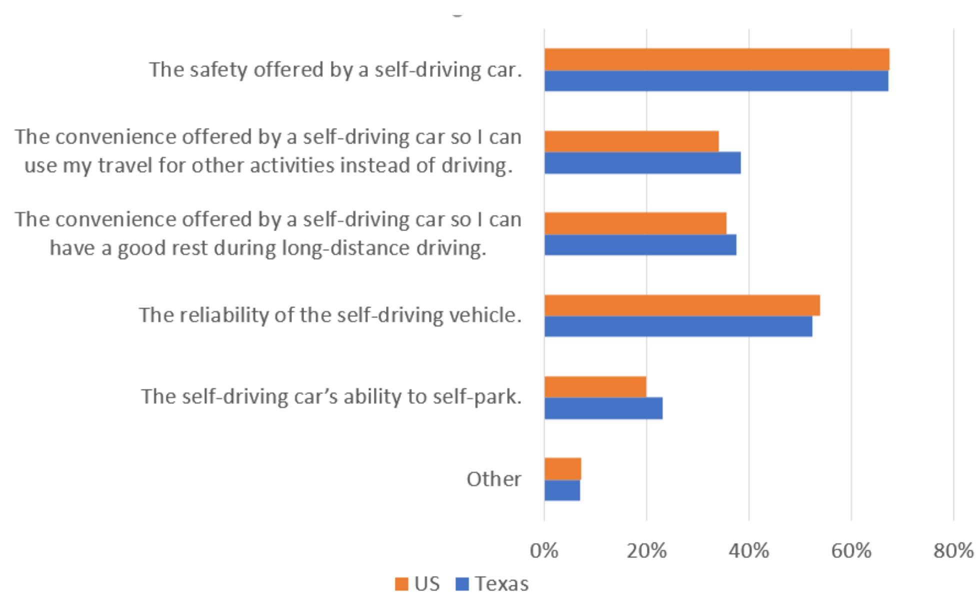


Figure 30. Reasons informing preference for long-distance trips using AVs

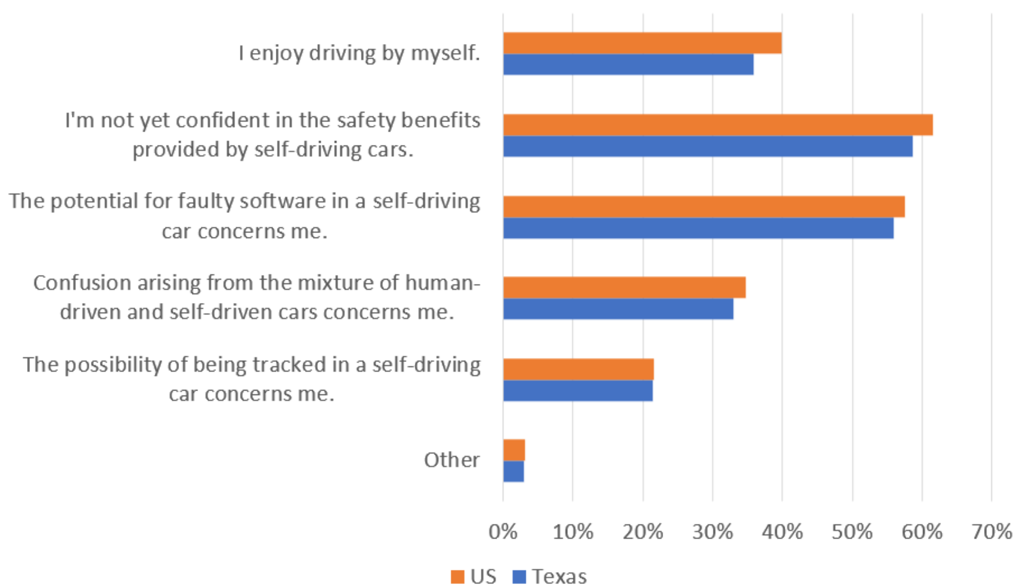
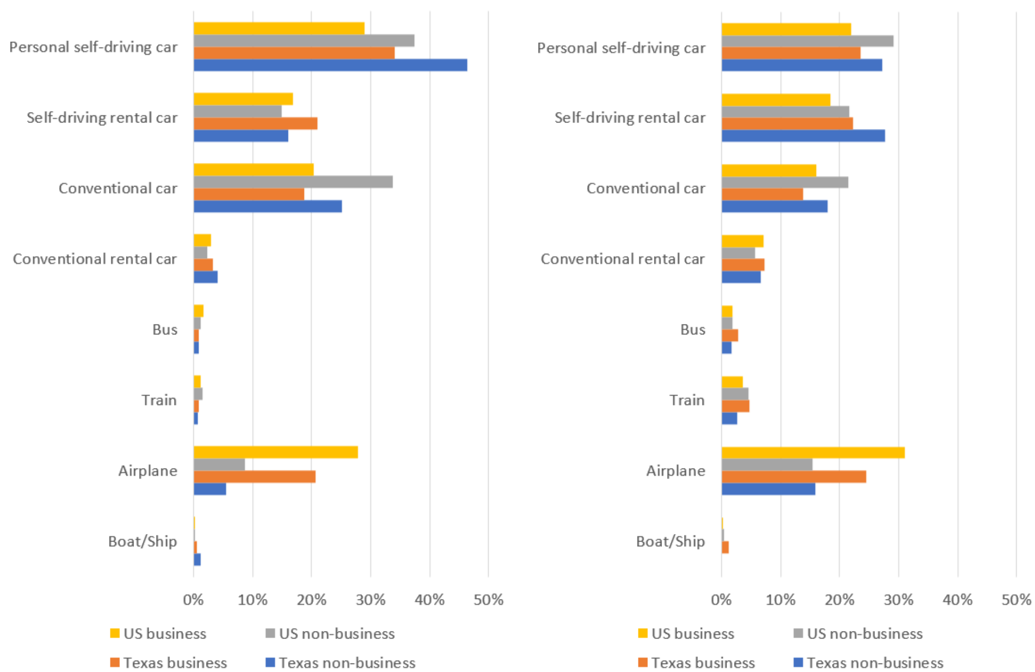


Figure 31. Reasons informing preference against long-distance trips using AVs

3.1.7. Future Scenario Questions about Long-Distance Travel

In this section of the survey, respondents were given a future scenario in which AVs are widely available and affordable. The questions were designed to provide insights on how this future mode choice of a self-driving vehicle would impact the frequency, duration, distance, destination, and departure time of possible long-distance trips.

Figure 32 shows respondents' mode choice considering two new AV choices: personal AVs and rental AVs. For long-distance trips shorter than 500 miles, personal self-driving cars dominate the market in the hypothetical scenario, for both business and non-business trips across the US and Texas. The conventional car is the second choice, in general, after the personal self-driving car. Air travel is more popular for business trips, compared to non-business trips. In terms of rental options, respondents preferred AVs significantly over conventional cars.

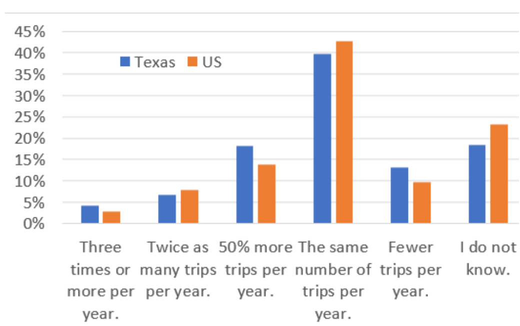


a) Trips between 75 and 500 miles *b) Trips longer than 500 miles*

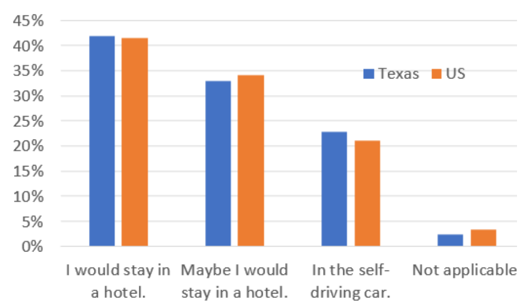
Figure 32. Mode choice for long-distance trips with AV choices

For long-distance trips that exceed 500 miles, the airplane is the most popular mode for business trips in the US, followed by personal AVs and rental AVs. For non-business trips in the US, the self-driving car is still the first choice. However, the pattern is different in Texas. The rental AV is the top choice for non-business trips, while airplanes, personal AVs, and conventional rental cars are all very popular for business trips. In contrast to long-distance trips shorter than 500 miles, the mode share for trips exceeding 500 miles decreased for personal cars but the share of conventional rental cars, buses, and trains all increased, which is expected because of the burden and cost of driving for such a long distance.

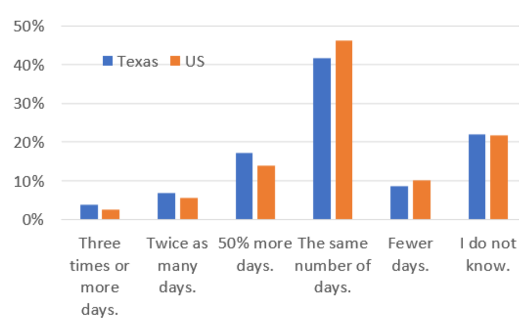
Figure 33 presents six different aspects of long-distance trip-making preferences when traveling in AVs. Figure 33a depicts the trip frequency preference if the respondents can travel with an AV. The majority chose to make the same number of trips per year. About a quarter of the population would like to make more trips per year, while about 12% of the population would make fewer trips. This could owe to their safety concerns about traveling in an AV or the presence of other AVs on the road. About 20% of respondents did not offer an answer, as they may be still uncertain about the exact changes that AVs may bring. US and Texas respondents presented a similar pattern, but more Texans are likely to make more long-distance trips per year. Figure 33c shows the changes in long-distance trip duration that respondents would make. The pattern is similar to Figure 33a, where over 40% would remain the same, with more Texans enjoying longer trip durations. Figure 33e illustrates a similar trend in trip distance. The pattern corresponds to when the respondents were asked about their preference of using AVs for their recalled trip in the previous section.



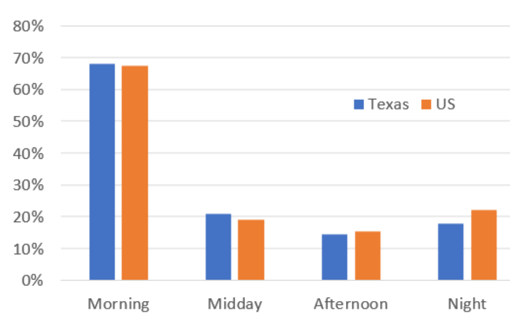
a) Trip frequency



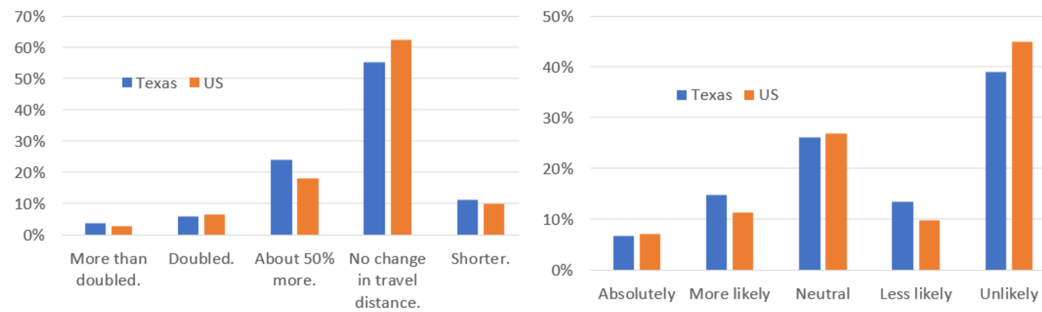
b) Overnight stay decisions



c) Trip duration



d) Departure time



e) Trip distance *f) Willingness to share rides*

Figure 33. Long-distance trip-making preference with AVs

Figure 33b shows the respondents' decisions about where to stay overnight when AVs are available. With an AV, one can just stay in the car overnight while it drives to the destination. However, about 40% of the population still preferred to stay in a hotel, although 50% would at least possibly remain overnight in self-driving cars. Figure 33d shows the departure time choice with AV travel. Although morning is the top choice, night was preferred to the afternoon, a more congested and busy time. Driving at night is a challenge for people who suffer from night vision problems, but AVs are anticipated to have technology that adequately supports night travel. The last question asked is about people's willingness to share a ride with someone they do not know under a social-distancing policy during a pandemic like COVID-19. As Figure 33f shows, over 40% of respondents would not like to share the ride, while about 20% may share. Texans were slightly more positive about sharing rides, compared to the broader US.

3.1.8. Limitations

The survey has demonstrated many useful and interesting results that help anticipate Americans' and Texans' long-distance travel choices. However, some limitations exist in the survey design and data collection process.

The survey respondent pool of 1,004 represents a small sample of both Texas and US residents and has been scaled proportionally to represent the entire state and country. These responses may include outliers despite all efforts to be as representative as possible of the population. More samples can help reduce the sample bias, which also means a higher cost for the data collection.

Due to the total time constraint on the survey questionnaire, only approximately 70 questions were asked. Considering the multiple topics involved, including automated technology, the COVID-19 pandemic, and long-distance travel, additional questions would help the team discover more in-depth results but would also increase the burden for respondents and thus produce an undesirable response quality.

The limited time to collect data also prevented the team from collecting survey responses that perfectly match the US and Texas demographics. Allowing more time to collect responses would help in collecting samples from demographic groups that are difficult to reach, but the benefit is marginal. The team chose to apply the IPF method to weigh the samples.

Presenting this survey in the unusual atmosphere of a pandemic may lead to bias in the responses, especially the stated preference questions. For example, those respondents who are still worried about the pandemic may be too conservative about long-distance travel. The team has made efforts to diminish such bias across the survey questions by clearly stating the background assumptions and ensuring the stated preferences are matched to trips in the pre-pandemic period.

Assembly of Other Datasets

Assembly of datasets is a vital step in creating a long-distance travel demand model that leverages this project's survey model. Because the LD-AV survey represents a subsample of the entire population, it is necessary to use other resources to properly model the entire population. After the data sources and models are identified, and their scales and caveats are understood, it is also necessary to consider how to combine different data sources using scientific, repeatable processes. Here, models using these components will be finalized, including identifying parameters for those models that will be true to the subsamples found in this and other projects' surveys on long-distance travel, vehicle ownership, and AV usage.

This section is organized to first introduce the model structure, followed by further explorations of each major dataset that feeds into it. This document then touches on prior work around models that predict AV ownership and usage, an important aspect that is also challenging to predict. Finally, a separate exploration is made into freight models to support the creation of an AV long-distance freight model.

4.1. Model Structure

The modeling framework of this study begins with a population generation stage, feeding into travel choice and then destination and mode choice. The final outcome is a set of disaggregate long-distance trips that represent all long-distance travel of a future US population that lives with a market where AVs are readily available as a viable mode choice. A separate freight module produces disaggregate long-distance freight trips.

While the Texas Department of Transportation (TxDOT) focuses on long-distance travel among Texans, the researchers found it important to model long-distance travel for the entire US due to the pervasive interconnections between their travel networks. This is because of the significant number of trips where Texans travel outside of the state, visitors come to Texas, and non-Texans pass through Texas. This also provides opportunities to contrast Texans' travel patterns with those of the rest of the US. When the nationwide model is completed, it will be straightforward to analyze results that affect Texas with an anticipated higher degree of accuracy.

A microsimulation model requires a representation of the full population of interest (i.e., the population of the United States). The PopGen synthesizer was used to generate a synthetic population. ACS Public Use Micro Sample (PUMS) data will be used as the population seed. Individuals within the synthesized population will be allocated according to the rJourney National Use Microdata Area (NUMA) travel analysis zone system, allowing us to adopt rJourney data and validate spatial patterns against their results. (Further detail about rJourney and its NUMA zone system is found later in this chapter).

The next stage of the model will represent a set of "pre-trip" decisions by households. Such processes include the decision to participate in long-distance travel, tour frequency over the course of a year, and party size for each tour. Results from the recently completed long-distance AV (LD-AV) survey by the project team will be combined with NHTS and regional long-distance travel data to estimate and validate these models.

The final stage of the model comprises destination and mode choice for each long-distance tour. The mode choice will be conditional upon a household AV ownership choice model. An existing AV ownership model for the US developed by the project team in 2017 will be updated for this purpose. A 2021 survey of Americans' preferences for AVs will be used to adjust the model, as it asked a parallel set of questions to the 2017 study. Mode and destination choice models will use a variety of data sources to generate alternative attributes for model estimation. The DB1B database from the Bureau of Transportation Statistics (BTS) provides both a source for air travel attributes (e.g., fare, travel time, and number of legs) and a validation of total annual travel through expansion of its 10% sample. Other datasets include Amtrak route details and NHTS long-distance trip records.

4.2. Data Sources

This section introduces each of the major data sources that the new long-distance travel model will heavily rely on, identifying key attributes for each and caveats that must be considered for appropriate usage.

4.2.1. Long-distance Automated Vehicle (LD-AV) Travel Survey of Americans

As presented in Chapter 2 and Chapter 3, the survey consisted of 70 questions (15 to 25 minutes), divided into three main topics involving seven sections, targeting different aspects of long-distance travel, AV and SAV usage, and effects of the COVID-19 pandemic. It included a mix of revealed and stated preference questions for current or recent trips and future scenarios. Questions related to the effects of COVID-19 were also included, and different scenarios were tested for a future COVID-19-like virus to understand the possible impacts of future pandemics.

The survey started with definitions of relevant concepts such as self-driving vehicles, one-way trips, round trips, and long-distance travel before respondents were shown the questions. For this study, long-distance travel is defined as a one-way trip over 75 miles from the origin to the destination (or a round trip over 150 miles in total). The first main topic investigated respondents' general long-distance trip-making patterns during 2019 and 2020. This offered a detailed comparison of long-distance trip patterns that could reveal the impacts of the pandemic. The second main topic inquired about details of the most recent pre-pandemic long-distance trip made by the respondent, followed by questions focusing on how their travel behavior would change if they can travel with an AV. The third main topic provided future scenarios when AVs are widely used, exploring respondents' preferences for future long-distance trip-making. The survey then ended with collecting demographic information.

The response collection process lasted three weeks, with rigorous scrutinizing of the responses over time. The targets were set for 50% of the sample being Texan and 50% from the rest of US, with individual targets for both Texas and non-Texas samples concerning gender, age, census region, and education. The responses were analyzed during the collection process so that the targets were adjusted daily based on data cleaning results. The demographic distribution of cleaned data is shown in Table 3.

A total sample of 1,004 responses was obtained after the filtering and cleaning process described. The final pool includes 451 (45%) Texans and 553 (55%) respondents from the rest of the nation, allowing both a detailed representation of the Texas region and a comparison with the rest of the continental US. The survey data incorporates many variables that can be valuable to the long-distance passenger travel demand model. While the detailed summary is shown in previous sections, Table 8 here shows the list of key variables.

Table 8 Key variables in LD-AV survey

Long-Distance Trips Frequency & Trip Purpose		
Variables	***Data type***	***Value list***
# LD business trips made in 2019	Numerical	Positive Integer
# LD non-business trips made in 2019	Numerical	Positive Integer
# LD business trips made in 2020	Numerical	Positive Integer
# LD non-business trips made in 2020	Numerical	Positive Integer

Primary LD mode choice for LD 75–500 mi. in 2019 & 2020	Categorical	Personal car; rental car; bus; train; airplane; boat/ship
Primary LD mode choice for LD over 500 mi. in 2019 & 2020	Categorical	
Avg LD trip-making frequency before COVID-19 pandemic	Ordinal	Never; less than 1 trip per year; every 9 months; every 6 months; every 3 months; every month; more than once a month; more than twice a month; every week
Avg LD trip-making frequency during COVID-19 pandemic	Ordinal	
LD trip-making frequency after COVID-19 pandemic is no longer a concern	Ordinal	
LD trip purposes before COVID-19 pandemic	Categorical	To/from work; work-related business; shopping; other family/personal business; school/church; medical/dental; visit friends/relatives; other social/recreational; other; refused/don't know
LD trip purposes during COVID-19 pandemic	Categorical	
LD trip purposes after COVID-19 pandemic is no longer a concern	Categorical	
Long-Distance Trip Revealed & Stated Preferences		
Variables	***Data type***	***Value list***
Exact or estimated date of travel	Date	Before March 2020
Type of LD trip	Categorical	Round trip; one-way trip; chained trip
Trip purpose	Categorical	— Business: — To visit customers; to meet other colleagues; to interview or to be interviewed; for marketing and advertising purposes; for consulting and advising purposes; to attend a convention, conference, and/or seminar; other — Non-business: — To visit friends; to visit family and/or to attend family-related events; for leisure and/or vacation purposes; for recreational purpose; for entertainment purposes; other
Origin city, state, country	Text	
Destination city, state, country	Text	
Estimated LD trip distance	Ordinal	75 mi. to 500 mi.; 500 mi. to 1,000 mi.; 1,000 mi.+
Primary mode choice	Categorical	Personal car; rental car; bus; train; airplane; boat/ship
Days staying at destination	Numerical	Positive integer
Days staying at destination if AVs are available	Numerical	Positive integer
Travel party size, including family members, friends, and colleagues	Numerical	Non-negative integer
# Children on the trip	Numerical	Non-negative integer
Future Scenario of Long-Distance Travel		
Variables	***Data type***	***Value list***
Primary mode choice for one-way LD business trip 75–500 mi.	Categorical	Personal self-driving car; self-driving rental car; conventional car; conventional rental car; bus; train; airplane; boat/ship; N/A

Primary mode choice for one-way LD non-business trip 75–500 mi.	Categorical	
Primary mode choice for one-way LD business trip over 500 mi.	Categorical	
Primary mode choice for one-way LD non-business trip over 500 mi.	Categorical	
Change in LD trip-making frequency if AVs are available	Ordinal	Fewer; about the same as current situation; 50% more as many/frequently/far as current situation; twice as many/frequently/far as current situation; three times as many/frequently/far as current situation
Change in LD total journey duration for leisure if AVs are available	Ordinal	
Change in LD travel distance if AVs are available	Ordinal	
Departure time preference if AVs are available	Categorical	Morning; midday; afternoon; night
Overnight stay decision change if AVs are available	Ordinal	Stay in a hotel; maybe stay in a hotel or self-driving car; stay in a self-driving car
Likelihood to ride a self-driving car with strangers for a reduced price under social-distancing policy during a pandemic	Ordinal	Unlikely; less likely; neutral; more likely; absolutely
Demographics		
Variables	***Data type***	***Value list***
Household size	Numerical	Positive integer
# Workers	Numerical	Non-negative integer
# Household vehicles	Numerical	Non-negative integer
Age	Categorical	18 to 24 years; 25 to 34 years; 35 to 44 years; 45 to 54 years; 55 to 64 years; 65 or more years
Gender	Categorical	Male; female; other; prefer not to say
Ethnicity	Categorical	White/European White/Caucasian, Hispanic/Latino/Mexican American, Asian/Asian American; Black/African American, American Indian/Native American; Mixed/Multiracial; other; prefer not to say
# Children	Numerical	Non-negative integer
Household's total annual income	Categorical	Less than \$10,000; \$10,000 to \$19,999; \$20,000 to \$29,999; \$30,000 to \$39,999; \$40,000 to \$49,999; \$50,000 to \$59,999; \$60,000 to \$74,999; \$75,000 to \$99,999; \$100,000 to \$124,999; \$125,000 to \$149,999; \$150,000 to \$199,999; \$200,000 or more
Education	Categorical	Didn't complete high school; completed high school, some college but no degree; associate or technical degree; bachelor's degree; master's degree; Ph.D.
Employment status	Categorical	Employed, 40 or more hours per week; employed less than 40 hours per week; student, working part time; student, not working; not employed, looking for work; not employed, not looking for work; retired; disabled, not able to work
Marital status	Categorical	Single; married; divorced; widowed

9-digit home zip code	Zipcode	
US home state	States	
Closest airport	Text	

4.2.2. RSG rJourney

Produced by RSG for the FHWA (FHWA, 2015), the rJourney long-distance nationwide travel demand model incorporates long-distance travel surveys (i.e., ATS, 2001 NHTS, and long-distances surveys in California, Colorado, and Ohio), socioeconomic data, and network data (FHWA, 2015). The 1.17 billion rJourney tours are generated from a synthesized household population of 31.5 million, representing all long-distance travel in the year 2010. All simulated tours consist of one outbound and one return trip over the same path, none shorter than 100 miles. As expected, car usage largely dominates shorter trips (less than or equal to 500 miles, or 805 km), while air travel dominates for longer ranges. Bus and rail consistently account for a small portion of all trips. The average party size in a tour is 2.15 people in this dataset.

Because rJourney was targeted for conventional mode choices in 2010, AVs were never considered. Although an experimental effort was conducted to add AVs as a mode choice in the downstream portion of the overall model structure (Perrine et al., 2020), this project's researchers surmise that the presence of AVs would impact auto ownership, trip generation, and "trip nights staying" outcomes enough that those respective modeling processes should be performed anew for this project to ensure better accuracy, rather than being "borrowed" from rJourney.

Despite that, rJourney still offers several features that can assist this project's efforts. In rJourney, long-distance travel is modeled among almost all pairwise combinations of 4,486 NUMA zones that are derived from both Census Bureau Public Use Microdata Areas (PUMAs) and county boundaries. The scale found with these NUMAs is not too large to obscure valuable finer details for long-distance travel demand modeling, nor too small to be infeasible for nationwide modeling efforts. As a result, NUMAs shall be leveraged for this project's modeling efforts.

rJourney provides skims matrices among all NUMAs for personally owned auto, bus, rail, and air travel with several items about travel time, distance, cost, air entry and egress overhead, and more, tuned to 2010. A subset of relevant summary statistics from this skim file are shown in Table 9.

Table 9. Summary statistics for the rJourney skim file for year 2010 LD trips within US

Variable	Mean	Std	Min	Max
Air File, *N* = 18,424,925 journeys				
Time (minutes)	218.7	97.94	25	812
Transfer time (minutes)	82.37	50.19	0	200
Frequency of direct journeys per week	10.58	24.07	0	339
Frequency of journeys with 1 stop per week	145.4	258.5	0	2,286
Frequency of journeys with 2 stops per week	348.8	932.7	0	10,968
Percent journeys within 30 minutes of scheduled arrival	88.79	4.00	0	100
Economy fare ($)	519.1	327.7	0	50,776
Business fare ($)	1,200	955.6	0	152,328
Access distance to airport from NUMA center, max. 100 (miles)	38.15	25.99	0	101
Egress distance from airport to NUMA center, max. 100 (miles)	38.22	26.34	0	102
Rail File, *N* = 8,010,759 journeys				
Time (minutes)	2,167	1,270	4	6,270
Transfers * 100	134.6	111.1	0	800
Frequency of departures per week	7.77	10.41	3	93
Economy fare ($)	131.8	39.51	9	181
Business fare ($)	340.6	132.4	18	605
Access distance to rail station from NUMA center, max. 50 (miles)	22.82	14.65	0	50
Egress distance from rail station to NUMA center, max. 50 (miles)	22.16	15.14	0	50
Road File, *N* = 19,727,179 journeys				
Time (minutes)	1,162	668.1	1	3,613
Distance (miles)	1,185	706.5	1	3,582
Toll (cents)	67.15	137.9	0	1,344
Bus time (minutes)	1,313	1,250	0	5,617
Bus fare (minutes)	94.71	85.72	0	383

4.2.3. National Household Travel Survey 2016/17

The 2016/17 NHTS surveyed travel behavior (e.g., travel mode and trip purpose) by residents in all 50 states and the District of Columbia. The travel data were collected with travel dates starting on April 19, 2016, and ending on April 25, 2017 (FHWA, 2017). Respondents in a stratified random sample of US households were designated a 24-hour travel day which started at 4:00 a.m. (local time) on the assigned travel day and ended at 3:59 a.m. of the following day. The daily travel data for each household includes all trips made by all household members aged five or older in that 24-hour period. Weights were utilized to produce well-balanced population-level estimates, including weights of household, trip, person, and vehicle (FHWA, 2017).

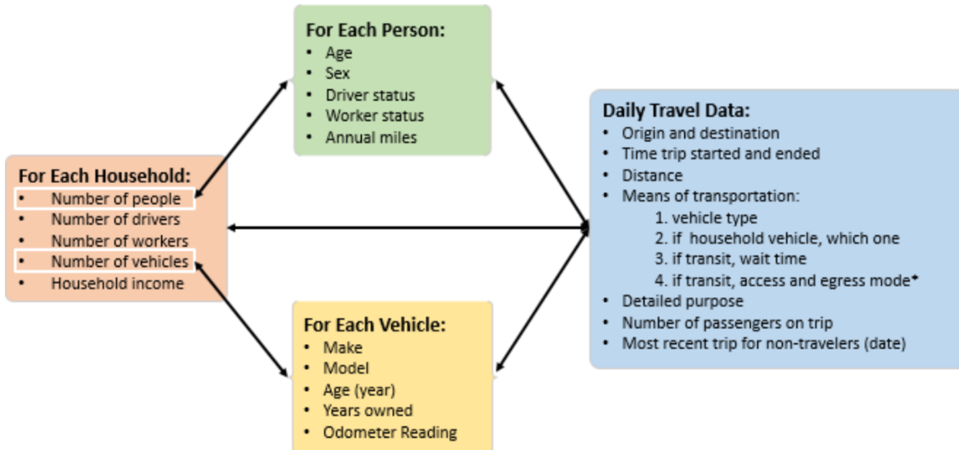


Figure 34. Core information in four components of NHTS data (FHWA, 2017)

Figure 34 shows how the household, person, vehicle, and trip datasets from NHTS are interrelated. The trip dataset offers information on all trips made by all household members, including the details of the trip (e.g., origin and destination, distance, etc.) associated with the vehicle, person, and household linked to more details in separate datasets.

The key variables at the household level and trip level that will be valuable to assemble the dataset that will be used for parameter estimations of the long-distance travel demand model can be found in Table 10. These variables will be further weighted, combined, and normalized with the survey dataset developed by the team.

Table 10. Key variables in NHTS 2016/17 data

Household	Trip
Household size # Workers in household # Persons aged between 0 and 4 in household # Drivers in household 2010 census division classification for the respondent's home address Count of household trips on travel day Urban / rural indicator Household income	• Respondent's age • Respondent's gender • Respondent drove on trip indicator • Respondent's educational attainment • Trip start time & trip end time • Trip origin and destination • Trip duration in minutes • Trip origin purpose • Trip destination purpose • Trip purpose summary • Generalized purpose of trip, home-based and non-home-based

Household	Trip
Household state # Household vehicles Homeownership # Adult household members at least 18 years old Frequency of paratransit use for travel Frequency of train use for travel Travel is a financial burden Date of travel day & travel day of week Land use variables in the census tract of the household’s home location: workers per square mile; percent of renter-occupied housing; population density; housing units per square mile	• Weekend/weekday trip indicator • Price of gasoline in cents on respondent’s travel day • Number of non-HH/HH members on trip • Primary activity in previous week • Household vehicle used on trip indicator • Trip distance in miles, adjusted for comparability to past surveys • Trip distance in miles, derived from route geometry returned by Google Maps API or from reported loop-trip distance • Trip mode, derived • Vehicle type • Public transportation used on trip • Count of transfers • Trip time to transit station in minutes • Bus/rail/other modes used to get to public transit indicator • Transit wait time in minutes

A data summary of the NHTS has been provided in Chapter 1. A few key findings are as follows:

- Automobile travel (i.e., single-occupancy vehicle [SOV] and high-occupancy vehicle [HOV]) dominates short-distance travel less than 400 miles (84% of trips for the US vs. 89% for Texas), while air becomes the dominant mode (72% for the US vs. 64% for Texas) for trips longer than 400 miles.

- The total PMT for one-way trips is 3,951.2 billion (352.1 billion in Texas) for US all-distance trips, among which 1,707.7 billion, about 43% (158.0 billion, about 45%, in Texas), are one-way long-distance trips over 50 miles and 1,356.6 billion, about 34% (130.3 billion, about 37%, in Texas), are one-way long-distance trips over 100 miles. Texas has a higher share of one-way long-distance trips over 50 miles and over 100 miles, respectively, compared with the US average.

- Friday has the biggest share of PMT for trips in the US across all distances and all seven days of the week, while Saturday has the top share of PMT in the Texas region. However, for long-distance trips, Sunday has the biggest share of PMT for both the US and Texas.

- For all trips in both the US and Texas, automobile (SOV + HOV) trips have the majority share of total PMT, with the percentage in Texas being slightly higher. As the trip distance increases, the percentage of airplane PMT increases in both the US and Texas, and the SOV PMT percentage decreases significantly.
- According to NHTS 2016/17 household travel survey data, we find that Americans are driving 2.106 trillion miles, or 9,441 miles per licensed driver, per year. The average American (age five and older) travels 13,179 miles a year, while those age 18 and up travel 16,031 miles a year. The average American flies about 2,215 miles per year (with the average adult American flying 2,694 miles per year, or 16.8% of his/her total PMT). About 44.0% of Americans' PMT each year occurs on trips over 50 miles each way, and 34.7% of PMT occur on trips over 100 miles each way.
- Long-distance trips tend to have more people in the vehicle compared to short-distance trips, shown by the average vehicle party size (for trips over 50 miles: 1.38 in the US vs. 1.66 in Texas, and for trips across all distances: 1.21 in the US vs. 1.42 in Texas). Looking at the party size variations by month for all trips, the party size for the US and Texas presents a similar trend, while Texas has a larger party size in June, July, and January than that of the US for long-distance trips.

4.2.4. Passenger Airline Ticket Sales (DB1B)

The Office of Airline Information of the BTS collects airline ticket data from reporting carriers and gathers the data in a dataset called the Origin and Destination Survey Databank or DB1B. DB1B is a 10% random sample of airline passenger tickets. This dataset, which is reported quarterly (i.e., four times per year), contains three tables: coupon, market, and ticket. The DB1B ticket table, which is of interest in this study, includes trip origin and destination data, year and quarter indicators, number of passengers, number of legs, and distance and fare information for each itinerary. This dataset has been published since 1993, providing 28 years of data. The team downloaded the data for the period between January 2019 and March 2021. This includes the first year of the COVID-19 pandemic and the year prior, so the team can understand the effect of COVID-19 on air travel frequency and other factors. The total number of itineraries available in this dataset over the course of 27 months (from January 2019 to March 2021) is almost 27 million, for around 58 million passengers. A subset of relevant summary statistics from this dataset is shown in Table 11. The reduction in air travel due to COVID-19 can be easily observed in the data reported in this table.

This dataset will be used to model the utility of air travel for long-distance trips in the mode choice model.

Table 11. Summary statistics for the DB1B air ticket data

Variable	Mean	Median	Std
2019 Quarter 1, *N* = 3,981,589			
Distance (mi)	2,164	1,850	1,485
Itinerary fare ($)	403	344	500
# of legs in the itinerary	2.47	2.00	1.11
Price per miles flown ($/mi)	0.26	0.18	0.59
2019 Quarter 2, *N* = 4,595,761			
Distance (mi)	2,166	1,834	1,499
Itinerary fare ($)	416	362	1435
# of legs in the itinerary	2.47	2.00	1.11
Price per miles flown ($/mi)	0.27	0.19	1.10
2019 Quarter 3, *N* = 4,390,179			
Distance (mi)	2,182	1,846	1,505
Itinerary fare ($)	404	354	453
# of legs in the itinerary	2.49	2.00	1.12
Price per miles flown ($/mi)	0.26	0.18	0.58
2019 Quarter 4, *N* = 4,480,274			
Distance (mi)	2,155	1,832	1,475
Itinerary fare ($)	417	359	1325
# of legs in the itinerary	2.48	2.00	1.00
Price per miles flown ($/mi)	0.27	0.19	3.52
2020 Quarter 1, *N* = 3,281,382			
Distance (mi)	2,157	1,840	1,487
Itinerary fare ($)	391	333	374
# of legs in the itinerary	2.47	2.00	1.11
Price per miles flown ($/mi)	0.25	0.18	0.56
2020 Quarter 2, *N* = 762,315			
Distance (mi)	1,992	1,740	1,237
Itinerary fare ($)	279	233	1,755
# of legs in the itinerary	2.39	2.00	1.09
Price per miles flown ($/mi)	0.18	0.13	2.16
2020 Quarter 3, *N* = 1,675,969			
Distance (mi)	2,102	1,837	1,302
Itinerary fare ($)	270	227	460
# of legs in the itinerary	2.56	2.00	1.13
Price per miles flown ($/mi)	0.17	0.12	2.40
2020 Quarter 4, *N* = 1,926,506			
Distance (mi)	2,157	1,874	1,385
Itinerary fare ($)	294	251	361
# of legs in the itinerary	2.54	5.00	1.12
Price per miles flown ($/mi)	0.18	0.13	0.56
2021 Quarter 1, *N* = 1,965,254			
Distance (mi)	2,137	1,850	1,422
Itinerary fare ($)	298	253	249
# of legs in the itinerary	2.46	2.00	1.12
Price per miles flown ($/mi)	0.19	0.13	0.19

4.2.5. TxDOT's Statewide Analysis Model (SAM)

The Texas statewide analysis model was initialized and built in 2001, and four versions have been developed so far. The SAM-V4 was designed to model 2015 as the base year and forecast travel demand for years 2025, 2035, and 2045.

Both SAM's passenger and freight models follow the four-step model structure, which is shown in Figure 35. After trip generation and distribution, the mode and destination choice results of passenger travel are transformed into trip tables or OD matrices for the final traffic assignment. The freight trip tables (in tons by commodity) are converted to trucks and rail cars, based on SAM weights. Feedback loops are performed to provide consistent results between travel time and cost skims and network assignment flows, feeding congested travel times back for subsequent iterations.

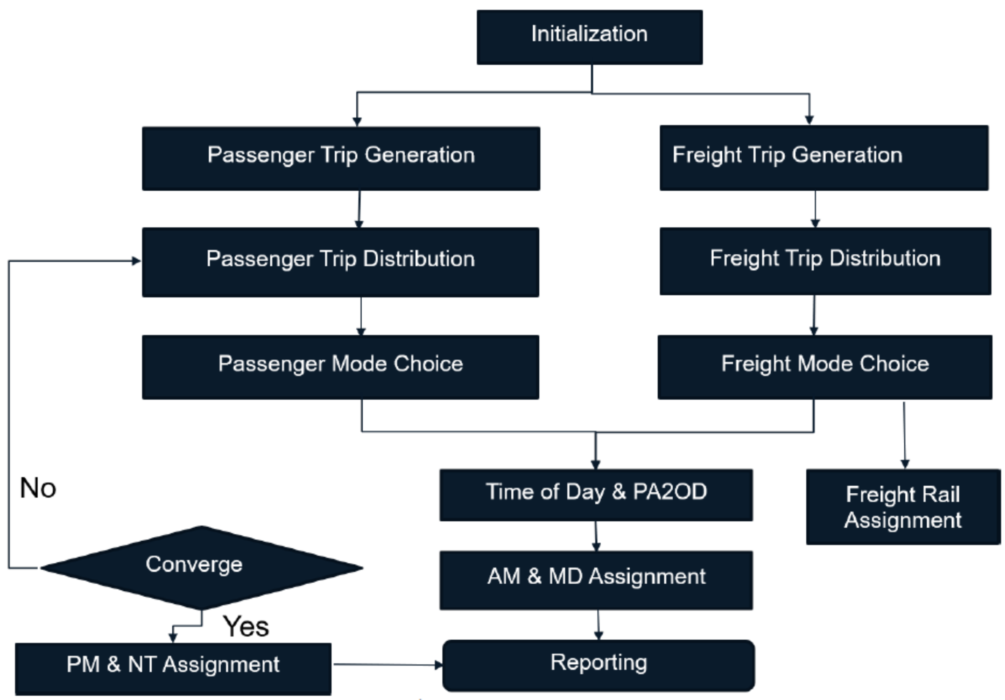
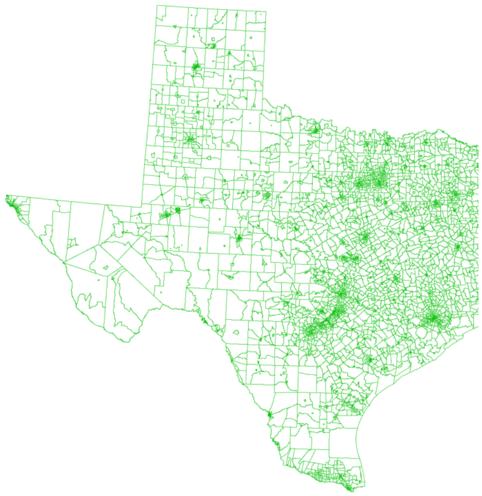
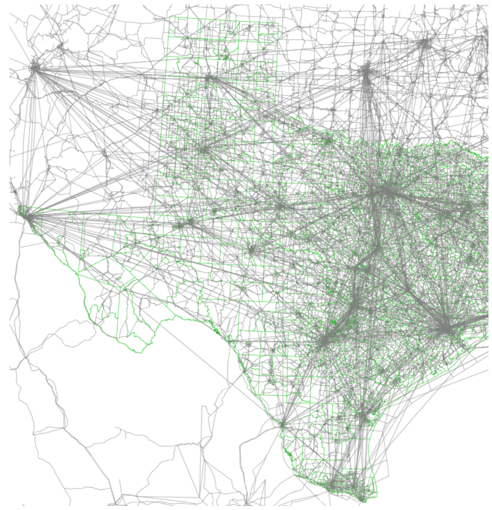


Figure 35. SAM four-step model of passenger and freight

The SAM network covers all of North America, with greater detail in and near Texas. Figure 36 shows the state's highway, railway, and airline networks, which contain 200,445 links and 168,507 nodes.



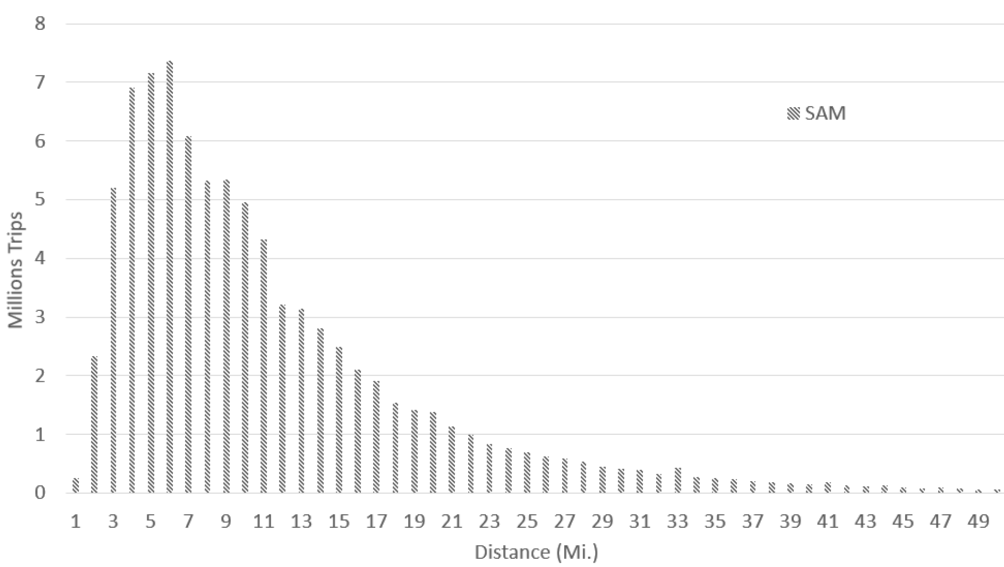
a. SAM Traffic Analysis Zones



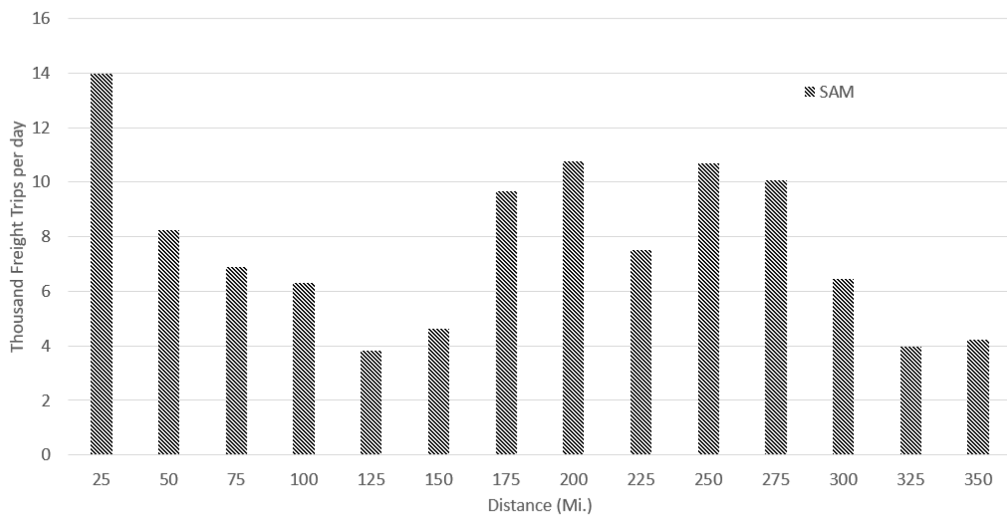
b. SAM Networks

Figure 36. SAM geographic data

The team summarized the data from the SAM-V3 output, which had already been obtained. SAM-V3 shows that the average trip length of passenger travel across Texas is about 12.0 miles. Around 60% of trips are under 10 miles, but these trip only account for 30% of total vehicle miles. Figure 37 charts the passenger and freight trip distribution across Texas. The trip distribution pattern of freight movement shows a close connection between Texas's major cities, since there is a peak in freight trips from 175 miles to 275 miles. Incidentally, 175 miles is about the distance from Austin to Houston, while 275 is about the distance from Houston to Dallas.



a. Passenger trip distribution



b. Freight trip distribution

Figure 37. Passenger and freight trip distribution based on SAM-V3 data

4.3. Automated Vehicle Ownership and Usage

This section explores research efforts to predict the future of AV ownership and usage—aspects that will heavily influence this project's models of AVs as a mode choice. We will identify the works that researchers anticipate leveraging as immediate inputs to AV ownership and destination and mode choice. This is supported by an overview of other surveys and models that will serve as guidance for considering the accuracy and validity of this project's outcomes.

There are several studies in the literature to model AV and conventional vehicle ownership with the advent of new technologies, including AVs. For example, Lavieri et al. (2017) modeled individual preferences for ownership and sharing of AVs using education, age, income, employment status, household composition, and experience of ride sharing and car crashing. The results showed that 15.4% of respondents were interested in AV ownership and sharing. A study conducted by Zhang et al. (2018) modeled vehicle ownership reduction and unoccupied VMT based on the advent of AVs using travel surveys and synthesized trip profiles from the Atlanta metropolitan area. The study developed a greedy algorithm, which showed that more than 18% of households can reduce vehicle ownership while maintaining their current travel patterns; that is equivalent to a 9.5% reduction in private vehicles in the study region. The study found that if travelers' schedules are relaxed by 15 minutes, up to 24.1% of households are likely to eliminate at least one of their current private vehicles. The study also applied a logistic regression model, which showed that people who live in suburban areas and have high income levels are more likely to reduce vehicle ownership as AVs become available. A more recent study by Lee et al. (2019) assessed factors that influence people's intentions to use AVs. The demographic factors used in the study include age, gender, marital status, monthly income, and current vehicle ownership. A structural equation model was estimated in this study to examine relationships between perceived usefulness, intention to use, perceived risks, and ease of use.

In another study by Kim et al. (2020), 3,106 respondents in Georgia were asked about short-term impacts of AVs on mode choice, medium-term impacts on activity patterns, and long-term impacts on behavior change. They used a cross-nested logit model for modeling the choice of home location and vehicle ownership. The results indicated that those with more vehicles than the number of household members of driving age are more likely to let go of vehicles. The younger pro-suburban individuals were more likely to move farther away from their current location after adopting an AV.

Furthermore, Bansal and Kockelman (2016) put forward a simulation-based fleet evolution network to forecast long-term adoption of connected and autonomous technologies by Americans. Respondents were asked about their preferences towards their household's annual vehicle transactions and CAV technologies. Similar activity pattern changes were studied by Le-Klähn et al. (2019), using SimMobility (an integrated simulation model) to assess how AV services change respondents' preferences and attitudes. This study also tried to determine changes in house location and vehicle ownership.

In addition to AV ownership models, a study conducted by Maeng and Cho (2022) aimed to assess consumer choice and usage patterns in anticipation of SAVs' upcoming launch into the market. The study used stated preference data from South Korea to reflect the mode choices at a disaggregated level. Usage of SAVs also leads to predicted future reduced demand for parking. Silva et al. (2021) used a case study conducted in Budapest, Hungary, to analyze the impacts of AV acceptability and usage of shared vehicles on parking demand. They used the responses collected in the survey to build present, transitional, and future scenarios to quantify the gradual changes in AV ownership. Similarly, Jiang et al. (2018) used a mixed logic model to analyze AV ownership behavior in Japan, while Saeed et al (2020) used a random parameter logit model on 2,097 survey responses collected across the US. The parameters used in the study included age, gender, education level, income level, and travel behavior.

Apart from demographic parameters and daily travel activities, it has been found that adoption of AVs and SAVs is quite susceptible to psychological parameters, such as enjoying driving and trust issues with upcoming technology, both of which make a personal car more attractive for trips. Such parameters are considered by Asgari and Jin (2019) while analyzing AV adoption rates and willingness to pay (WTP) for AVs on four levels of automation. They found that AVs were quite readily accepted by technologically savvy people. However, it was hard to convince the people who enjoyed driving to adopt the new technology. Also, people are more likely to accept increased costs and the new automation features when they believe that they will improve quality of life in terms of time and cost savings, stress reduction, and more convenience.

Wang et al. (2021) also conducted a stated preference experiment in the greater Toronto area to understand Torontonians' WTP for different vehicle automation levels and their preferences towards shared vehicle ownership. A survey was conducted that received 190 usable responses. The authors used the dataset from this survey to estimate WTP for different automation levels and users' vehicle ownership choices. The results showed a strong inertial tendency among respondents to stay with conventional private vehicles. In addition, the private car buyers showed a higher tendency towards Level 4 automation, while car-sharing service users preferred Level 5. The parameters of this study's vehicle ownership model can be used for this project's mode choice modeling.

The research team conducted a survey in 2017 to measure the impacts of technology availability and costs on vehicle ownership in the US. In this study, Quarles et al. (2018) surveyed adult Americans about their and their households' willingness to buy or use electric, self-driving, and shared vehicle types. After cleaning the dataset and removing responses that were nonsensical or contradictory, the final sample contained the responses of 1,426 Americans. In this survey, respondents were asked about the number of vehicles owned by their household and their willingness to purchase a new vehicle type or release an existing vehicle next year. If the respondent was willing to do so, he/she was asked about the likeliness of this plan and the new vehicle type he/she plans to acquire.

To track vehicle ownership in later years of the survey, Quarles et al. (2021) used a household-level micro-simulation to model vehicle ownership for 2017–2050. The micro-simulation is calibrated with the survey data and assumes that the trend of acquiring a new vehicle will remain constant in the future. For this purpose, 2017 survey data is first used to model the households' WTP for AVs, where human-driven vehicle capability is maintained in all fully autonomous vehicles. The coefficients of this model are presented in Table 12.

Table 12. Regression coefficients for annual application of household WTP with human-driven option (Quarles et al., 2020)

Variable	Parameter	t-statistic
Intercept	5124.40	8.94
Age	-53.87	-7.22
HH children	210.16	1.73
HH income	7.34	3.20
Student or not	-1127.40	-1.87
Unemployed or not	-1127.3	-4.01
Married or not	544.63	2.35
HH vehicles	-271.91	-2.07
Vehicle purchase year probability	31.06	10.74
Grocery distance	34.52	1.83
Public transit distance	-14.95	-1.41
No disability	-837.69	-2.17
$N = 1426$, $R^2 = 0.2025$		

The study then assumes that the WTP for AV technology will increase 5% each year. Using the calibrated model and this assumption, the authors simulated a base scenario considering a constant 5% annual decline in an AV's purchase price premium (maintaining human-driven capability) over the simulation period (i.e., from 2017 to 2050). The AV ownership for each household is defined by comparing WTP with the AV technology price in the year of interest. Table 13 shows that total vehicle ownership increases during the simulation period under this scenario. AV ownership does not increase until 2040, when it then begins increasing dramatically, reaching almost 40% of private vehicle ownership by year 2050. The project team used the survey data and the vehicle ownership model in the study by Quarles et al. (2021) in this project.

Table 13. US privately owned fleet composition with 5% AV premium decline and human-driven capability (Quarles et al., 2020)

Year	Vehicle/HH (Std)	Vehicles/HH of AV or Conventional Vehicle	
		% AV	% Conventional
2020	1.95 (0.030)	0	100
2025	1.78 (0.038)	0	100
2030	1.62 (0.041)	0	100
2035	1.54 (0.039)	0	100
2040	1.48 (0.041)	0.97	99.03
2045	1.46 (0.044)	13.66	86.34
2050	1.45 (0.037)	36.41	63.59

4.4. Freight Data: Freight Analysis Framework

The Freight Analysis Framework 4 (FAF4) data integrates different data sources like the 2012 Commodity Flow Survey (CFS) and international trade data from the Census Bureau to provide 30-year estimates of tonnage, value, ton-miles, and value-miles by FAF regions of origin and destination and by transportation modes for different industry sectors like agriculture, extraction, utilities, construction, service, and others.

Two types of freight flow are defined in FAF4 data: domestic flow and international flow. The domestic flows are the freight trips from domestic origins and destinations, while the international flows are either import flow from foreign origins arriving at domestic destinations or export flow from domestic origins arriving at foreign destinations. A flow of freight movement can be a combination of international flow and domestic flow. The origins and destinations in the FAF are at the state level, or at the FAF region level, which has a better resolution of 132 domestic regions and 8 international regions. In terms of the commodity, FAF4 data uses the two-digit level of the Standard Classification of Transported Goods (SCTG) to define the commodity categories.

Table 14 shows the definitions of transportation modes that are used to categorize the freight flow, which is documented in the FAF4 user guide (BTS, 2017).

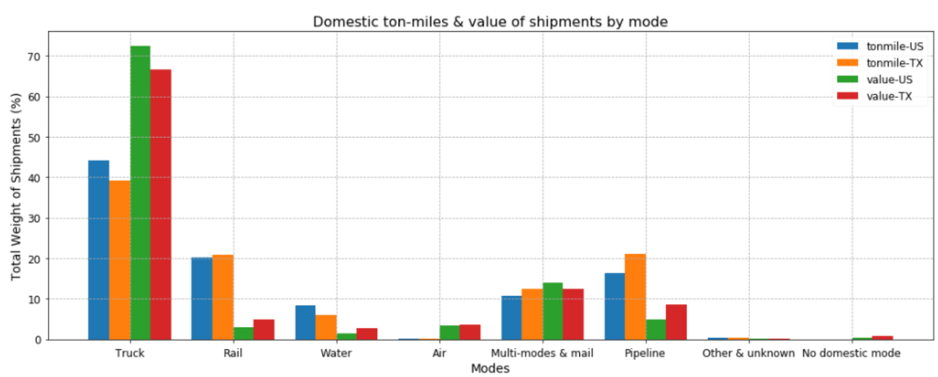
Table 14. FAF modes (BTS, 2017)

Code	Mode	Description
1	Truck	Includes private and for-hire truck. Does not include truck that is part of **Multiple Modes and Mail** or truck moves in conjunction with domestic air cargo.
2	Rail	Includes any common carrier or private railroad. Does not include rail that is part of **Multiple Modes and Mail**.
3	Water	Includes shallow draft, deep draft, Great Lakes and intra-port shipments. Does not include water that is part of **Multiple Modes and Mail**.
4	Air (includes truck-air)	Includes shipments move by air or a combination of truck and air in commercial or private aircraft. Includes air freight and air express. In the case of imports and exports by air, domestic moves by ground to and from the port of entry or exit are categorized with **Truck**.
5	Multiple Modes and Mail	Includes shipments by multiple modes and by parcel delivery services, US Postal Service, or couriers (capped at 150 pounds). This category is not limited to containerized or trailer-on-flatcar shipments.
6	Pipeline	Includes crude petroleum, natural gas, and product pipelines. Note: Does include flows from offshore wells to land which are counted as **Water** moves by the US Army Corps of Engineers. Does not include pipeline that is part of **Multiple Modes and Mail**.
7	Other and Unknown	Includes movements not elsewhere classified such as flyaway aircraft,

Code	Mode	Description
		and shipments for which the mode cannot be determined.
8	No Domestic Mode	Includes shipments that have an international mode, but no domestic mode and is limited to import shipments of crude petroleum transferred directly from inbound ships to a US refinery at the zone of entry. This is done to ensure a proper accounting of import flows, while avoiding assigning flows to the domestic transportation network that do not use.

Based on the definition above, Figure 38 shows the mode shares of domestic, import, and export freight movement in ton-mile and value for the whole US as well as Texas. Excluding truck trips involved in the multi-modes, truck dominates domestic freight movements in both tonnage and value transported, accounting for over 40% in ton-miles and almost 70% in value. Following truck, rail is the second mode choice in the US, transporting about 20% of miles. Compared to the US overall, Texas has a higher usage of pipeline, while truck's mode share is lower.

International flow shows a different pattern from domestic flow. Water is the dominating mode choice of export and import flow in the US, for both ton-miles and value transported. In terms of export flows, air freight trips rank second in total value transported in the US, mostly due to electronic goods. However, in Texas, truck ranks first in total value transported, followed by water. Compared to export flow, import flow has higher water in ton-miles transported and lower air mode shares in value.



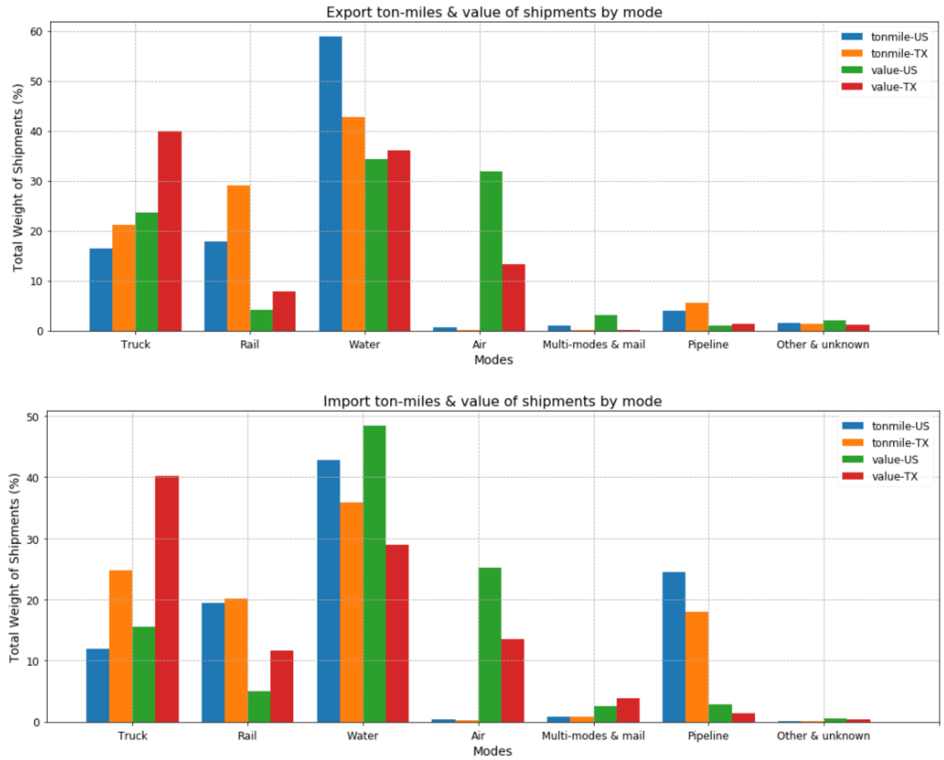


Figure 38. Freight mode shares by weight and value (Source data: FAF5)

The nation's and state's top 10 commodities transported, in tonnage and value, are also presented in Appendix E. Electronics are always the top commodity transported in value, for both Texas and nation, for domestic or international trips. In general, the tonnage freight movement pattern is different from the freight movement pattern for value transported. High-ranking goods in transported tonnage include more essential and large-size raw materials, like coal and gravel, while goods that rank high in transported value are mostly delicate and high-value electronics or machinery. As expected, Texas presents a similar pattern to the US, but fuel oil and petroleum rank higher in tonnage compared to the US.

Passenger Travel Models and Application

This chapter continues efforts that were undertaken in the previous chapter to create a nationwide travel model to predict mode and destination choices for long-distance passenger travel in a future where AVs are readily available in the marketplace. This chapter documents efforts to synthesize the population and use this synthesis to generate disaggregate long-distance trip models for scenarios before and after AV market penetration. Figure 39 provides the model framework, delivering a set of individual long-distance trip estimates to reflect all long-distance travel by future US populations. Pre-trip models include the decision to participate in long-distance travel and its purpose, tour frequency over the course of a year, trip season, and party size for each tour. The outputs of these models are used as inputs to estimate the destination and mode choice models, with mode choice conditioned on household vehicle ownership decisions. US household and person synthetic data at census tract level (10% of the entire population) as well as 2019 PUMS data were used to run these models and generate disaggregate long-distance trips for the sampled population.

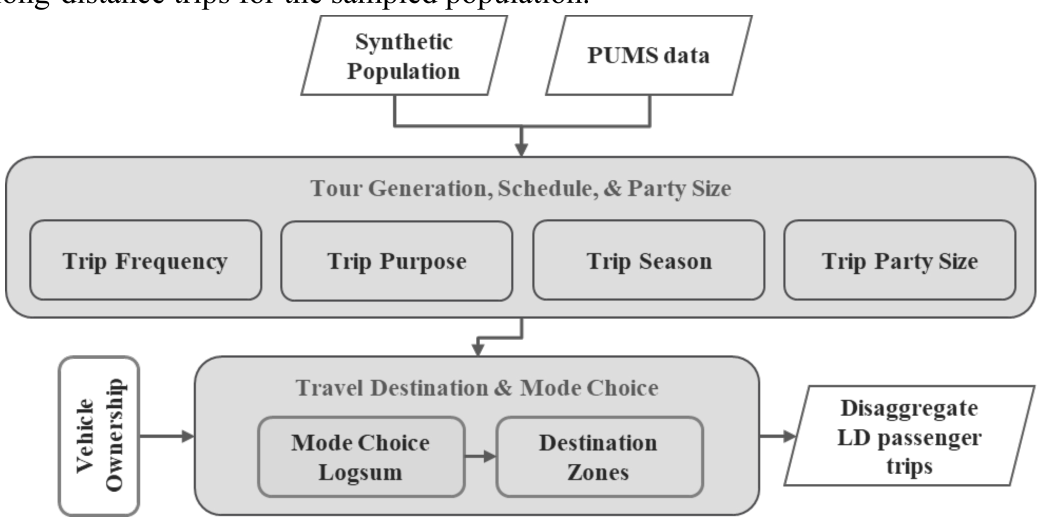


Figure 39. Long-distance travel model framework for domestic US passenger travel

Figure 40 illustrates the steps required to use synthetic populations and estimated models to generate disaggregated long-distance trips for the sample US population. The model application uses different datasets in addition to the synthetic population, including the 2016/17 NHTS, PUMS, EPA Smart Location, and FHWA rJourney datasets. Other data sources map different geographic codes among these datasets. The model application starts by applying the vehicle ownership models prior to and after AV market penetration to different households. The number of human-driven vehicles (HVs) and AVs owned by each household is then used as an input to person-level models to find the specifications of long-distance trips. Note that some variables in the estimated models, such as driver's license possession and distance from home to transit and grocery stores, were not available in the synthetic population and PUMS dataset. Thus, the research team estimated models for these variables using the 2016/17 NHTS data.

Person-level models start by finding long-distance trip frequency using the zero-inflated negative binomial model. Trip season and trip purpose are estimated for each trip using multinomial logit models. The results of all these models are then used to estimate the destination choice of each traveler. The destination choice model then requires mode choice logsums for all OD pairs, which were found using FHWA rJourney skims and mode choice models. Next, party sizes for each trip are estimated, followed by the mode choice. The details of each model as well as the outputs are explained in the following sections.

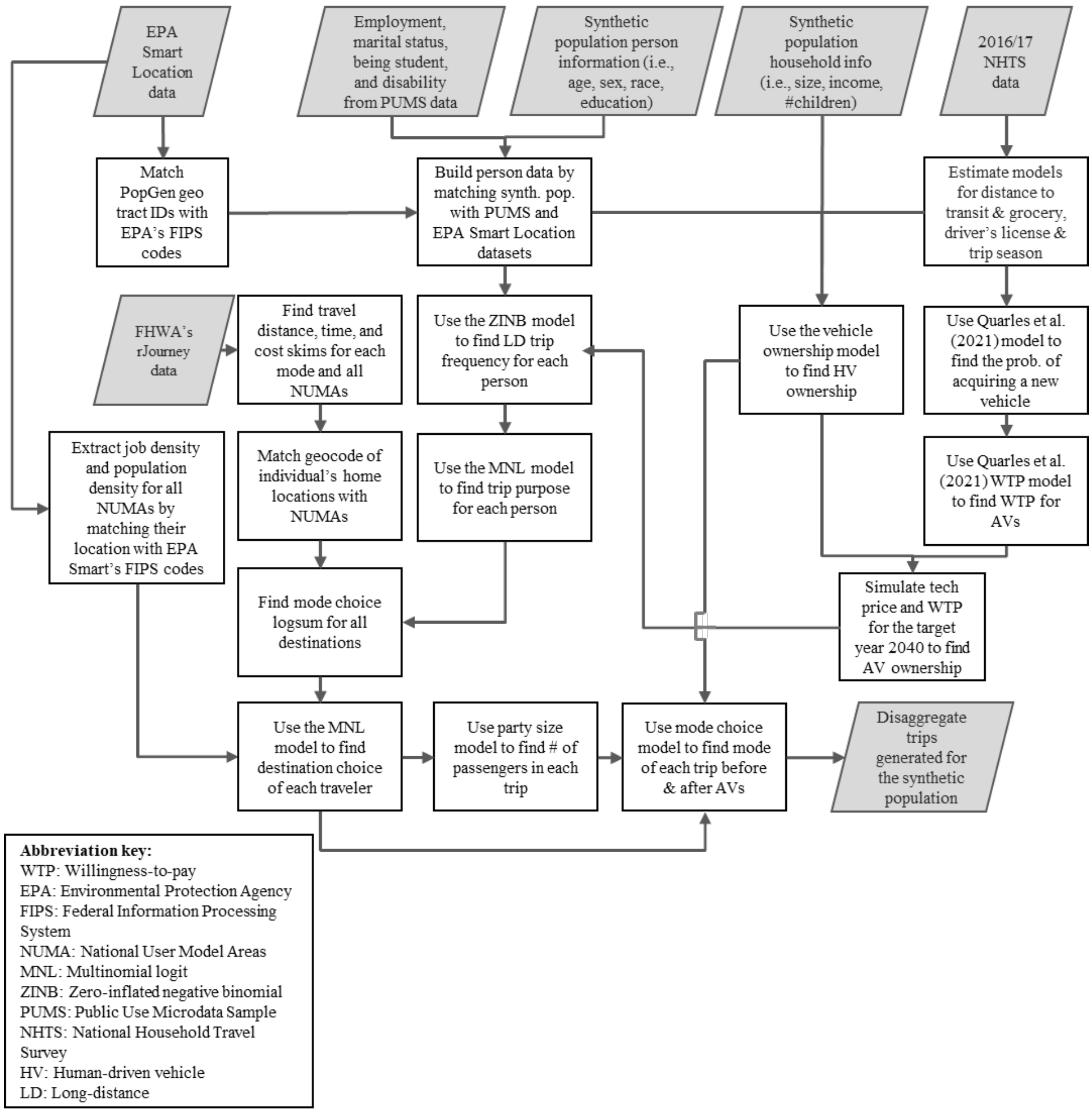


Figure 40. Detailed steps of the model application for disaggregated long-distance trip generation

5.1. Synthetic Population

To simulate US travel, the research team created a 10% synthetic population at the level of the nation's 73,056 census tracts. The team had to run this massive data synthesis on the Texas Advanced Computing Center's (TACC) super-computers over several days. The synthesized population is based on marginals from 2019 five-year ACS data using PopGen 2.0 software, developed by Pendyala et al. (2011) and Ye et al. (2009).

The household and person data were synthesized across 2,351 PUMAs to mimic the population distributed across the US (50 states and the District of Columbia), consistent with census datasets and geographic-correspondence files. Margins of households' income and size and individuals' gender, age, race, and education were scraped from the Census Bureau and processed as the input for PopGen. The output of the synthesis is a 10% sample of Americans' household and person data that are well-matched to each other and to the control margins. The personal and household margins from ACS at census travel level were used; detailed categories and the share of each category in the synthesized population are summarized in Table 15. The synthetic households and persons are sampled from PUMS. The PUMS 2019 data and dictionary were obtained from the US Census Bureau FTP site (US Census Bureau, 2022). The household IDs and person IDs in the synthetic population correspond to the IDs in PUMS files. Thus, these data were used to extract variables that were not reported in the synthetic population files. Overall, the synthetic population provided the specification of 12,082,535 households and 28,097,623 individuals across 73,056 census tracts.

Table 15. Summary statistics of the synthesized population (2019, 28.1M persons and 12.1M households)

Variable	Category	2019 Synthetic Population	2016/17 NHTS
PERSON			
Sex	Male	47.43%	49.07%
	Female	52.56%	50.93%
Race	White	73.54%	72.49%
	Black or African American	12.23%	12.71%
	Asian	5.38%	5.33%
	American Indian or Alaska Native	0.76%	0.86%
	Native Hawaiian/Pacific Islander	0.16%	0.28%
	Multiple responses selected	3.19%	3.96%
	Some other race	4.73%	4.37%
Education	High school graduate or GED	52.71%	33.51%
	Some college or associate degree	23.91%	28.56%
	Bachelor's degree	14.72%	21.02%
	Graduate or professional degree	8.66%	16.90%
Age	Younger than 10 years old	11.99%	8.37%
	11–17 years old	10.10%	9.69%
	18–24 years old	8.49%	10.37%
	25–34 years old	13.79%	14.07%
	35–44 years old	12.80%	14.03%
	45–54 years old	13.34%	13.43%
	55–64 years old	13.29%	14.45%
	65–74 years old	9.44%	10.05%
	75 years or older	6.75%	5.12%
HOUSEHOLD			
Household Size	1-person HH	27.86%	27.88%
	2 persons in HH	33.93%	33.88%
	3 persons in HH	15.59%	15.67%

Variable	Category	2019 Synthetic Population	2016/17 NHTS
	4 persons in HH	12.90%	14.33%
	5 persons in HH	5.97%	5.42%
	6 persons in HH	2.30%	1.93%
	7 or more persons in HH	1.44%	0.89%
Annual Household Income	Less than $10,000	5.87%	7.51%
	$10,000–$14,999	4.33%	6.02%
	$15,000–$24,999	8.95%	9.78%
	$25,000–$34,999	8.97%	10.01%
	$35,000–$49,999	12.30%	12.37%
	$50,000–$74,999	17.26%	16.54%
	$75,000–$99,999	12.77%	12.30%
	$100,000–$124,999	9.17%	9.38%
	$125,000–$149,999	6.07%	5.35%
	$150,000–$199,999	6.84%	5.22%
	$200,000 or more	7.49%	5.50%
#Children	0 children	70.60%	69.92%
	1 child	9.69%	12.13%
	2 children	11.96%	12.29%
	3 children	5.18%	3.94%
	4 children	1.84%	1.22%
	5 or more children	0.72%	4.93%

5.2. Trip Frequency

The long-distance trip frequency model was developed using the zero-inflated negative binomial distribution based on population-weighted 2016/17 NHTS trip data for domestic travels. A zero-inflated model assumes that a “zero outcome” comes from two different processes. In this case, it starts with a decision process about whether or not a person is making a long-distance trip (>75 miles). If the person decides not to make a long-distance trip, the only possible outcome is zero. If the person is making a long-distance trip, then a count model is utilized to predict how many trips will be made.

The two parts of the zero-inflated negative binomial model are, first, a logit model to decide whether a person will make a long-distance trip or not, and second, a negative binomial count model to find the number of trips. Table 16 presents the summary statistics of all tested variables in the zero-inflated negative binomial model. These variables include respondents’ and households’ information, such as age, sex, employment status, household income, and number of workers in the household.

Table 17 presents the negative binomial regression coefficients for each variable along with t-stat, p-values, and practical significance. The practical significance is defined as the change in percentage of the estimated predictor resulting from increasing the value of variables by one standard deviation (SD). This table also shows the inflation model's coefficients for predicting excess zeros (along with their t-stats and p-values). The count model's parameter estimates show that long-distance trip rates rise almost 51% following a 1 SD increase in the natural logarithm of household annual income (measured in US dollars). Shifting the population-weighted sample toward men by 1 SD increased the average long-distance trip frequency 21.6%. A 1 SD increase in households' vehicles increased long-distance trip-making rates by 66%. Applying the trip frequency model to the synthetic population indicated that 4.51% would take long-distance trip on a given day. The estimated number of long-distance trips per month is 2.003 per person, while this number was 2.03 per month per individual in the NHTS data. Following Huang et al. (2020), it is assumed in this research that trip frequency rises 15% after AVs are in market.

Table 16. Summary statistics of variables used in zero-inflated negative binomial (ZINB) model (*N* = 189,718)

Variable	Description	Mean	Median	SD	Min	Max
LD_trips	Daily long-distance trips per person >75mi	0.08	0	0.387	0	6
Age	Age of the household head?	52.15	55	17.838	16	92
Veh_per_Adults	#Vehicles per adults in household	1.13	1	0.573	0	12
Worker_ per_Adults	#Workers per adults in household	0.60	0.67	0.425	0	3
HH_Income	Household income ($1000)	83.78	62.50	55.28	10	200
#Adults	# Adults in household	2.05	2	0.790	1	10
#Workers	# Working adults in household	1.26	1	0.985	0	7
HH_Veh_Count	# Household vehicle count	2.25	2	1.023	0	12
Education_high school or higher	Household head has completed high school?	0.76	1	0.499	0	1
Male?	Household head is male?	0.48	0	0.499	0	1
Worker	Household head is worker?	0.59	1	0.490	0	1

Table 17. ZINB model for long-distance trip frequency using 2016/17 NHTS household data

Negative binomial (NB) model coefficients				
Variable	**Estimate**	**t-stat**	**P-value**	**Pract. Sign.**
(Intercept)	0.799	3.62	0.000	
Male	0.172	7.85	0.000	0.216
Age	-0.002	-3.52	0.000	-0.099
Ln (HH income) ($)	-0.079	-2.72	0.006	0.507
Education associate degree or higher	0.191	6.84	0.000	0.216
#Adults	-0.228	-14.71	0.000	-0.460
Worker	-0.080	-3.95	0.000	-0.077
HH vehicle count	0.141	12.40	0.000	0.657
ln(theta)	15.45	6.44	0.017	

Zero-inflation (ZI) model coefficients				
Variable	Estimate	t-stat	P-value	Pract. Sign.
(Intercept)	7.125	31.49	0.000	
Ln (HH income) ($)	-0.043	-4.04	0.000	0.507
HH vehicle count	-0.410	-19.80	0.000	0.657

N = 201,820, Pseudo-R^2 = 0.015

5.3. Vehicle Ownership

The number of vehicles owned by each household is a key variable in different travel models, including the trip frequency and trip purpose models. Thus, the research team estimated two household vehicle ownership models: prior to and following the mainstream introduction of AVs into the marketplace. Households' vehicle ownership prior to AVs is predicted via a Poisson regression model using population-weighted 2016/17 NHTS household data. A negative binomial model was also used to estimate households' vehicle ownership, but it was found that the Poisson model was sufficient. Table 18 presents the summary statistics of the tested variables in the conventional car ownership model. Table 19 shows the Poisson model estimates for vehicle ownership prior to AVs as well as the practical significance of the variables (shown in the last column) in this model.

Results show that the population density of the household home location, number of drivers and workers in the household, number of children, and household average income are practically significant variables. Vehicle ownership rises 85% when the number of drivers in the household is increased by 1 SD. It also rises 24% and 12% by increasing income per household size and number of workers per number of adults in the household, respectively. Vehicle ownership falls by 27% if the population density of households' home location rises by 1 SD.

Table 18. Summary statistics of variables used in household vehicle ownership model (*N* = 125,217)

Variable	Description	Mean	Median	SD	Min	Max
HH_Veh_Count	Household vehicle count	1.98	2	1.18	0	12
Income ($1000)	Household income ($1000)	76.03	62.50	57.96	0	225
Income_per_HHSize ($1000)	Household income per household size	40.30	31.25	33.04	0	225
Workers_per_Adults	# of workers in the HH per # of adults	0.55	0.50	0.45	0	1.00
HHSize	# Household members	2.14	2	1.17	1	13
#Children < 18 yo	# of members younger than 18 years old	0.35	0	0.82	0	8
#Children < 4 yo	No. of household members younger than 4 years old	0.09	0	0.36	0	5

Variable	Description	Mean	Median	SD	Min	Max
#Adults	# of adults in the household (>18 years old)	1.78	2	0.71	1	10
White	Household head is white?	0.83	1	0.38	0	1
#Drivers	# of drivers in Household	1.68	2	0.77	0	9
ln(Home_pop_dens_tract)	Pop. density (persons per sq mi) in the census tract of the household home	7.08	7.31	1.77	3.91	10.31

Table 19. Poisson model to estimate vehicle ownership using 2016/17 NHTS household data

	Estimate	t-Stat	P-value	Pract. Sig.
(Intercept)	0.042	3.84	0.000	
Income_per_HH_Size	0.0032	28.62	0.000	24.49
Workers_per_Adults	0.120	14.52	0.000	12.22
#Children<18 yo	0.013	3.37	0.001	2.32
White	0.087	10.12	0.000	7.41
#Drivers	0.434	92.02	0.000	85.77
ln(Home_pop_dens_tract)	−0.073	−37.33	0.000	−27.09

N = 125,217, Pseudo-R^2: 0.112

To track vehicle ownership in later years of the model, Quarles et al. (2021) used a household-level micro-simulation to model vehicle ownership for 2017–2050. The micro-simulation was calibrated with the survey data and assumes that the rate of acquiring a new vehicle will remain constant in the future. For this purpose, 2017 survey data were first used to model households' WTP for AVs, and all capabilities found in today's HVs are maintained in all AVs. Vehicle purchase year probability is a statistically significant variable inside this model. This variable finds the probability of replacing an HV with a newer HV or an AV in the upcoming year. Thus, a binary logit model was used to find the probability of acquiring or releasing a vehicle.

The study assumed that the WTP for AV technology will increase 5% each year (Quarles et al., 2021). Using the calibrated model and this assumption, the research team simulated a base scenario with a consistent 10% annual decline rate in an AV's purchase price premium (assumed to be $40,000 in 2017) over the simulation period, 2019–2040 (Quarles el al., 2021). The number of new purchased AVs is determined by comparing WTP with the AV technology price in the year of interest.

Figure 41 shows the number of vehicles owned by households in 2019, estimated by the vehicle ownership model prior to AV market penetration. More than 40% of the synthetic households are estimated to have two vehicles and more than 30% are estimated to have one. The mean number of vehicles owned by 2019 synthetic population households is 2.01, compared to almost 1.98 in the 2016/17 NHTS data. The 1.5% difference in the mean values can be attributed to the difference between the year of estimation and the calibration data.

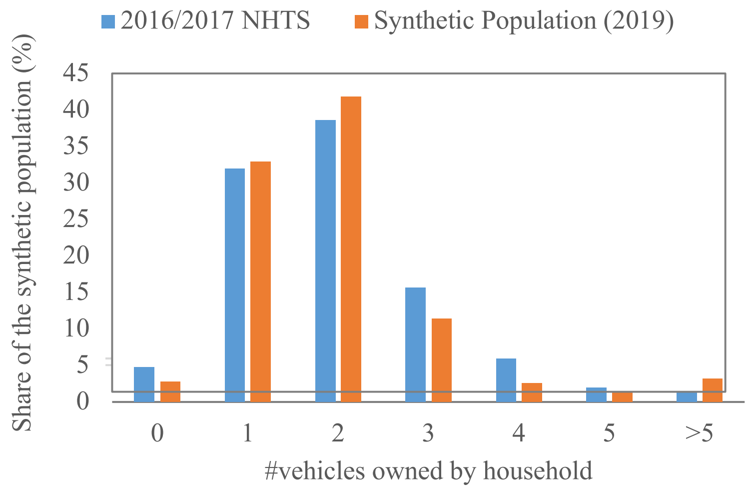


Figure 41. Passenger vehicles owned by synthetic US households in 2019 compared to the 2016/17 NHTS data

5.4. Trip Season

Trip seasonality is tied to trip purpose, mode, and (implicitly) destination choice. For this model, 2016/17 NHTS data were used to predict the season in which the long-distance domestic passenger trips were taken. Table 20 shows multinomial model parameter estimates, with summertime travel as the base alternative. The application of the synthetic population to this model resulted in trips being divided as follows: 30% in summer, 28% in fall, 22% in winter trips, and 19% in spring. The distribution of trips in the NHTS data is 31% summer, 25% fall, 20% winter, and 24% spring (Figure 42).

Table 20. Specifications of the multinomial logit model for trip seasons using 2016/17 NHTS data

	Fall Trip			Winter Trip			Spring Trip		
	Estimate	**t-Stat**	**P-value**	**Estimate**	**t-Stat**	**P-value**	**Estimate**	**t-Stat**	**P-value**
(Intercept)	0.034	0.341	0.733	-0.630	-6.92	0.000	-0.828	-6.55	0.000
Male	0.270	6.16	0.000	0.270	6.16	0.000	0.270	6.16	0.000
Age	-	-	-	-	-	-	0.010	7.55	0.000

	Fall Trip			Winter Trip			Spring Trip		
	Estimate	**t-Stat**	**P-value**	**Estimate**	**t-Stat**	**P-value**	**Estimate**	**t-Stat**	**P-value**
College Educated or Higher	0.167	2.49	0.013	0.217	3.07	0.002	0.117	1.775	0.076
Income ($1000)	0.001	1.45	0.147	-	-	-	-	-	-
HH Size	-0.097	-5.03	0.000	-0.097	-5.03	0.000	-0.097	-5.03	0.000
#Vehicle Owned	0.091	4.88	0.000	0.091	4.88	0.000	0.091	4.88	0.000
Employed?	-0.250	-5.56	0.000	-	-	-	-0.250	-5.56	0.000
#Adults	-0.113	-3.54	0.000	-	-	-	0.084	2.73	0.006

N = 10,455, Adj. Rho^2: 0.0013

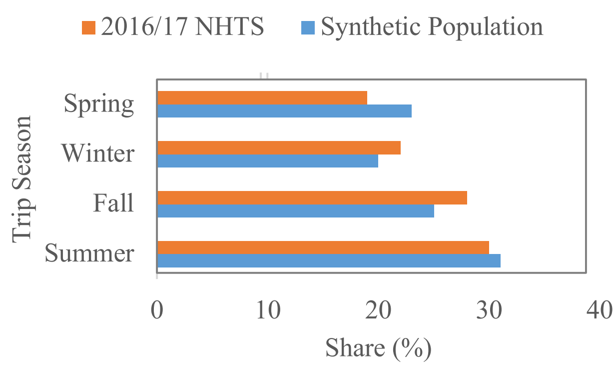


Figure 42. Trip season share in the NHTS data vs synthetic population

5.5. Trip Duration Model

The trip duration model predicts the number of nights that the traveler spends on the journey before heading back home. The outlier for the 85th percentile is 9 nights or longer, while the outlier for the 95th percentile is 24 nights or longer. Several entries show travel durations over one month. These outliers may be explained by home relocation (the traveler never returns) or temporary work relocations. Only trips shorter than 30 days are considered in the model. Table 21 shows the descriptive statistics of the variables that are used in the negative binomial count model.

Table 21. Descriptive statistics of variables used in journey duration model

Variables	Mean	Median	SD	Min	Max
Nights	4.14	3	3.79	0	28
Travel Mode Is Personal Car	0.67	1	0.47	0	1
Travel Mode Is Rental Car	0.08	0	0.26	0	1
Household Annual Income ($1000)	79	68	51.8	5	200
Age	47	50	16	21	70

Variables	Mean	Median	SD	Min	Max
Education—Associate Degree or Higher	0.32	0	0.47	0	1
Number of Vehicles	1.71	2	0.92	0	5
Full-Time Employed	0.45	0	0.50	0	1

Table 22 shows the parameter estimates of the trip duration model. The model shows that married and female travelers have longer journey durations while employed travelers are likely to have shorter trips. Households that are traveling by air or own more vehicles have longer trip durations.

Table 22. Parameter estimates in trip duration (in nights) prediction using population-weighted negative binomial count model

Variable	Coefficient	Std. Err.	P Value
(Intercept)	2.829	0.1083	0.000
Married	1.129	0.091	0.184
Female?	1.403	0.093	0.000
Employed?	0.775	0.0914	0.005
#Vehicles Owned	1.069	0.0481	0.164
Trip Mode Air	1.754	0.0959	0.000

5.6. Trip Purpose

Multinomial logit models were estimated to model purposes of long-distance trips using 2016/17 NHTS trip data. The filtered dataset of long-distance trips included 11,414 observations. The trip purposes considered in the model include commute (9%), business (6%), shopping (16%), personal business (10%), school (1%), medical or dental (3%), religious (1%), visits to friends and relatives (18%), other social recreation (26%), and other purposes (10%). Table 23 presents the summary statistics of all tested variables considered in the multinomial model. These variables include respondents' information such as age, sex, and employment status. It also includes household information, such as household income, number of vehicles owned, household size, and number of employed adults in the household, along with land use variables, such as population density.

Table 23. Summary statistics of variables used in multinomial model using 2016/17 NHTS data (*N* = 11,414)

Variable name	Mean	Median	SD	Min	Max
Trip Distance	189.84	120.13	302.68	75.00	6830.957
Age	52.24	55.00	16.88	12	92
Male?	0.53	1.00	0.50	0	1
Worker?	0.60	1.00	0.49	0	1

Variable name	Mean	Median	SD	Min	Max
Driver?	0.97	1.00	0.17	0	1
College degree or higher?	0.53	1.00	0.50	0	1
HH White?	0.87	1.00	0.33	0	1
HH Black?	0.05	0.00	0.22	0	1
Urban?	0.73	1.00	0.45	0	1
Summer Trip	0.30	0.00	0.46	0	1
Fall Trip	0.28	0.00	0.45	0	1
Winter Trip	0.22	0.00	0.41	0	1
Spring Trip	0.19	0.00	0.39	0	1
#Adults	2.04	2.00	0.66	1	4
#Workers	1.23	1.00	0.95	0	4
HH Size	2.49	2.00	1.14	1	6
#Vehicles	2.48	2.00	1.18	0	6
HH Income ($1000)	95.45	87.50	55.86	10	200
Friday to Sunday Trips	0.48	0.00	0.50	0	1

Multinomial logit models with 10 alternatives are presented in Table 24, which keeps commute trips as the base. These models show that as household income increases, the probability of making long-distance business, shopping, personal, and recreational trips increases as compared to daily long-distance work (commute) trips. However, household income has an inverse impact on medical and dental trips. The most common long-distance trip purpose for employed family heads is a commute. There is a high probability of making business trips in the spring and fall seasons relative to other trip seasons. As their age increases, family heads tend to make more medical or dental and personal trips as compared to commute trips.

Table 24. Specifications of the multinomial logit mode choice model for trip purpose using 2016/17 NHTS data (commute trip is the base purpose)

	Work business trips			Shopping trips		
Response Variable	Estimate	t-stat	P-value	Estimate	t-stat	P-value
(Intercept)	−0.543	−2.53	0.011	2.916	15.04	0.000
Worker?	-	-	-	−2.178	−18.97	0.000
Age	0.012	4.26	0.000	0.007	3.74	0.000
Male?	-	-	-	−0.499	−7.54	0.000
#Workers	-	-	-	-	-	-
Fall trip?	0.738	7.79	0.000	-	-	-
Winter trip?	-	-	-	−0.602	−7.98	0.000
Spring trip?	0.683	7.53	0.000	−0.374	−5.41	0.000
College degree or higher?	0.422	3.60	0.000	0.279	3.84	0.000
HH size	−0.074	−1.94	0.052	−0.126	−4.32	0.000
#Adults	−0.858	−10.18	0.000	−0.436	−8.76	0.000

HH income ($1000)	0.014	14.84	0.000	0.007	9.12	0.000
HH race white?				0.273	3.87	0.000
#Vehicles	−0.101	−2.79	0.005	-	-	-
	Other family/personal business trips			**School trips**		
Response Variable	Estimate	t-stat	P-value	Estimate	t-stat	P-value
(Intercept)	2.498	12.15	0.000	2.051	3.66	0.000
Worker?	−1.870	−15.45	0.000	−3.997	−14.02	0.000
Age	0.013	6.13	0.000	−0.130	−7.82	0.000
Male?	−0.658	−8.97	0.000	-	-	-
#Workers	-	-	-	-	-	-
Fall trip?	−0.247	−3.04	0.002	1.018	4.03	0.000
Winter trip?	−0.556	−6.39	0.000	−0.567	−1.65	0.098
Spring trip?	−0.679	−7.79	0.000			
College degree or higher?	-	-	-	1.980	6.33	0.000
HH size	−0.103	−2.99	0.003	-	-	-
#Drivers	-	-	-	-	-	-
#Adults	−0.188	−3.10	0.002	-	-	-
HH income ($1000)	0.006	6.71	0.000	0.016	7.73	0.000
HH race white?	-	-	-	−0.548	−2.04	0.041
#Vehicles	−0.115	−3.84	0.000	−0.255	−2.63	0.008
	Medical/dental trips			**Religious trips**		
Response Variable	Estimate	t-stat	P-value	Estimate	t-stat	P-value
(Intercept)	−0.156	−0.51	0.613	−1.665	−3.73	0.000
Worker?	−3.244	−16.93	0.000	−2.013	−9.63	0.000
Age	0.041	9.93	0.000	0.020	3.77	0.000
Male?	−0.197	−1.42	0.155	-	-	-
Driver?				-	-	-
#Workers				-	-	-
Fall trip?	0.202	1.34	0.180	-	-	-
Winter trip?	-	-	-	-	-	-
Spring trip?	-	-	-	-	-	-
College degree or higher?	-	-	-	-	-	-
HH size	-	-	-	−0.106	−1.46	0.145
#Adults	-	-	-			
HH income ($1000)	−0.018	−7.68	0.000	0.009	5.90	0.000
HH race white?	-	-	-	-	-	-
#Vehicles	-	-	-	-	-	-
	Social/recreational trips			**Friend/relative visit trips**		
Response Variable	Estimate	t-stat	P-value	Estimate	t-stat	P-value
(Intercept)	3.237	20.47	0.000	2.806	16.11	0.000
Worker?	−2.392	−21.83	0.000	−2.131	−18.84	0.000
Age	-	-	-	0.005	3.19	0.001
Male?	−0.622	−10.38	0.000	−0.731	−11.45	0.000
#Workers	-	-	-	-	-	-

Fall trip?	-	-	-	0.337	5.48	0.000
Winter trip?	−0.616	−9.53	0.000	-	-	-
Spring trip?	−0.663	−10.50	0.000	-	-	-
College degree or higher?	0.391	5.99	0.000	0.358	5.05	0.000
HH size	-	-	-	−0.205	−7.96	0.000
#Adults	−0.419	−11.33	0.000	-	-	-
HH income ($1000)	0.009	12.04	0.000	0.007	9.67	0.000
HH race white?	0.396	6.43	0.000	-	-	-
#Vehicles	-	-	-	−0.202	−8.77	0.000
Other trip purposes						
Response Variable	Estimate	t-stat	P-value			
(Intercept)	−11.123	−6.26	0.000			
Worker?	-	-	-			
Age	0.093	4.20	0.000			
Male?	-	-	-			
#Workers	-	-	-			
Fall trip?	-	-	-			
Winter trip?	-	-	-			
Spring trip?	3.172	3.90	0.000			
College degree or higher?	-	-	-			
HH size	-	-	-			
#Adults	-	-	-			
HH income ($1000)	0.022	5.08	0.000			
HH race white?	-	-	-			
#Vehicles	−0.990	−3.72	0.000			

$N = 11{,}414$, $\rho^2 = 0.2501$

Figure 43 shows the distribution of trip purposes for the synthetic population and the 2016/17 NHTS data. The application of the trip purpose model to the synthetic population resulted in the following division of trip purposes: 11% commute, 7% business, 16% shopping, 4% personal business, 0.02% school, 0.81% medical or dental, 1.37% religious, 31% friend and relative visits, 29% other social recreation, and 0.01% other purposes.

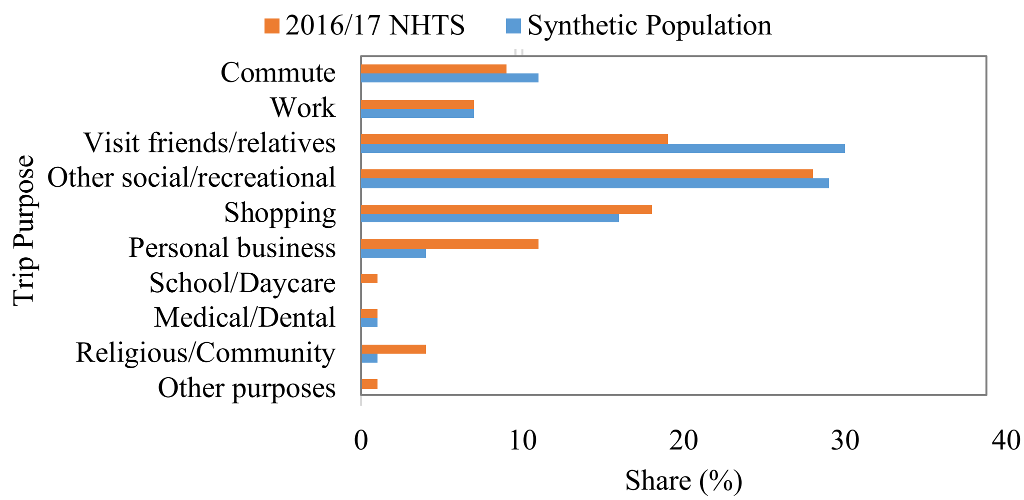


Figure 43. Distribution of trip purposes for the synthetic population and NHTS data

5.7. Party Size

The party size model predicts the number of members, including non-household members, in a travel party. A negative binomial model is used to estimate the number of passengers on each trip (i.e., party size minus one). For this model, 2016/17 NHTS trip data were used, after being filtered to trips longer than 75 miles one-way to be consistent with the definition of long-distance trips in the LD-AV survey. In addition to the NHTS data, the EPA Smart Location data were used to extract job density values in the origin and destination locations of each trip. To align the locations of origins and destinations with the EPA Smart data, the former's coordinates were matched with the Federal Information Processing System (FIPS) codes used in the latter's dataset. For consistency with the NHTS dataset, the party size model controls for the ten trip purpose categories listed in 5.6 Trip Purpose. We will note that in the 2016/17 NHTS dataset, and consistent with the 1990 Nationwide Personal Transportation Survey design, school and religious or community trips are combined into one trip purpose group. However, in this study, the research team divided this group into two trip categories, school/daycare and religious/community, based on the trip origin and destination purposes.

Table 25 summarizes the variables and parameters tested for the party size model using the 2016/17 NHTS and EPA Smart Location datasets. Sample weights were used in the modeling process to ensure that the regression analyses reflect the broader population of interest. Note that the trips reported in the NHTS data were mostly one-way trips. A few long-distance trips (60 among 11,162) were flagged as round trips and reported once in the dataset. For these trips, the distances were divided by two to be consistent with the definition of long-distance trips (i.e., 75 miles one-way). Table 26 presents the negative binomial model to estimate travel party size (minus one, to be consistent with the assumptions of a negative binomial model). Variables' practical and statistical significance are reported in this table. Other tested variables, which exhibited correlation or insignificance, are also presented in Table 25. The last column of Table 26 illustrates the practical significance of the travel party size model estimates. Practical significance values show that party size falls by 3% if age rises 1 SD. It also rises 10% if female variable increases by 1 SD. Table 26 indicates that party size is smaller for commute, work business, shopping, and visiting friends or relatives compared to other trip purposes (i.e., family/personal business, other social/recreational, medical/dental, school, and religious).

Table 25. Summary statistics of variables used in the shifted travel party size model from the 2016/17 NHTS and EPA Smart Location datasets (*N* = 13,665)

Variable	Description	Mean	Median	SD	Min	Max
Party_Size	Party size **minus 1**	1.36	1.0	2.16	0	46
Age	Age	52.30	55	16.92	16	92
Female	Respondent is female?	0.48	0	0.50	0	1
College_Education or higher	Respondent has at least an associate degree?	0.83	1	0.38	0	1
White	Respondent is white?	0.87	1	0.33	0	1
Employed	Respondent is employed?	0.60	1	0.49	0	1
Workers_per_Adults	#Workers in household per # adults	0.60	0.5	0.42	0	2
Income_per_HH_Size	Household income per household size	46.39	42.50	33.51	0.63	225
#Children	Number of children in the household	0.44	0	0.87	0	6
Land Use Data						
ln(O PopDens_tract) (persons per sq mi)	Population density in the census tract of origin	6.39	6.62	1.88	3.91	10.31
ln(D PopDens_tract) (persons per sq mi)	Population density in the census tract of destination	6.43	6.62	1.88	3.91	10.31
Trip Time						
Weekend	Trip is on weekend?	0.32	0	0.47	0	1
Summer_Trip	June, July & August trips	0.35	0	0.48	0	1

Variable	Description	Mean	Median	SD	Min	Max
Fall_Trip	September, October & November trips	0.26	0	0.44	0	1
Spring_Trip	March, April & May trips	0.12	0	0.32	0	1
Winter_Trip	December, January & February trips	0.23	0	0.42	0	1
		Trip Purposes				
Commute_Trip	Commute (to/from work)	0.09	0	0.28	0	1
Work_Business_Trip	Business	0.07	0	0.21	0	1
Shopping_Trip	Shopping	0.18	0	0.37	0	1
Personal_Business_Tri p	Personal business other than shopping	0.11	0	0.29	0	1
Visit_Friends_Relativ es	Visiting friends or relatives	0.28	0	0.37	0	1
School_Trip	To/from school or daycare	0.01	0	0.10	0	1
Religious_Community _Trip	To/from church/community/re ligious activities	0.04	0	0.15	0	1
Medical_Dental_Trip	Medical or dental treatments	0.01	0	0.16	0	1
Other_Social_Recreati onal	Other social or recreational	019	0	0.44	0	1
Other_Trip_Purposes	Any other trip purpose not listed above	0.01	0	0.33	0	1

Table 26. Negative binomial model for shifted party size (Y = party size minus 1)

	Estimate	Stand. err.	P-Value	Pract. Sig.
Theta	2.087	0.395	0.000	
(Intercept)	0.628	0.065	0.000	
Age	−0.004	0.001	0.000	−3.07
Female	0.204	0.065	0.002	10.61
Commute_Trip	−1.305	0.120	0.000	−24.56
Business_Trip	−1.151	0.191	0.000	−18.65
Shopping_Trip	−0.220	0.068	0.001	−6.71
Visit_Friends_Relatives	−0.226	0.064	0.000	−6.73

N = 13,665, Pseudo-R^2 = 0.0396

Figure 44 illustrates the estimated party size for the synthetic population prior to and after AVs are in the market. The application of this model showed no significant change in party size after AVs are introduced. The mean party size both prior to and after AVs are in the market is 2.04.

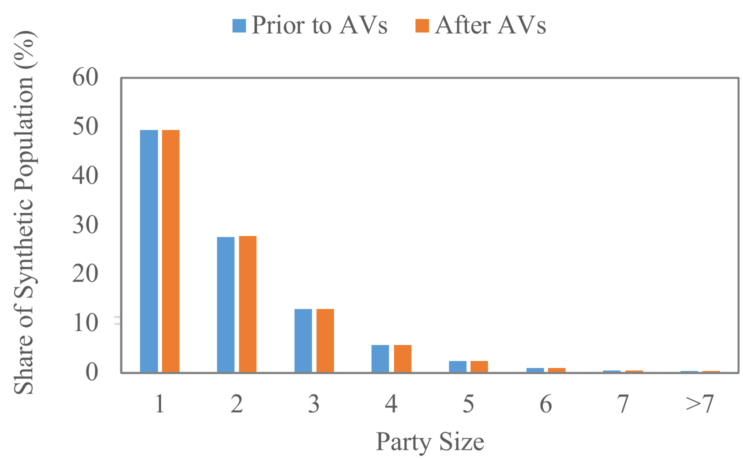


Figure 44. Distribution of party sizes for the synthetic population

5.8. Mode Choice

The mode choice models were estimated using the LD-AV survey data collected in this project. The revealed mode preferences (RP) for the reported trips in the LD-AV survey were first used to estimate mode choice among the currently available choices. Stated preferences (SP) for future long-distance travel scenarios were used to estimate mode choice for business and non-business trips once AVs are an option, in a joint RP-SP model. Here, we present the joint mode choice models for business and non-business trips. The research team excluded bus, rail, and boat/ship from the choices due to the insufficient data available for these three modes. As the SP questions involving AVs did not present specific trip scenarios, the origin and destination locations, which were needed to extract travel time and cost skims, were taken from the respondents' RP scenarios.

Thanks to geocoding that provides a match between the rJourney zones (FHWA, 2015) and the reported cities of origin and destination in the LD-AV survey, time and cost estimates for different modes were obtained from the rJourney skims dataset. The availability of different modes between origin and destination locations was also extracted from the rJourney data, under the assumption that OD pairs without any distance and time skims for a specific mode are not accessible through that mode. The summary of variables used in the mode choice models is presented in Table 27.

Table 27. Summary statistics of variables used in the mode choice model (*N* = 809)

	Mean	Median	SD	Min	Max
Age	47.3	50	16	21	70
White	0.8	1	0.42	0	1
Education_College or Assc. higher	0.3	0	0.46	0	1
#Adults	1.4	1	1.32	1	9
#Children < 18 yo	0.4	0	0.73	0	6
HH_Veh_Count	1.0	1	0.19	0	1
Disabled	0.1	0	0.34	0	1
Employed	0.6	1	0.49	0	1
Married	0.5	1	0.5	0	1
Income ($1000)	79	68	51.8	5	200
O PopDens (/sq mi of county)	3,292	1,380	10,392	0.4	87,552
O JobDens (/sq mi of county)	2,819	704	14,843	0.0	134,124
D PopDens (/sq mi of county)	1,417	1,391	557	417.8	5,799
D JobDens (/sq mi of county)	8,861	557	41,391	3.1	798,859
PartySize	0	0	0.25	0	4
MoreThan_1_night_Trip	0.8	1	0.37	0	1
TravelTime Car (hr)	13.6	9.5	14.63	0	193.5
Distance Car (mi)	838.4	574	9.15	0	12,124
TravelTime Air (hr)	2.6	2.1	2.59	0	31
AccessDistance Air (mi)	24.8	13	30.5	0	294
Toll Car ($)	0.7	0	1.51	0	17.7
LongDistance (>500 mi)	0.6	1	0.48	0	1
Fare Air ($)	691.7	550	658	0	6903

Mode and destination choices for long-distance domestic trips were estimated for business and non-business trips in a joint model prior to and after AVs become available, using the 2021 LD-AV survey. Less common existing long-distance modes (including bus, rail, and boats) were not included, so only air, rental car, personal car, and AVs were permitted. To consider chain trips, we summed the time and costs of all trip legs. The specifications of the joint RP-SP logit models for non-business trips with AVs are presented in Table 28. The operational cost for different AVs was assumed to be $0.70 per mile. The operational cost of personal HVs was assumed to be $0.50 per mile, while $50 per driving day (minimum 1 day) in addition to $0.10 per mile were assumed for a rental car. To avoid the correlation between travel costs and times, the residuals of travel cost from travel times were considered in the mode choice models. The specifications of the mode choice model for non-business trips show that users are more willing to use airplanes for trips longer than 500 miles. In addition, AV choice has an inverse relationship with driver age and a direct relationship with drivers having at least a college or associate degree. Rental cars had a higher utility for trips with larger party size.

Table 28. Specifications of the logit mode choice model after AV introduction using joint SP-RP LD-AV survey, EPA Smart, and RSG rJourney data

	Estimate	t-ratio	P-value
ASC car	0	-	-
ASC air	−1.187	−7.464	0.000
ASC rental car	−0.710	−10.803	0.000
ASC AV	−0.090	−0.291	0.385
Travel time car	−0.281	−5.469	0.000
Travel time air	−0.270	−2.282	0.011
Travel time rcar	−0.103	−3.618	0.000
Travel time AV	−0.113	−4.815	0.000
Access distance air	−0.028	−3.666	0.000
Residual of cost from travel time	−0.002	−3.777	0.000
Long-distance>500 mi air	1.914	4.120	0.000
Party size rental car	0.129	2.591	0.005
Female car	−0.207	−1.336	0.091
Age AV	−0.023	−3.472	0.000
Associate degree AV	0.725	2.459	0.007
μ revealed preference	1.000	-	-
μ stated preference	0.752	11.398	0.000

N = 584, R^2: 0.3513

After AVs are in market in the future (e.g., in year 2040) with AV technology premium of $3,500, 61% of households are estimated to have AVs; the mode shifts are shown in Figure 45. Note that the technology premium changes over time and the $3500 price is based on the assumption that AV technology premium was $40,000 in 2017 and drops 10% annually.

Figure 45. Mode share shift after AVs are in market with technology cost of $3,500

5.9. Destination Choice Model

Multinomial logit models were estimated to predict destination choice of domestic trips using 2016/17 NHTS trip data. The possible destination zones were the 4,484 NUMAs as defined in rJourney (FHWA, 2015), comprised of combinations of census tract and county boundaries. Mode choice logsums were calculated using the before-AV mode choice model. To match the NHTS trip data with the NUMA zones, the coordinates of the origin and destination locations of trips further than 75 miles reported in the 2016/17 NHTS survey were geocoded to the NUMA centers. There are 15,120 long-distance trips in the 2016/17 NHTS trip data, which includes 1,802 work-related trips and 10,637 personal trips, with the rest categorized as other trip purposes. These trips are made to almost 2,000 destination zones. An aggregate model, including attractions (eight different employment counts at the destination location), distance, mode choice logsum, and population and housing density, was tested in the destination choice models. Note that only the logsum of the generalized cost part of the mode choice utility function is used for the destination choice model. Two categories were considered for trip purposes, work-related and personal. Land use data were extracted from the EPA Smart Location data by mapping NUMAs to county FIPS codes. The summary of land use data in destination choice models is presented in Table 29. In addition, travel time and cost skims from rJourney data (FHWA, 2015) were used for the mode choice logsum estimations.

Given the large choice set, i.e., the many destinations available, computational time was one of the main concerns in the estimation of this model. Thus, a strategic sampling approach, presented by Lemp and Kockelman (2012), was employed, in which the alternatives are drawn in proportion to updated choice-probability estimates. This approach has two iterations. In the first iteration, a simple random sampling was done among all alternatives. The second part of this approach uses the choice probability estimates for a strategic sampling of alternatives. The sampling was conducted for 20, 50, and 100 alternatives. The destination choice models with the strategic sampling of 100 alternatives are presented in Table 30. These models show that entertainment, retail, service, and education jobs are important contributors to destination choice for non-work trips. Meanwhile, the number of retail, office, service, industrial, and public administration jobs are significant in the work trip model.

Table 29. Summary of EPA Smart Location data used in the destination choice model

	Mean	Median	SD	Min	Max
Total Population in Tract	4,496	4,153	2,650	75	51,536
Destination Population Density (persons per acre in tract)	2.97	0.11	11.42	0	237
Tract's Land Area (sq mi)	389	65	2,036	0.02	90,576
Total Employment in Tract	1,967	985	4,407	0	139,713
#Retail Jobs in Tract	191	77	352	0	4,740
#Office Jobs in Tract	163	34	1,032	0	37,702
#Industrial Jobs in Tract	618	234	1,550	0	34,471
#Service Jobs in Tract	277	76	1,347	0	59,864
#Entertainment Jobs in Tract	212	70	1,306	0	81,796
#Education Jobs in Tract	195	69	756	0	26,559
# Medical Jobs in Tract	223	80	539	0	14,718
#Public Administration Jobs in Tract	87	13	276	0	7,114

Table 30. Destination choice model parameter estimates

	Non-Business Trips			Business Trips		
	Estimate	t-Stat	P-Value	Estimate	t-Stat	P-Value
Mode choice logsum	0.017	122.96	0.000	0.011	50.49	0.000
Destination population density at the tract level (persons/sq mi)	0.002	1.60	0.109	0.005	2.61	0.009

	Non-Business Trips			Business Trips		
	Estimate	t-Stat	P-Value	Estimate	t-Stat	P-Value
#Retail jobs in tract	−0.068	−8.62	0.000	−0.049	−2.38	0.017
#Industrial jobs in tract	0.027	3.20	0.001	0.021	1.04	0.297
#Service jobs in tract	0.019	2.17	0.030	0.057	2.56	0.010
#Public administration jobs in tract	−0.019	−3.90	0.000	-	-	-
#Medical jobs in tract	-	-	-	−0.044	−2.81	0.005
	N = 9,325, Pseudo-R^2: 0.0597			*N* = 1,802, Pseudo-R^2: 0.060		

Figure 46 summarizes the results of the destination choice model for the synthetic population. Assuming a $3,500 technology cost premium (e.g., in year 2040), total PMT per capita in long-distance trips is estimated to rise 24% (from 250 to 309 miles per month). The PMT in HVs is estimated to fall from 171 to 101 per capita per month, rental cars from 46 to 28, and airplanes from 33 to 31, while the PMT in AVs and SAVs are predicted to be 83 and 66 per capita per month, respectively.

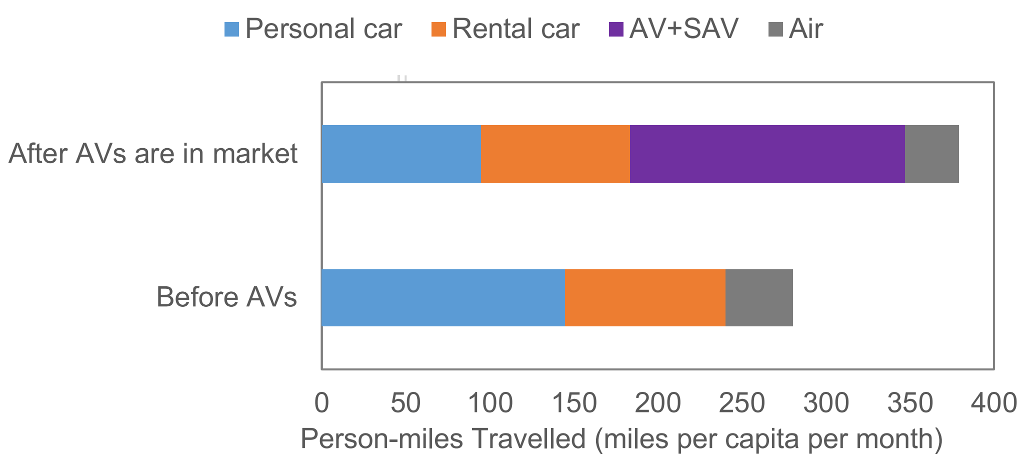


Figure 46. Shift in PMT after AVs are in market

5.10. Other Models

A selection of variables was used as inputs for different models, as explained earlier. Some variables such as number of drivers in the household, retirement status, and distance from home to transit, grocery store, and work were not available in the synthetic population. Thus, the research team used different datasets to estimate such variables. Table 31 summarizes the estimated binomial logit model for being retired using 2016/17 NHTS data.

Table 32 summarizes the Poisson model for the number of drivers in a household. Table 33 presents the negative binomial model for the age of the oldest vehicle owned by a household using the 2017 AV survey data (Quarles et al., 2021).

Table 34 shows the ordinary least squares models estimated for distance to transit and grocery store from a household's location. Table 34 shows the ordinary least squares models estimated for distance to transit and grocery store from a household's location. Table 35. Binary logit model to predict probability of a person making LD trip on any given day using 2016/2017 datagives estimates for binary logit model to predict probability of a person making LD trip on any given day using 2016/2017 data.

Table 31. Binomial logit model for retirement status using 2016/17 NHTS person data

	Estimate	t-Stat	P-Value
(Intercept)	−9.264	−151.20	0.000
Employed?	−3.583	−172.80	0.000
Age	0.159	165.70	0.000

N = 235,904, R^2: 0.559625

Table 32. Count model for number of drivers in the household using 2016/17 NHTS household data

	Estimate	t-Stat	P-Value
(Intercept)	−0.312	−16.86	0.000
Income_per_HH_Size ($1000)	0.0012	15.59	0.000
Workers_per_Adults	0.239	37.41	0.000
Num_Adults	0.404	76.22	0.000
White	0.109	15.62	0.000
ln(Home_Pop_Dens_tract)	−0.034	−23.63	0.000

N = 125,217, R^2: 0.242754

Table 33. Negative binomial model for age of the oldest vehicle using 2017 AV survey data (Quarles et al., 2021)

	Estimate	t-Stat	P-Value
(Intercept)	1.601	35.34	0.000
HH_Veh_Count	0.596	23.52	0.000
Income ($1000)	-0.005	-10.57	0.000

N = 1414, R^2: 0.2542631, Theta: 2.317

Table 34. Ordinary least squares models for distance from home to transit and grocery store using 2017 AV survey data (Quarles et al., 2021)

	Distance to transit (mi)			Distance to grocery (mi)		
	Estimate	t-Stat	P-Value	Estimate	t-Stat	P-Value
(Intercept)	7.907	16.04	0.000	5.151	18.63	0.000
Income ($1000)	−0.018	−3.11	0.002	−0.005	−1.67	0.095

	Distance to transit (mi)			Distance to grocery (mi)		
	Estimate	t-Stat	P-Value	Estimate	t-Stat	P-Value
#Children	0.835	2.86	0.004	0.620	3.79	0.000

N = 1414, R^2: 0.01107

Table 35. Binary logit model to predict probability of a person making LD trip on any given day using 2016/2017 data

	Estimate	P-Value	Pract. Sig
Intercept	-4.240	0.000	
Male	0.177	0.000	0.159
College education	0.369	0.000	0.292
Employed	0.162	0.002	0.145
HH income ($)	3.94E-06	0.000	0.391
# Vehicles owned	0.087	0.000	0.193
February	0.197	0.061	0.097
March	0.277	0.006	0.150
April	0.469	0.000	0.185
May	0.252	0.036	0.101
June	0.298	0.014	0.116
July	0.686	0.000	0.374
August	0.610	0.000	0.338
September	0.340	0.001	0.186
October	0.332	0.002	0.155
November	0.394	0.000	0.203
Monday	-0.653	0.000	-0.424
Tuesday	-0.880	0.000	-0.573
Wednesday	-0.809	0.000	-0.525
Thursday	-0.500	0.000	-0.328
Friday	-0.247	0.002	-0.161
Saturday	-0.142	0.074	-0.079
Holiday	0.297	0.023	0.087
Close2holiday	0.214	0.001	0.156
Christmas Week	0.540	0.000	0.164

N = 228,595, *Rho*^{2}: 0.035

Figure 47. Number of Long-Distance Trips per Day over the Year 2016 shows variations of number of long-distance trips over each day from April 2016 to April 2017 while Figure 48. Percentage of Long-Distance Trips over the Day of Week for Year 2016shows percentage of long-distance trips over the day of week for year 2016. Figure 49. Percentage of Long-Distance Trips over the Day of Week for Year 2016 shows percentage of long-distance trips over the month's year 2016/17.

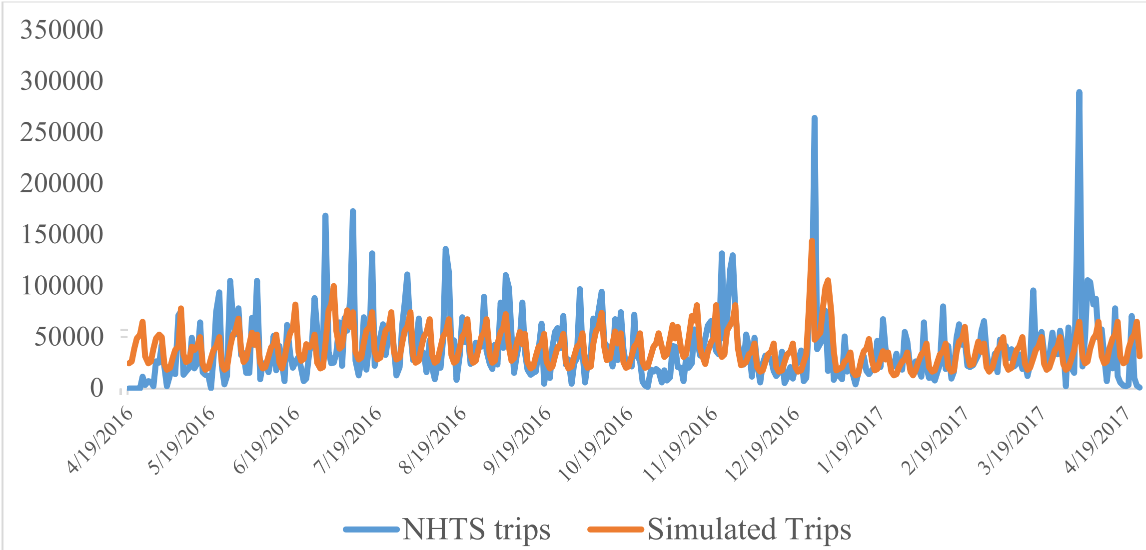


Figure 47. Number of Long-Distance Trips per Day over the Year 2016

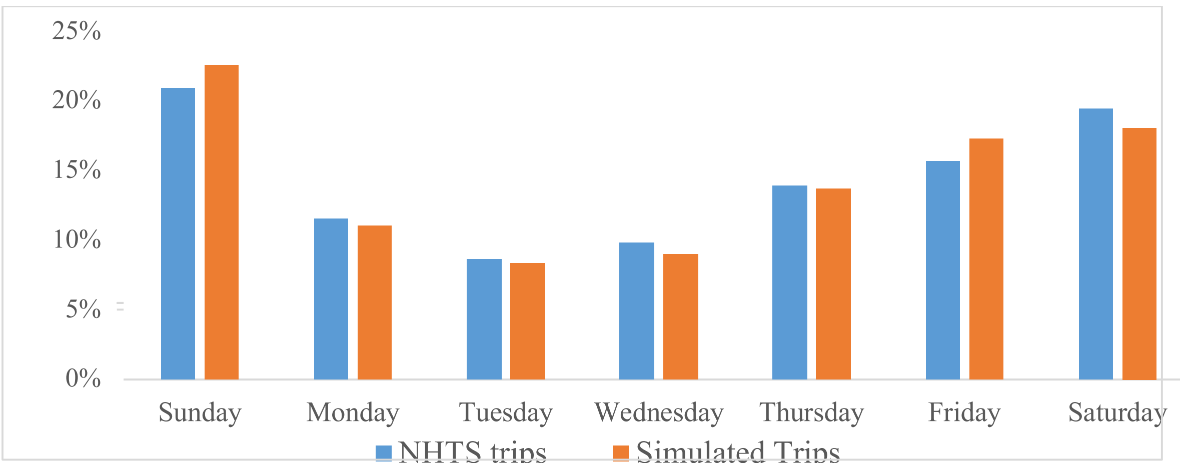


Figure 48. Percentage of Long-Distance Trips over the Day of Week for Year 2016

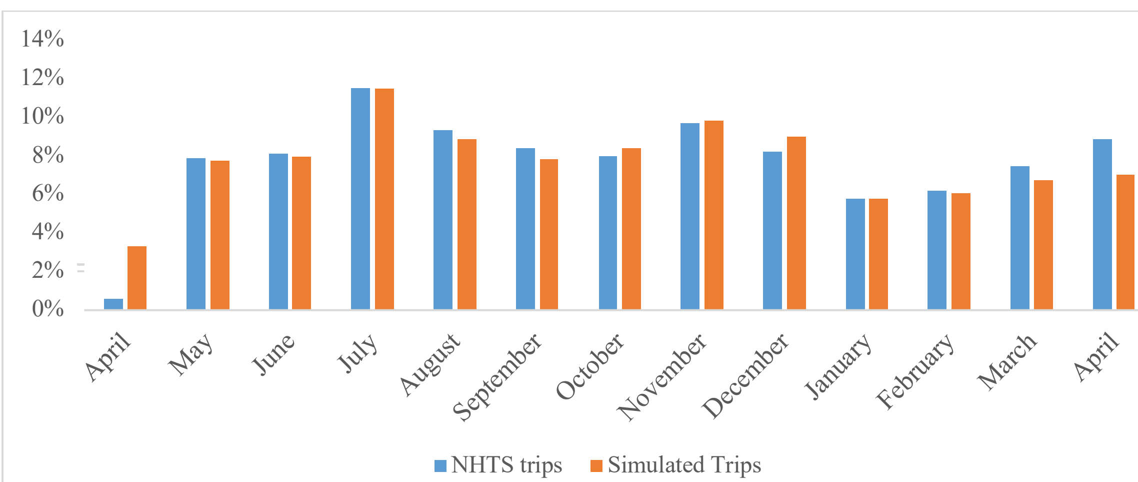


Figure 49. Percentage of Long-Distance Trips over the Day of Week for Year 2016

5.11. Trip-Chaining in NHTS Sample

The aforementioned model outcomes are applicable to individual trip links exceeding 75 miles from the NHTS trip file. In the course of this investigation, we have treated each trip segment from the NHTS trip file as an individual trip, including those segments that entail brief stops en-route to the primary destination. For example, a 150-mile trip from Austin to Houston that involves a refueling stop shows as two separate trips (one for “shopping” midway, and one for the true purpose of the trip at the final destination). Long-distance (LD) trips are defined as those with the real destination (not an en route “pit stop” for gas or food) more than 75-miles away (on the travel network). To address the limitation in segments vs true trips, we developed an algorithm to identify intermediate “trips” that are really just pit stops in a longer-distance trip-chain en route to a final destination (resulting in fewer true LD trips). The algorithm is also designed to fuse “short” (less than 75-mile one-way) trip segments into long-distance (LD) trips (more than 75-miles one-way) as well as “long” (>75 miles) and “short” (<75 miles) trip segments with each when they can be considered part of chain to the final destination. Thus, the number of LD trips may rise or fall, relative to a simple counting of segments that are 75-miles or longer (which is the technique that most analysts use). The algorithm reflects trip coordinates (to appreciate trip direction, thereby avoiding back and forth trips or tours to many true destinations [like delivery chains]), dwell times at “destinations” (at the end of every trip segment, to avoid counting relatively short “pitstops”), and trip purpose (to distinguish refueling and meals en route, for example, from a longer-duration final-destination activity).

After applying the algorithm’s many rules, the NHTS 2016/17 sample’s LD trips (i.e., those more than 75 miles one-way) fell by just 3.4% (from 1.84 one-way LD trips per American per month to 1.78). This brings the total number of NHTS person-trips down by 0.88%, with the average American making 3.47 trips per day instead of the previously estimated 3.50 trips. Notably, the algorithm was only applied to trips or chains of trips that met the criteria for long-distance travel (those segments more than 75 miles each or a series of related segments adding to more than 75 miles one-way), ensuring that it did not impact shorter trips.

The best predictor in distinguishing pit stops are the coordinates (lat and long) of stopping points Successive trip segments by each NHTS respondent were used to discern whether distance from the origin kept rising in a directed way, or started pivoting (or even falling), indicating changes in direction. If the Euclidean distance from the chain origin started falling (after rising after earlier stops), it indicated a return trip. Combined with other factors (including mode changes, site activity/trip-end purpose, and short-activity durations), these falling distances or changes in direction helped distinguish trip chains. Mode shifts (like driving to or taking a bus to an airport, and changing planes at a hub airport) are not real destinations. Additionally, stopping away from the origin during a long-distance trip, to purchase food or gasoline, is often not a true destination - especially when the stop is short and trip direction unchanged.

Meal and carry-out stops shorter than 90 minutes and general errand stops (like to a post office) shorter than 15 minutes are assumed to be part of a longer trip. While the algorithm is designed to identify and classify long-distance trips more accurately, there are limitations to pitstop inference. For example, it is challenging to determine whether a person stopped to purchase gas or a breakfast coffee when setting out on a long-distance trip. The NHTS "WHYTO" purpose categories (shown above) do not distinguish fuel stops from grocery or clothing-shop stops, and do not distinguish errand types (like a visit to the post office versus a library), making it difficult to categorize these as true destinations or pit stops along the way. To address this challenge, a destination dwell time threshold of 30 minutes (for fuel/shopping) and 15 minutes (for errands) is assumed to help identify longer stops that may indicate a true purpose. Some of these stops should be considered as necessary, separate, or true destinations that would have been made regardless of the longer trip that day. Similarly, stops near the end of a long-distance trip may be trips that would have been made anyway upon arrival at the destination. Travelers have many options in how they chain trips, and some important destinations may be along one's long-distance trip route.

When "pit stops" (short stops, typically to eat, refuel, change modes, etc.) on long-distance trips are no longer counted as destinations (thanks to the algorithm's application), only the attributes of the final leg of a trip chain (variables of travel day trip purpose [WHYTRP90] and trip purpose summary [WHYTRP1S]) are used to determine the LD trip purpose. This approach reduces the shares of LD trips taken for commutes plus work trips, shopping, meals/food, and other volunteer activities/change in mode (indicated by the "97= Something else" purpose) by 10.8, 35.5, 74.2, and 54.9 percentage points, respectively. Removing such pit stops raises the shares of (1) school plus religious trips, (2) medical trips, (3) transporting someone, and (4) social trips (visiting friends and relatives) plus recreational trips, by 14.8, 16.7, 20.8 and 21.0 percentage points, respectively. Average and median LD person-trip lengths also rise (after applying the algorithm), by about 10 percent: from 268.5 to 289.9 (average LD trip) miles and from 129.2 to 138.5 (median) miles, respectively. This algorithm considers the location, timing, and sequence features of sample "trips" to infer the real reasons behind long-distance travels (where many "long" [> 75-mile] and/or "short" [< 75-mile] segments may be describing a single long-distance trip). Overall, this algorithm provides a more accurate and comprehensive understanding of true long-distance travel patterns, accounting for the complex nature of multi-stop trips and ensuring that the resulting data is meaningful for transportation planning and policy. This discrepancy has the potential to alter the outcomes of trip frequency and trip purpose models. Nevertheless, we anticipate that the increase on long-distance trips is approximately 10%.

Table 36. Purposes Before and After Applying the Trip Chaining Algorithm

		Without Trip Chaining (Prev Method)		With Trip Chaining		% Change
		# of trips	% of Total	# of trips	% of Total	
# of LD Trips (> 75 miles one-way) in NHTS Sample		15972	1.73%	15434	1.69%	-3.3684%
# of Person-Trips in NHTS Sample		923573		915457		-0.879%
# Trips per Day per American		3.495		3.465		
WHYTRP90 (Travel day trip purpose)	01=To/From Work	1146	7.2%	1342	8.7%	17.1%
	02=Work-Related Business	809	5.1%	941	6.1%	16.3%
	03=Shopping	2350	14.7%	1562	10.1%	-33.5%
	04=Other Family/Personal Business	1368	8.6%	1609	10.4%	17.6%
	05=School/Church	421	2.6%	506	3.3%	20.2%
	06=Medical/Dental	351	2.2%	421	2.7%	19.9%
	08=Visit Friends/Relatives	2534	15.9%	3025	19.6%	19.4%
	10=Other Social/Recreational	3900	24.4%	3594	23.3%	-7.8%
	11=Other (such as change of mode)	3083	19.3%	2423	15.7%	-21.4%
	99=Refused / Don't Know	10	0.1%	11	0.1%	10.0%
WHYTRP1S (Trip purpose summary)	01=Home	4537	28.4%	6010	38.9%	32.5%
	10=Work (and work related business)	2138	13.4%	1870	12.1%	-12.5%
	20=School/Daycare/Religious activity	271	1.7%	311	2.0%	14.8%
	30=Medical/Dental Services	228	1.4%	266	1.7%	16.7%
	40=Shopping/Errands	1854	11.6%	1196	7.7%	-35.5%
	50=Social/Recreational	3316	20.8%	4013	26.0%	21.0%
	70=Transport Someone	607	3.8%	733	4.7%	20.8%
	80=Meals	1693	10.6%	436	2.8%	-74.2%
	97=Something Else (such as unpaid volunteer activities and change of transportation mode)	1328	8.3%	599	3.9%	-54.9%

Freight Model Specification

The implementation of autonomous trucking will bring sweeping changes to the world of freight transport. Semi-automated trucks may enable automated driving under supervision and limited circumstances, such as driving long distances on an interstate highway. Eventually, fully automated self-driving trucks (or ATrucks) may leave the truck terminal and travel to a destination without human intervention or presence in the truck cab (Viscelli, 2018). ATrucks may be able to automate other functions in addition to onboard tasks like drop-offs and pickups, but most experts expect an attendant will remain on board, doing other types of work, sleeping as needed, and ensuring thoughtful deliveries and pickups (Yankelevich, et al., 2018). Such multi-tasking of vehicle attendants will allow for extended use of commercial trucks (i.e., every day, almost 24 hours a day) and greater labor productivity, resulting in lower per-mile and per-ton-mile freight delivery costs.

In 2014, trucks carried 1.996 trillion ton-miles of freight around the US, 37.7% of the nation's total ton-miles transported that year (BTS, 2017). Investment in and use of ATrucks will affect not only national and regional economies (Clements and Kockelman, 2017) but trade patterns, production levels, and goods pricing. Commercial trucks consume about 20% of the nation's transportation fuel, and self-driving technologies are predicted to reduce those diesel fuel bills by 4–7%

(Liu and Kockelman, 2017; Barth et al., 2004; Shladover et al., 2006). ATrucks can reduce some environmental impacts, lower crash rates, and increase line-haul transportation. Platooned convoys should enable following truck attendants to avoid certain restrictions on service hours, enabling longer driving distances. Uranga (2017) predicts greater use of ATrucks before passenger vehicle automation, thanks to the more obvious economic benefits of self-driving trucks.

While there is active investigative interest in the travel and traffic effects of self-driving cars, research into the travel and traffic impacts of ATrucks is dearly lacking. This project leverages Freight Analysis Framework 5 (FAF5) to estimate the mode and origin choices of freight carriers; then, model parameters are fed into the random-utility-based multi-regional input-output (RUBMRIO) model to investigate how patterns of freight flow will change based on automation technology's impacts on truck cost and operation. ATrucks' per-mile operating cost is assumed to be 50% of that of HTrucks', factoring in the benefits of increased safety, a lower wage for truck drivers, and a higher initial cost (e.g., purchase of tractor) to introduce ATrucks.

The remainder of this chapter proceeds through each component of the freight models, showing how it is estimated and parameterized and offering insight into significant trends that will affect the overall RUBMRIO model.

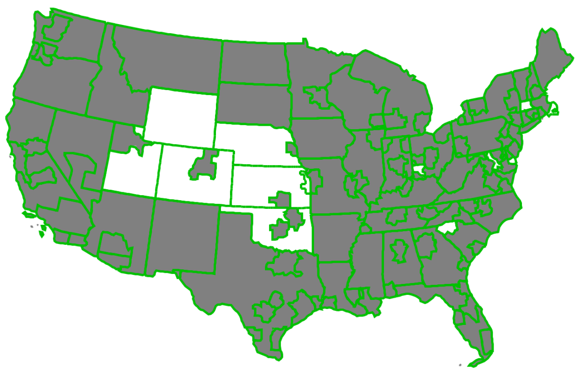
6.1. Datasets and Model Parameter Estimation

This section introduces the prepared datasets for the freight model, as well as the model estimations. The estimated models will be used in the RUBMRIO model specified in section 6.2.

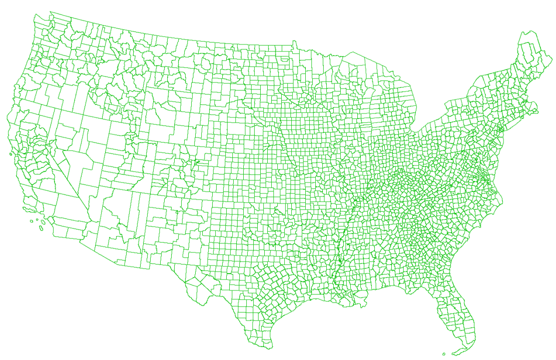
6.1.1. Freight Analysis Framework (FAF5) Data

FAF5 integrates trade data from a variety of industry sources, with emphasis on the Census Bureau's 2017 Commodity Flow Survey (CFS) and international trade data (Census Bureau, 2021). It provides estimates of US trade flows (in tons and dollar value) by industry, across 8 modes (truck, rail, water, air, multiple modes and mail, pipeline, non-domestic, and others), and between FAF5's 129 aggregate zones within the US. FAF5's origin-destination-commodity-mode annual freight flows matrices were used to predict domestic and export trade flows by zones. FAF5 data show foreign export flows exiting the US from 117 of these 129 zones, as shown in gray in Figure 50a. Consequently, these same 117 zones serve as both production and export zones in this paper's trade modeling system.

FAF5 zones were then disaggregated into county-level matrices using the 2017 CFS boundary data (which identify the counties belonging to each FAF5 zone). Ten metro areas were also added to the CFS data in 2017, leading to 3,109 contiguous counties (as shown in Figure 50b) after excluding Hawaii and Alaska. Interzonal travel times and costs by rail, ATruck, and HTruck were all computed for the 3,109-by-3,109-county matrix using the shortest highway and railway paths in terms of free-flow travel time. All intracounty travel distances were assumed to be the radii of circles having that county's same area.



(a)



(b)

Figure 50. Continental US domestic and export zones for trade modeling: (a) FAF5's 129 zones, including the 117 export zones (shown in gray), and (b) its 3,109 domestic freight counties.

6.1.2. Freight Mode Choice Model

The freight mode choice model serves as a key component in the RUBMRIO model, as it distributes the freight flow for an OD pair by mode. The freight mode choice model was estimated by leveraging data assembled from different sources. FAF5 freight flow data provides freight flow records, with skims supported by FAF4, rJourney, and county-level population data. Due to the large gap in magnitude of the transported values of different commodities, one mode choice model was estimated for each commodity. FAF5 provides 42 different types of commodities. They are aggregated further to 20 types, which can be matched to the input-output table. Sector/Industry 3 is considered similar to sector 2, so they are estimated as the same category. Ton-miles of each commodity transported between OD pairs by mode are used as the weights for each freight flow record, and weights are further normalized and transformed using a log function to maintain a reasonable scale. Sectors 14 to 20 do not have specific SCTG code, so their parameters are the average of all other sectors. Table 35 also shows the IMPLAN (Impact Analysis for Planning) code matching with other freight code to utilize the input-output (IO) table.

Table 37. Description of economic sectors in RUBMRIO model

Sector	Description	IMPLAN Code	NAICS Code	SCTG Code	Ton-Miles Transported in 2017 (billion)
1	Agriculture, Forestry, Fishing, and Hunting	1 to 19	11	1	16.6
2	Mining	20 to 30	21	10 to 15	791.4
3	Construction	34 to 40	23		
4	Food, Beverage, and Tobacco Product Manufacturing	41 to 74	311, 312	2 to 9	965.2
5	Petroleum and Coal Product Manufacturing	115 to 119	324	16 to 19	337.8
6	Chemicals, Plastics, and Rubber Product Manufacturing	120 to 152	325, 326	20 to 24	343.9
7	Primary Metal Manufacturing	170 to 180	331	32	104.8
8	Fabricated Metal Manufacturing	181 to 202	332	33	57.4
9	Machinery Manufacturing	203 to 233	333	34	58.6
10	Computer, Electronic Product, and Electrical Equipment Manufacturing	234 to 275	334, 335	35, 38	44.7
11	Transportation Equipment Manufacturing	276 to 294	336	36, 37	83.7
12	Other Durable and Non-Durable Manufacturing	75 to 114, 153 to 169, 295 to 304	313 to 316, 321 to 323, 327, 337	25 to 31, 39	481.6
13	Miscellaneous Manufacturing	305 to 318	339	40, 41, 43	119.8
14	Transportation, Communication and Utilities	31 to 33, 332 to 353	22, 48, 49, 51	--	--

Sector	Description	IMPLAN Code	NAICS Code	SCTG Code	Ton-Miles Transported in 2017 (billion)
15	Wholesale Trade	319	42	--	--
16	Retail Trade	320 to 331	44, 45	--	--
17	FIRE (Finance, Insurance and Real Estate)	354 to 366	52, 53	--	--
18	Services	367 to 440	54 to 56, 61, 62, 71, 72, 81, 92	--	--
19	Household	--	--	--	--
20	Government	--	--	--	--

For each model, four skim tables are used as variables in the utility function: truck travel time, truck cost, rail travel time, and rail cost. Truck travel time was derived from the 2010 rJourney data, which provides passenger travel times between NUMAs across the US. NUMAs are zones utilized for the rJourney model that are comprised of counties or PUMAs. The origin and destination's population-weighted travel times at the NUMA level are aggregated into FAF zone level to offer an average passenger travel time between FAF zones. Since trucks require more highway travel time than passenger vehicles, the following equation is used to convert a passenger vehicle's highway travel time to a truck's. (Cambridge Systematics, 2002):

$$t_{ij,truck} = \text{(highway distance in miles / 45 mph)} + \text{Floor ((highway distance in miles / 45 mph) / 10 hours)} \cdot \text{14 hours} \qquad (1)$$

Rail time is calculated based on rail distance, adjusted following the equation in Texas's statewide analysis model version 4 (SAM4) to show a fixed 30 hours' dwelling time and an average speed of 21.72 miles per hour on railways:

$$t_{ij,rail} = \text{30 hours} + \text{rail distance / 21.72 mph} \qquad (2)$$

In addition, to maintain the travel time variable under a reasonable scale for both truck and rail, two transformations were applied. The first was to change from hours to minutes by multiplying a factor of 60, and a log transformation was further applied to travel time in minutes.

Truck cost and rail cost are calculated based on travel distances. The National Private Truck Council (2021) gives an average truck cost for 2021 of $2.90 per mile, with drivers' wages accounting for 40% and fuel for 12%. Rail's cost was about $1.59 per mile on average in 2019 (Ashe, 2022). Since ton-mile data is not available in FAF5, truck and rail distances are calculated based on FAF4, by dividing total ton-miles by total tons transported to get the average distance per ton for each FAF's OD pair by commodity. The cost terms are used directly, without a log transformation, as it is easier to normalize the cost to a dollar unit when using the equations in the RUBMRIO model.

With freight flow records and skim tables obtained, the model was estimated in the R computer language using the Apollo package. Table 38 presents model results with generic time and cost coefficients for both truck and rail (β^m_{cost} and β^m_{time}, respectively), setting truck as the baseline (β^m_0 = 0). This gives the following utility function for truck and rail:

$$V_{ij}^{m,truck} = \beta_0^m + \beta_{cost}^m \cdot s_{ij,truck} + \beta_{time}^m \cdot t_{ij,truck} \quad (3)$$

$$V_{ij}^{m,rail} = \beta_0^m + \beta_{cost}^m \cdot s_{ij,rail} + \beta_{time}^m \cdot t_{ij,rail} \quad (4)$$

Results show a disutility with increased travel times and costs, and trucks are preferred in general for every commodity. Most coefficients have significance levels of 0.05, except those for the commodity groups Agriculture, Forestry, Fishing, and Hunting; Primary Metal Manufacturing; and Miscellaneous Manufacturing. This may be due to one or more commodities within the category showing a pattern distinct from the rest.

Table 38. Freight model choice parameter estimates using FAF5 data

Sector	Parameters	Estimate	Std. Err.	t-stat	P-value
(1) Agriculture, Forestry, Fishing, and Hunting	β^1_{rail}	−5.085	6.442	−0.789	0.430
	β^1_{time}	−2.390	3.465	−0.690	0.491
	β^1_{cost}	−0.827	1.449	−0.571	0.568
(2) Mining and Construction	β^2_{rail}	−0.693	0.079	−8.766	0.000
	β^2_{time}	−0.283	0.028	−10.186	0.000
	β^2_{cost}	−0.084	0.029	−2.894	0.004
(4) Food, Beverage, and Tobacco Product Manufacturing	β^4_{rail}	−1.336	0.050	−26.462	0.000
	β^4_{time}	−0.270	0.019	−14.026	0.000
	β^4_{cost}	−0.143	0.015	−9.264	0.000
(5) Petroleum and Coal Product Manufacturing	β^5_{rail}	−0.780	0.123	−6.357	0.000
	β^5_{time}	−0.379	0.042	−8.960	0.000
	β^5_{cost}	−0.277	0.047	−5.867	0.000
(6) Chemicals, Plastics, and Rubber Product Manufacturing	β^6_{rail}	−0.774	0.049	−15.660	0.000
	β^6_{time}	−0.348	0.019	−18.116	0.000
	β^6_{cost}	−0.043	0.016	−2.715	0.007
(7) Primary Metal Manufacturing	β^7_{rail}	−1.510	0.096	−15.722	0.000
	β^7_{time}	−0.030	0.036	−0.832	0.405
	β^7_{cost}	−0.189	0.031	−6.036	0.000

Sector	Parameters	Estimate	Std. Err.	t-stat	P-value
(8) Fabricated Metal Manufacturing	β^{8}_{rail}	−1.798	0.107	−16.750	0.000
	β^{8}_{time}	−0.088	0.042	−2.089	0.037
	β^{8}_{cost}	−0.186	0.033	−5.694	0.000
(9) Machinery Manufacturing	β^{9}_{rail}	−1.770	0.109	−16.214	0.000
	β^{9}_{time}	−0.166	0.044	−3.794	0.000
	β^{9}_{cost}	−0.199	0.032	−6.207	0.000
(10) Computer, Electronic Product, and Electrical Equipment Manufacturing	β^{10}_{rail}	−2.120	0.097	−21.928	0.000
	β^{10}_{time}	−0.133	0.038	−3.477	0.001
	β^{10}_{cost}	−0.228	0.028	−8.229	0.000
(11) Transportation Equipment Manufacturing	β^{11}_{rail}	−1.637	0.086	−19.117	0.000
	β^{11}_{time}	−0.131	0.033	−3.949	0.000
	β^{11}_{cost}	−0.136	0.026	−5.182	0.000
(12) Other Durable and Non-Durable Manufacturing	β^{12}_{rail}	−1.806	0.044	−40.837	0.000
	β^{12}_{time}	−0.114	0.017	−6.694	0.000
	β^{12}_{cost}	−0.250	0.013	−18.782	0.000
(13) Miscellaneous Manufacturing	β^{13}_{rail}	−5.002	0.866	−5.779	0.000
	β^{13}_{time}	−0.039	0.315	−0.122	0.903
	β^{13}_{cost}	−0.858	0.233	−3.685	0.000

6.1.3. Freight Origin Choice Model

The freight origin choice model is also a key component in the RUBMRIO model, as it distributes freight flow across different potential origins. The freight origin choice model uses similar freight flow records as the mode choice model estimation but aggregates records by mode. Two variables are used in the origin choice model: the logsum estimated from the mode choice model, and the population of the origin FAF zone with a log function transformation. The utility function to transport commodity m from i to j is as follows:

$$V_{ij}^{m} = \gamma^{m}\log(pop) + \delta^{m}\ln\left(\sum_{d\in D}\exp\left(V_{ij}^{md}\right)\right) \quad (5)$$

where γ^{m} and λ^{m} are parameters to be estimated for log of population and mode choice logsum, respectively. The model was also estimated using the Apollo package in R, with iterative coding for 132 different FAF zone as origins, while excluding any origins that are unavailable. “Unavailable” here indicates FAF OD pairs that do not have freight flow between them. Table 39 shows the model estimates.

Table 39. Freight origin choice model estimates

Sector	Parameters	Estimate	Std. Err.	t-stat
(1) Agriculture, Forestry, Fishing, and Hunting	γ^1	0.260	0.055	4.7
	δ^1	0.347	0.017	20.8
(2) Mining and Construction	γ^2	0.283	0.021	13.6
	δ^2	2.517	0.039	64.1
(4) Food, Beverage, and Tobacco Product Manufacturing	γ^4	0.502	0.011	47.7
	δ^4	1.858	0.016	116.3
(5) Petroleum and Coal Product Manufacturing	γ^5	0.322	0.030	10.8
	δ^5	2.457	0.046	53.7
(6) Chemicals, Plastics, and Rubber Product Manufacturing	γ^6	0.533	0.011	48.5
	δ^6	1.842	0.018	104.0
(7) Primary Metal Manufacturing	γ^7	0.440	0.022	20.4
	δ^7	2.452	0.056	44.0
(8) Fabricated Metal Manufacturing	γ^8	0.531	0.020	26.2
	δ^8	1.818	0.042	43.1
(9) Machinery Manufacturing	γ^9	0.529	0.020	27.1
	δ^9	1.329	0.031	42.4
(10) Computer, Electronic Product, and Electrical Equipment Manufacturing	γ^{10}	0.748	0.016	47.7
	δ^{10}	1.121	0.024	47.0
(11) Transportation Equipment Manufacturing	γ^{11}	0.582	0.016	37.3
	δ^{11}	2.023	0.034	59.8
(12) Other Durable and Non-Durable Manufacturing	γ^{12}	0.605	0.009	69.8
	δ^{12}	1.668	0.014	118.0
(13) Miscellaneous Manufacturing	γ^{13}	0.624	0.015	40.8
	δ^{13}	0.737	0.011	68.7

6.1.4. Economic Interaction Data

The model's embedded IO matrices' technical coefficients and regional purchase coefficients (RPCs) were obtained through IMPLAN's transaction tables, as derived from US inter-industry accounts. Technical coefficients reflect production technology or opportunities (i.e., how dollars of input in one industry sector are used to create dollars of product in another sector) and are core parameters in any IO model. RPCs represent the share of local demand that is supplied by domestic producers. RPC values across US counties are assumed to be constant here, since variations are unknown. However, counties closer to international borders are more likely to "leak" sales (as exports) than those inland, everything else constant. And production processes or technologies can vary across counties (and within industries, across specific manufacturers and product types, of course). This application assumes that all US counties have access to the same production technologies or technical coefficients table. IMPLAN's 440-sector transaction table was also split into 20 industry sectors, plus Household and Government sectors, to represent the US economy in this trade-modeling exercise.

6.2. Random-Utility-Based Multiregional Input-Output Model Specifications

This section introduces the different components of the RUBMRIO model, including the disutility function, production function, and trade flows. Proof is also given of the existence and the uniqueness of the RUBMRIO variant model.

6.2.1. Disutility Function

In the RUBMRIO model, both internal trade flows and external trade flows (from counties to export zones/customs districts) are based on the disutility of acquiring some commodity m from origin zone i and consuming it in zone j, shown in equation (6) (or exporting it to zone k, shown in equation (7)).

$$V_{ij}^{m} = -p_i^m + \gamma^m \log(pop) + \lambda^m \ln\left(\sum_{d \in D} \exp\left(V_{ij}^{md}\right)\right) \quad (6)$$

$$V_{ik}^{m} = -p_i^m + \gamma^m \log(pop) + \lambda^m \ln\left(\sum_{d \in D} \exp\left(V_{ik}^{md}\right)\right) \quad (7)$$

where $D = \{Truck, Rail\}$, with $V_{ij}^{m,truck}$ and $V_{ij}^{m,rail}$ defined in equations (3) and (4), p_i^m is the price of purchasing \$1 of commodity m in zone i (in units of utility), and γ^m and λ^m are estimated parameters based on Ben-Akiva and Lerman (1985) of input origin and shipping-mode choice by zone and sector.

6.2.2. Production Function

The behaviors of land and transport markets are highly affected by the components' market prices, including land rents and transport costs, which in turn affect production, consumption, and location decisions. The cost of producing one unit of commodity n in zone i is a function of the cost of inputs from other firms at other locations and the corresponding transport costs. The form of the overall manufacturing cost and ultimate sales price for a unit is shown in equation (8).

$$p_j^n = \sum_m \left(a_{0j}^{mn} \cdot c_j^m\right) \ \forall j, n \quad (8)$$

where a_{0j}^{mn} is the technical coefficient for zone j, which defines the fractional amount of commodity m required to produce one unit of commodity n in zone j, and c_j^m is the weighted-average cost of input m in zone j. These technical coefficients, a_{0j}^{mn}, come from the original IMPLAN transactions tables (Minnesota IMPLAN Group, 1997) for total purchases, both local and imported. IMPLAN is a social accounting and impact analysis software, developed by the Minnesota IMPLAN Group. The input costs, c_j^m, are a weighted average of input purchase price p_i^m for commodity *m* from all input zones *i* plus the associated generalized transport costs b_{ij}^{md} (from each zone *i* to zone *j* using mode *d*), as shown in equations (9) and (10). The weight factors are the interzonal trade flows by mode (X_{ij}^{md}).

$$B_{ij}^{md} = p_i^m + b_{ij}^{md} \ (9)$$

$$c_j^m = \frac{\sum_i \sum_d \left(X_{ij}^{md} \cdot B_{ij}^{md} \right)}{\sum_i \sum_d X_{ij}^{md}} \ (10)$$

6.2.3. Trade Flows

Trade flows can be calculated when all the other values are given, including export demands, production costs, technical coefficients, and transport costs. Under an assumption of profit-maximizing and cost-minimizing behavior, with unobserved heterogeneity in alternatives, consumers (both final and intermediate) will buy from the producer that can supply the lowest total price (including transport costs) of any input. Unobserved heterogeneity introduces the random element, which, under an assumption of IID Gumbel distribution, leads to the multinomial logit model for origin and mode choices. Two kinds of trade flow are estimated in the current RUBMRIO model; these are the interzonal trade flows by mode, X_{ij}^{md}, and the flows to export zones by mode, Y_{ik}^{md}, as shown here:

$$X_{ij}^{md} = C_j^m \frac{\exp\left(V_{ij}^m\right)}{\sum_i \exp\left(V_{ij}^m\right)} \frac{\exp\left(V_{ij}^{md}\right)}{\sum_d \exp\left(V_{ij}^{md}\right)} \ \forall i, j, m, d \ (11)$$

$$Y_{ik}^{md} = Y_k^m \frac{\exp(V_{ik}^m)}{\sum_i \exp(V_{ik}^m)} \frac{\exp\left(V_{ik}^{md}\right)}{\sum_d \exp\left(V_{ik}^{md}\right)} \ \forall i, k, m, d \ (12)$$

where C_j^m is the total volume of *m* consumed in zone *j*, which can be calculated based on equation (13):

$$C_j^m = \sum_n \left(a_j^{mn} \cdot x_j^n \right) \ \forall j, m \ (13)$$

Here, a_j^{mn} is the technical coefficient matrix (following leakage considerations) for zone *j*, which defines the amount of commodity *m* required (from within the state) to produce one unit of commodity *n* in zone *j*. And x_i^m is the total production of commodity *n* in zone *i*, which is the sum of the trade flows leaving zone *i* to meet the demands of other producers and export zones.

$$x_i^m = \sum_j \sum_d X_{ij}^{md} + \sum_k \sum_d Y_{ik}^{md} \ \ \forall i, m \ (14)$$

Equations (6) through (14) constitute the majority of the RUBMRIO model; these equations are solved iteratively to achieve an equilibrium trade pattern. To resolve this set of equations (and achieve a convergent solution), the iterations begin by setting all prices to zero, solving for trade-flow probabilities, and generating an initial pattern of trade. This alters the price structure, and thus the trade pattern. We continue updating prices and patterns until convergence. Zhao and Kockelman (2002) describes this process.

6.2.4. Solution Existence and Uniqueness

This section presents the fixed-point RUBMRIO variant problem that incorporates the cost of modes in the average cost calculations for commodities, compared to the general form in Zhao and Kockelman (2004), which uses an average cost term as logsum.

Define P_{ij}^m as the probability that region *j* purchases input *m* from region *i* and P_{ij}^{md} as the probability of choosing mode *d* for transport given that region *j* purchases input *m* from region *i*:

$$P_{ij}^m = \frac{\exp\left(V_{ij}^m\right)}{\sum_i \exp\left(V_{ij}^m\right)} \ (15)$$

$$P_{ij}^{md} = \frac{\exp\left(V_{ij}^{md}\right)}{\sum_d \exp\left(V_{ij}^{md}\right)} \ (16)$$

Then we can reformulate the average cost:

$$p_j^n = \sum_m a_{0j}^{mn} \cdot c_j^m = \sum_m a_{0j}^{mn} \cdot c_j^m = \sum_m a_{0j}^{mn} \cdot \frac{\sum_i \sum_d \left(X_{ij}^{md} \cdot B_{ij}^{md}\right)}{\sum_i \sum_d X_{ij}^{md}}$$

$$= \sum_m a_{0j}^{mn} \cdot \frac{\sum_i \sum_d \left[X_{ij}^{md} \cdot \left(p_i^m + b_{ij}^{md}\right)\right]}{\sum_i \sum_d X_{ij}^{md}}$$

$$= \sum_m a_{0j}^{mn} \cdot \frac{\sum_i \sum_d \left[C_j^m \frac{\exp\left(V_{ij}^m\right)}{\sum_i \exp\left(V_{ij}^m\right)} \frac{\exp\left(V_{ij}^{md}\right)}{\sum_d \exp\left(V_{ij}^{md}\right)} \cdot \left(p_i^m + b_{ij}^{md}\right) \right]}{\sum_i \sum_d X_{ij}^{md}}$$

$$= \sum_m a_{0j}^{mn} \cdot \frac{C_j^m \sum_i \sum_d \left[\frac{\exp\left(V_{ij}^m\right)}{\sum_i \exp\left(V_{ij}^m\right)} \frac{\exp\left(V_{ij}^{md}\right)}{\sum_d \exp\left(V_{ij}^{md}\right)} \cdot \left(p_i^m + b_{ij}^{md}\right) \right]}{C_j^m}$$

$$= \sum_m a_{0j}^{mn} \cdot \sum_i \sum_d \left[\frac{\exp\left(V_{ij}^m\right)}{\sum_i \exp\left(V_{ij}^m\right)} \frac{\exp\left(V_{ij}^{md}\right)}{\sum_d \exp\left(V_{ij}^{md}\right)} \cdot \left(p_i^m + b_{ij}^{md}\right) \right]$$

$$= \sum_m a_{0j}^{mn} \cdot \sum_i \sum_d \left[P_{ij}^m P_{ij}^{md} \cdot \left(p_i^m + b_{ij}^{md}\right) \right]$$

We then denote:

$$\vec{p} = \{p_j^n\}$$

Therefore,

$$f_j^n(\vec{p}) = \sum_m a_{0j}^{mn} \cdot \sum_i \sum_d \left[P_{ij}^m(\vec{p}) P_{ij}^{md} \cdot \left(p_i^m + b_{ij}^{md}\right) \right]$$

$$= \sum_m a_{0j}^{mn} \cdot \sum_i P_{ij}^m(\vec{p}) \cdot \left[p_i^m + \sum_d \left(P_{ij}^{md} \cdot b_{ij}^{md}\right) \right]$$

Therefore, we have a fixed-point problem as follows:

$$\vec{p} = \vec{f}(\vec{p}) \qquad (17)$$

Compared to Zhao and Kockelman (2004), this fixed-point problem variant replaces the generic transportation price (regardless of mode) with the probability-weighted transportation cost for different modes. The proof of existence and uniqueness of the solution to this fixed-point model follows Zhao and Kockelman (2004):

- Existence condition for the price solution

  First, we impose a rather weak condition on the feasible set to ensure the existence of a solution. Let $K_p = \{p_{ij}^n | 0 \le p_{ij}^n \le p_{ij}^{n*}, \forall i, j, n\}$, where $\{p_{ij}^{n*}\}$ are upper bounds that we assume can be determined a priori (in practice, one can usually choose very large numbers as upper bounds). Then K_p is a bounded and closed convex subset (therefore, a compact set) on the space R^{MJJ}. We can easily observe that if the prices are bounded, the function $\vec{f}$ also can be considered bounded, since it is a convex combination of prices (plus transportation costs) across space (i.e., $\sum_i P_{ij}^m = 1, \forall m$) and economic sectors (i.e., $\sum_m a_j^{mn} \leq 1, \forall n, j$). If one assumes that $\vec{f}$'s upper bounds are also $\{p_{ij}^{n*}\}$, one essentially assumes that the upper bounds are large enough to accommodate the transportation prices' contributions to $\vec{f}$. Then, $\vec{f}$ is a mapping $K_p \rightarrow K_p$, and it is continuous. Following Brouwer's theorem (see Khamsi and Kirk, 2001), we then have the following condition:

The fixed-point problem (17) provides at least one solution if and only if there exist positive constants $\{p_{ij}^{n*}\}$ such that the fixed-point problem (17) provides at least one feasible solution in the space K_b.

- Uniqueness condition for the price solution

  Sufficient conditions for the uniqueness of the solution of a fixed-point problem are given by Banach's theorem (see Border, 1985), which requires that the function be contractive over a complete set or quasi-contractive (implying monotonicity) over a compact set. We consider that K_p is in a normed space, due to the mean-value theorem (see Khamsi and Kirk, 2001), if $\|\nabla \vec{f}(\vec{p})\| < 1$; then the fixed-point problem has a unique solution, and the sequence $\vec{p}^{(t+1)} = \vec{f}(\vec{p}^{(t)})$ converges on the unique solution $\vec{p} = \vec{f}(\vec{p})$, if $\vec{p}^{(o)} \in K_p$.

Now consider the general case of a dispersion parameter λ^m for the origin choice model:

$$P_{ij}^m = \frac{\exp(\lambda^m V_{ij}^m)}{\sum_i \exp(\lambda^m V_{ij}^m)} \quad (18)$$

The same process in Zhao and Kockelman (2004) can be followed, when the probabilities are determined by relative disutilities, which depend on prices:

$$\frac{\partial f_j^n(\vec{p})}{\partial p_i^m} = \frac{\partial}{\partial p_i^m}\left[\sum_m a_j^{mn} \sum_k P_{kj}^m(\vec{p}) \cdot \left(p_k^m + p_k^m + \sum_d \left(P_{ij}^{md} \cdot b_{ij}^{md}\right)\right)\right]$$

$$= a_j^{mn} \frac{\partial}{\partial p_i^m}\left[\sum_k P_{kj}^m(\vec{p}) \cdot \left(p_k^m + \sum_d \left(P_{ij}^{md} \cdot b_{ij}^{md}\right)\right)\right]$$

$$= a_j^{mn} \cdot P_{ij}^m \cdot \left\{1 - \lambda^m \left[\left(p_i^m + \sum_d \left(P_{ij}^{md} \cdot b_{ij}^{md}\right)\right) - c_j^m\right]\right\} (19)$$

Letting $d_{ij}^d = \sum_d \left(P_{ij}^{md} \cdot b_{ij}^{md}\right)$, equation (19) can be written as:

$$\frac{\partial f_j^n(\vec{p})}{\partial p_i^m} = a_j^{mn} \cdot P_{ij}^m \cdot \left\{1 - \lambda^m \left[p_i^m + d_{ij}^d - c_j^m\right]\right\}$$

which is the same equation as (3.14) in Zhao and Kockelman (2004). This proof then merges with the proof in Zhao and Kockelman (2004) (equation 3.14 forward) to show that $\left\|\nabla \vec{f}(\vec{p})\right\| < 1$, and we reach the following restrictive uniqueness condition for price solution:

The fixed-point problem (17) results in at most one equilibrium price solution if the dispersion parameters $\{\lambda^m\}$ are sufficiently small such that the inequality $\lambda^m < 1/ \max_{1 \le i,j \le J} \left(p_i^m + d_{ij}^m - c_j^m\right)$ $\forall m$ holds.

6.3. Scenario Experiment and Analysis

The base case scenario for the model was set based on the US's export flow from 2017, when only HTrucks were available. In the ATrucks scenario, they were added as an additional mode nested within the truck mode. The alternative specific constants (ASCs) for Atrucks are set as −0.1 to recognize the initial high cost and the gradual adoption and preference for Atrucks. The operating cost of ATrucks is taken to be 50% of that of HTrucks, based on the assumption that no drivers are needed in the vehicles. More sensitivity analysis will be conducted to also create a model where ATrucks have drivers to react to emergencies. Figure 51 shows the nesting structure of the mode and origin choices. The new equations are from Huang and Kockelman (2020) and the parameters are shown in Table 40.

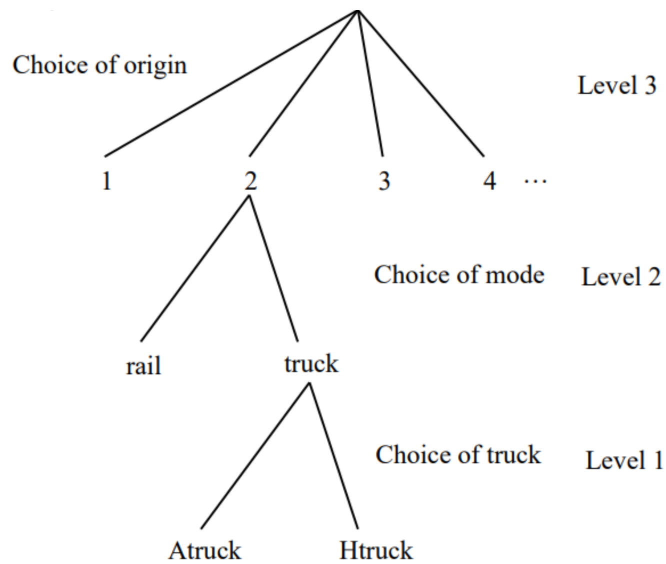


Table 40. Parameter estimates for origin, mode, and truck choice equations

Sector	Origin Choice Parameters $\theta^m_{ij}=1$		Mode Choice Parameters $\theta^m_{ij,mode}=1/1.2$			Truck Choice Parameters $\theta^m_{ij,truck}=1/1.4$		
	γ^m	λ^m	$\beta^m_{1,\ rail}$	$\beta^m_{rail,time}$	$\beta^m_{rail,cost}$	$\beta^m_{1,\ ATruck}$	$\beta^m_{truck,time}$	$\beta^m_{truck,cost}$
1	0.26	0.35	−5.09	−2.39	−0.83	−0.10	−2.79	−0.96
2	0.28	2.52	−0.69	−0.28	−0.08	−0.10	−0.33	−0.10
4	0.28	2.52	−0.69	−0.28	−0.08	−0.10	−0.33	−0.10
5	0.50	1.86	−1.34	−0.27	−0.14	−0.10	−0.32	−0.17
6	0.32	2.46	−0.78	−0.38	−0.28	−0.10	−0.44	−0.32
7	0.53	1.84	−0.77	−0.35	−0.04	−0.10	−0.41	−0.05
8	0.44	2.45	−1.51	−0.03	−0.19	−0.10	−0.04	−0.22
9	0.53	1.82	−1.80	−0.09	−0.19	−0.10	−0.10	−0.22
10	0.53	1.33	−1.77	−0.17	−0.20	−0.10	−0.19	−0.23
11	0.75	1.12	−2.12	−0.13	−0.23	−0.10	−0.16	−0.27
12	0.58	2.02	−1.64	−0.13	−0.14	−0.10	−0.15	−0.16
13	0.61	1.67	−1.81	−0.11	−0.25	−0.10	−0.13	−0.29

The base case scenario without ATrucks shows a total \$1.06 trillion export demand and \$11.1 trillion domestic demand, with trucks dominating the market, generally moving 93% of product value while rail moves the other 7%. If measuring by transported ton-mile, truck accounts for 83% and rail 17%. Since the model was driven by export demand, it remains the same in both the base case and the ATruck scenarios, and the domestic flow is also very similar across scenarios (dropping slightly to \$10.9 trillion in the ATruck scenario).

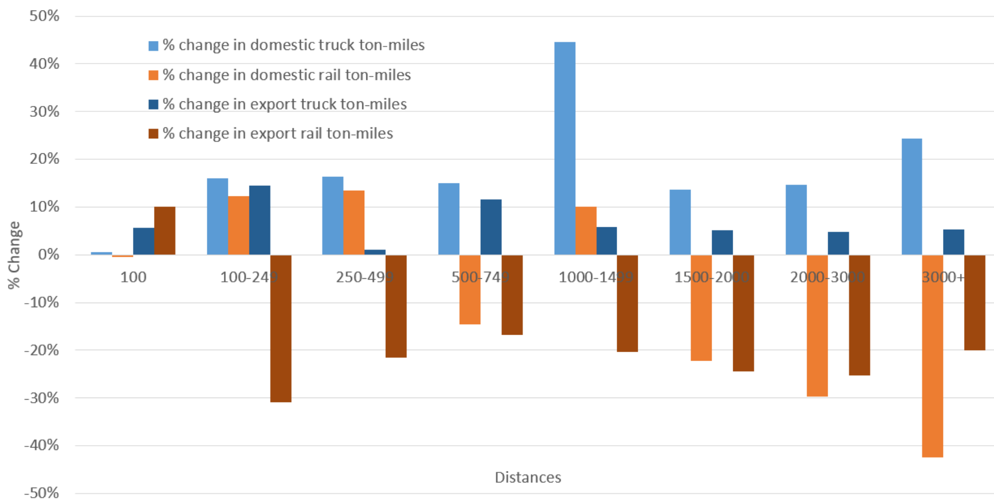


Figure 52. Change in domestic and export ton-miles by mode after the introduction of ATrucks

Figure 52 shows the percentage change in truck (sum of ATruck and HTruck) and rail mode choice for domestic and export flow in ton-miles after the introduction of ATrucks. Their introduction causes an increase in domestic and export truck ton-miles for trips of all distances, and especially those between 1,000 and 1,500 miles. Export rail ton-miles shift to trucks, but domestic rail's ton-miles still increase for trips between 1,000 and 1,500 miles, or shorter than 500 miles.

Figure 53 shows the mode split (in terms of transported ton-miles) between HTrucks, ATrucks, and rail in the ATruck scenario. HTrucks' mode share rises as the distance increases but drops after 500 miles, while rail share is fairly stable across all distances.

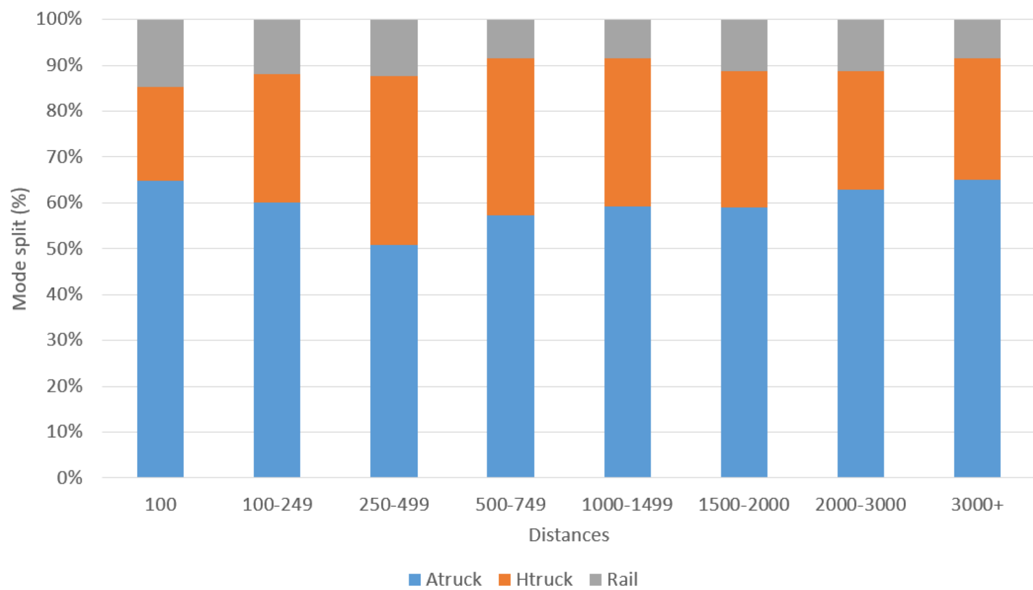


Figure 53. Mode share (in ton-miles) with introduction of ATrucks

As mentioned, a major benefit of ATrucks is that unlike HTrucks, whose drivers need to rest nightly, they can keep driving overnight. In the mode choice model estimates, an extra 14 hours of non-driving time is assumed for every 10 hours of an HTruck's on-road travel time. Figure 54 provides a comparison of two ATruck scenarios, one where ATrucks reduce cost by 50% and also eliminate overnight resting time, and another where ATrucks only reduce the cost without realizing any time savings. Figure 54 demonstrates that if the travel time saved by ATrucks is taken into consideration, they attract 25% more value and ton-miles, especially for trips between 500 and 750 miles. This increase drops off when the distances become longer, only comparatively raising ATrucks' ton-miles by about 10% and their value transported by 5% for trips over 1,500 miles.

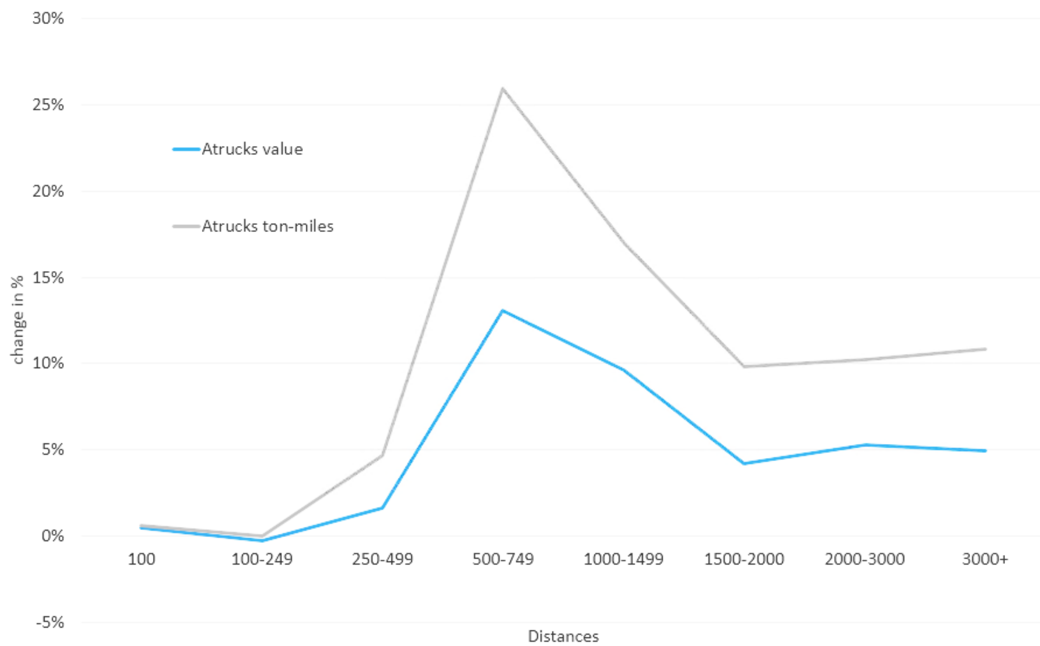
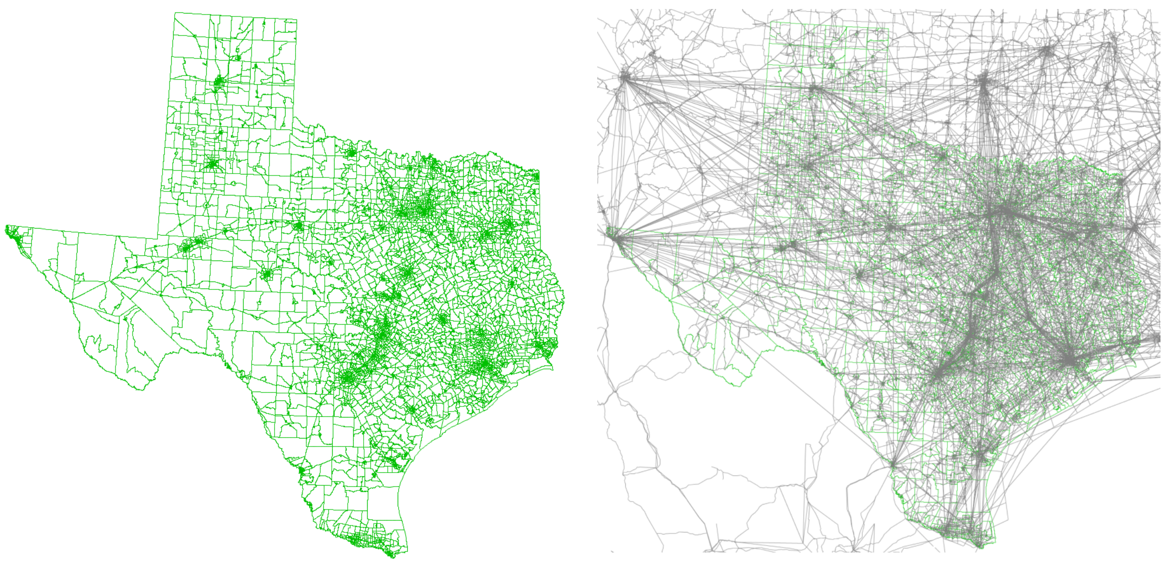


Figure 54. Change in value and ton-miles transported when ATrucks' overnight time savings and cost savings are taken into account (compared to cost savings only)

Texas Freight Demand Model Application

In this chapter, we leverage TxDOT's SAM-V4 to conduct freight travel predictions for 2045 accounting for the use of ATrucks. Both ATrucks and automated passenger cars are modeled to reflect the future use of automation technology, with the analysis emphasis on freight patterns.

The SAM network covers all of North America, with greater detail in and near Texas. Figure 55b shows the state's highway, railway, and airline networks, which contain 200,445 links and 168,507 nodes. There are 6,860 traffic analysis zones (TAZs) in SAM-V4, significantly more detail compared to the 4,400 TAZs in SAM-V3. As illustrated, dense networks exist within Houston, San Antonio, Austin, and the Dallas-Fort Worth metroplex.



a. SAM-V4 TAZs *b. SAM-V4 Networks*

Figure 55. Geographic data used in SAM model

7.1. Travel Demand Model Methods

A four-step travel demand modeling process with a feedback loop is used here to model traffic patterns across the entire state of Texas, including trip generation, trip distribution, mode choice, and traffic assignment. For passenger travel's four-step model, the traditional trip distribution table was obtained from the SAM-V4 model, and the production-attraction matrix was then converted into an OD matrix. The model uses one time of day (24 hours) as the simulation horizon, recognize that many trips, especially freight trips, are long-distance, spanning many times of day and many different congestion settings. Computation time is another concern that is addressed by using only one time-of-day simulation. For the freight model, a doubly constrained trip distribution procedure was used, based on SAM-V4's 2045 freight trip generation parameters. A mode choice model was then applied, reflecting truck, carload rail (CL), and intermodal transport (IM) alternatives. A base case scenario, without AV, SAV, and ATruck modes, was run first, to compare against the self-driving scenarios. Various parameter settings were also tested, using sensitivity analysis, to provide a sense of prediction variability.

7.2. Trip Generation

Trip generation data were obtained from the SAM-V4 2045 scenario results, based on underlying population and jobs forecasts by zone (Alliance Transportation Group, 2018), using 2016/17 NHTS data. Passenger trip types here include home-based work, home-based other, home-based school, non-home-based other, and non-home-based visitor. In terms of freight trip generations, Table 41's 15 commodity groups are based on US SCTG code. SAM-V4 freight transport attraction and production are conducted across all Texas counties and also non-Texas US states.

7.3. Trip Distribution

As previously mentioned, the passenger trip distribution table was obtained from SAM-V4 directly; however, this procedure can be replaced with a destination choice model in future work. Freight trips are distributed by ton of each commodity, using a doubly constrained gravity model, to keep values in strong alignment with current freight production and consumption levels across the state of Texas and beyond. The associated utility function is as follows:

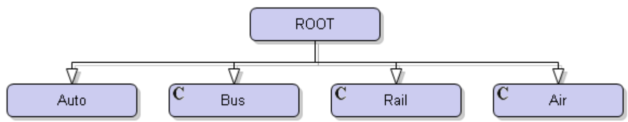
$$V_{ijc}=\exp\left(-1/\left(D_c \cdot D_{ij}\right)+\delta \cdot \ln(pop_i)+\tau \cdot \log\left(\sum_m \exp\left(V_{ij}^m\right)\right)\right)$$

where D_c is the average travel distance for commodity group c and D_{ij} is the distance from zone i to zone j.

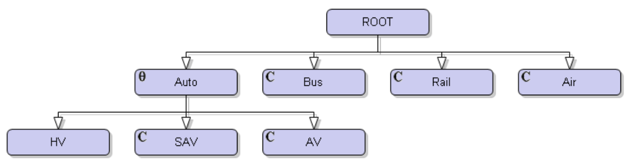
7.4. Mode Choice

Four passenger modes exist in the base case (Year 2040) scenario: conventional automobile ("HV"), bus, rail, and air. Three freight modes exist: truck, (carload) rail, and IM. These choice models were expanded to accommodate AV, SAV, and

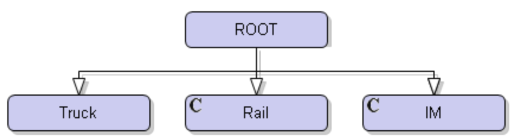
ATruck modes, as shown in Figure 56. Trip costs, fares, and in-vehicle travel times for bus, rail, and air all come from SAM-V4 outputs. Rail's values are the average of all SAM rail modes for each OD pair (including urban rail, intercity rail, and high-speed rail alternatives in many OD cases). When AVs and SAVs are added to the set of alternatives, they, along with HVs, are nested under the auto mode (Figure 56b). There is no parking cost for SAV use (much like a taxi), and privately owned AVs are assumed to face the same parking costs as HVs (since AVs are not expected to be allowed to drive empty, as this would create additional congestion for cities and regions).



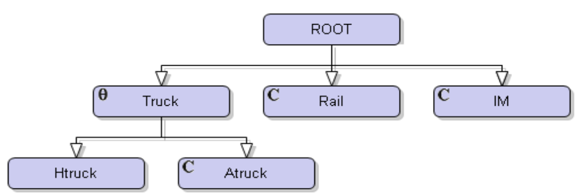
a. Passenger mode choice structure without AVs



b. Passenger mode choice structure with AVs



c. Freight mode choice structure without ATrucks



d. Freight mode choice structure with ATrucks

Figure 56. Mode choice structures for passenger and freight transport, before and with AVs

Operating costs of bus, rail, and air modes come directly from SAM-V4 outputs, while several assumptions are used for vehicle costs. Litman (2018) anticipates AV operating costs to be $0.80–1.20 per mile in the early years of AV availability, before declining to \$0.60–1.00 per mile, compared to \$0.40–\$0.60 per mile for HVs. Johnston and Walker (2017) expected SAVs to debut in some cities in 2018 at \$0.86 per mile, or \$0.84 per mile for shared autonomous electric vehicles (SAEVs). They expect traditional transportation network company vehicles (like today's Lyft and Uber rides) to cost \$2.04 per mile, and SAEV fees to fall to \$0.51 per mile in 2025, \$0.36 in 2030, and \$0.33 in 2035. Bösch et al. (2017) predict that SAVs may cost \$0.44 per mile to cover operating costs and deliver a very healthy 30% profit margin, while a dynamic ride-sharing (en route carpooling) service may cost \$0.20–0.30 per passenger mile. They also suggest that SAVs purpose-built for use as pooled taxis may lower fares to just \$0.16 per mile, long term.

Perrine et al.'s (2018) model of long-distance US travel assumed AV costs would range from \$0.10 to \$1.65 per mile and value of travel time (VOTT) would be \$3.00 to \$9.00 per hour for AV occupants, with the base case scenario of \$0.20 per mile operating cost and VOTT of \$6.00 across six distinct scenarios. Fagnant and Kockelman (2016) estimated that SAV pricing at \$1.00 per mile could generate a 19% annual return on investment if each AV's purchase price is \$70,000. This return varied from 12.3% to 38.8% for operating costs of \$0.50 and \$0.25 per mile, respectively. Arbib and Seba (2017) envision internal-combustion SAVs will cost roughly \$0.38 per mile, while SAEVs may be much cheaper, at \$0.16 per mile in 2021 and \$0.10 per mile in 2030. They posit that government subsidies or advertising may one day make SAEVs free to most or all riders.

Based on all these estimates, this work assumes that both AVs and HVs carry operating costs of \$0.60 per mile and SAVs cost either \$1.50, \$1.00, or \$0.50 per mile (across scenarios). These are combined with parameter assumptions from Zhao and Kockelman (2017) to create the mode choice parameters used here. The parameters are summarized in Table 41, with several varying later during sensitivity analyses. The ASCs for AVs and SAVs are set to be negative, at −0.05 and −0.2, respectively, to reflect some consumer hesitation. This is based on surveys and other work by Casley et al. (2013), Schoettle and Sivak (2014), and Bansal and Kockelman (2017) suggesting that AVs and SAVs will improve travelers' safety and mobility but may generate some acquisition cost, privacy, and controllability concerns (especially when the vehicle is not privately owned).

Table 41. Passenger and freight model parameters

(a) Passenger Model

	Base Case	Automobile	Bus	Rail	Air
Mode Choice	Constant	0	−2.8	−2.8	−2.8
	Operating Cost Coefficient	0.072	−0.14	−0.14	−0.14

	In-Vehicle Time Coefficient	−0.019			−0.019	−0.019	−0.019
	Operating Cost ($/mile)	0.6			N/A	N/A	N/A
	Parking Cost	✓			N/A	N/A	N/A
	VOTT	15.83			8.14	8.14	8.14
	AV Case	HV	AV	SAV	Bus	Rail	Air
	Nesting Coefficient	λ = 0.6*			N/A	N/A	N/A
	Constant	0	−0.05	−0.2	−2.8	−2.8	−2.8
	Operating Cost Coefficient	−0.072	−0.072	−0.072	−0.14	−0.14	−0.14
	In-Vehicle Time Coefficient	−0.019	−0.015*	−0.015*	−0.019	−0.019	−0.019
	Operating Cost ($/mile)	0.6	0.8*	1*	N/A	N/A	N/A
	Parking Cost	✓	✓	×	N/A	N/A	N/A
	VOTT ($/hr)	15.83	11.08*	11.08*	8.14	8.14	8.14

(b) Freight Model (Adapted from TxDOT SAM4)

Trip Distribution	Mode Choice Logsum			Log of Population	
	τ = 0.5			δ = 0.1	
Mode Choice	**Rail Constant**	**IM Constant**	**Cost Coefficient**	**Time Coefficient**	**Average Travel Distance (mi.)**
Agriculture	−1.343	−5.224	−0.018	-	1,539
Mining	−2.291	−6.111	−0.033	-	888
Coal	3.316	-	−0.007	-	1,175
Nonmetallic Minerals	−1.441	−8.469	−0.031	-	670
Food	−2.237	−6.430	−0.016	-	1,715
Consumer Manufacturing	−6.742	−4.233	−0.012	-	2,174
Non-Durable Manufacturing	−5.941	−5.345	−0.019	-	1,837
Lumber	−2.253	−6.053	−0.029	−0.021	1,437
Durable Manufacturing	2.407	−2.771	−0.008	−0.064	1,828
Paper	−1.772	−4.420	−0.013	-	1,463
Chemicals	−0.874	−6.644	−0.011	-	1,322
Petroleum	−2.529	−8.443	−0.030	-	935
Clay, Concrete, Glass	−2.668	−6.520	−0.019	-	1,414
Primary Metal	−0.609	−7.263	−0.010	-	1,661
Secondary & Misc. Mixed	−4.143	−4.457	−0.016	-	1,902

*Note: Numbers marked with * are modified during sensitivity analysis.*

As shown in Figure 56c and d, the HTruck and ATruck alternatives are nested under the truck mode after AVs are introduced to the market. The air and water modes are ignored here, since they are considered fixed in the SAM-V4. (In reality, some air-freight and water-born-freight trips will probably be replaced by ATruck trips, due to their convenience, cost, and speed.) An ATruck is assumed to cost 1.5 times as much as an HTruck per mile because of the cost of automation equipment and training for the drivers who attend the truck, but is also assumed to save some connecting (uploading or downloading) time at origins and destinations. The nesting coefficient is set to 0.7, recognizing that HTrucks and ATrucks have more relative substitutability as their costs and times are similar. Travel time and travel cost skims of the IM mode were obtained from SAM-V4 (Alliance Transportation Group, 2018).

7.5. Traffic Assignment and Feedback Loop

Mode and destination choice results are transformed into trip tables or OD matrices and round-trip tours are split in two for the final traffic assignment. Based on 2009 NHTS data (Santos et al., 2011), HV, AV, and SAV occupancies are set to 1.5 persons. The freight trip table (in tons by commodity) is converted to trucks and rail cars, based on SAM weights. Feedback loops are performed to provide consistent results between travel time and cost skims and network assignment flows, feeding congested travel times back for subsequent iterations, using the method of successive averages.

A multi-modal, multi-class assignment was conducted in each scenario to reflect large differences in VOTT between human drivers and self-driving vehicles. The feedback loop was set to perform 20 iterations, with a stopping criterion of a relative gap below 10^{-4}, to try and achieve a stable, convergent equilibrium.

7.6. Model Calibration

To appreciate how parameter and model specification changes affect predictions, the revised model's results (for the before-ATruck base case) were compared to the original SAM-V4 outputs, with histograms of trip distances shown in Figure 57. The correlation is 0.899 at 1-mile distance bins for trip counts between every OD pair. Truck and CL volumes exhibit relatively high correlations in each of the 15 commodity classes, while IM results (for intermodal assignments) are relatively uncorrelated. Fortunately, the IM mode accounts for a relatively small amount of Texas trade for most commodities.

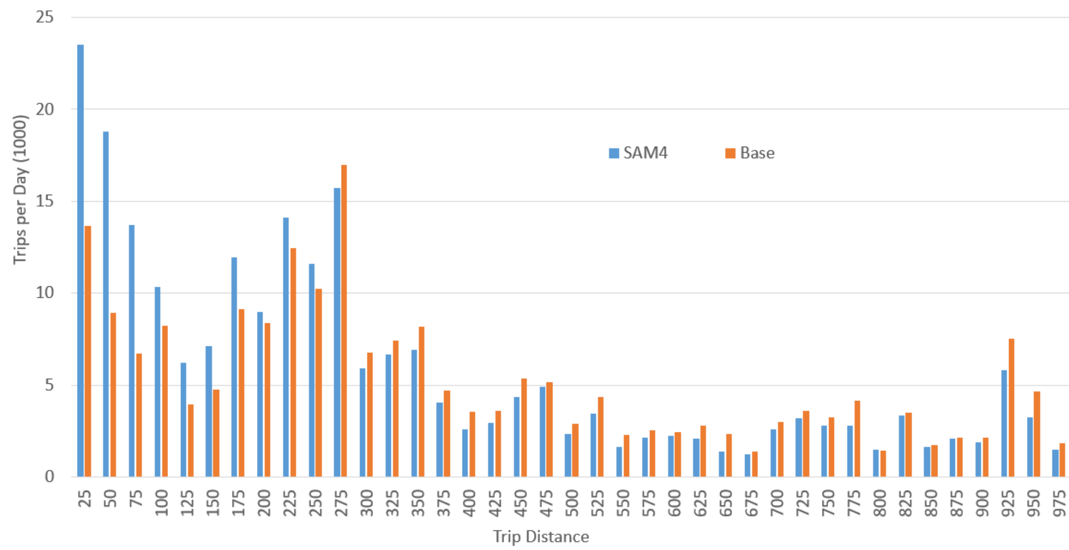
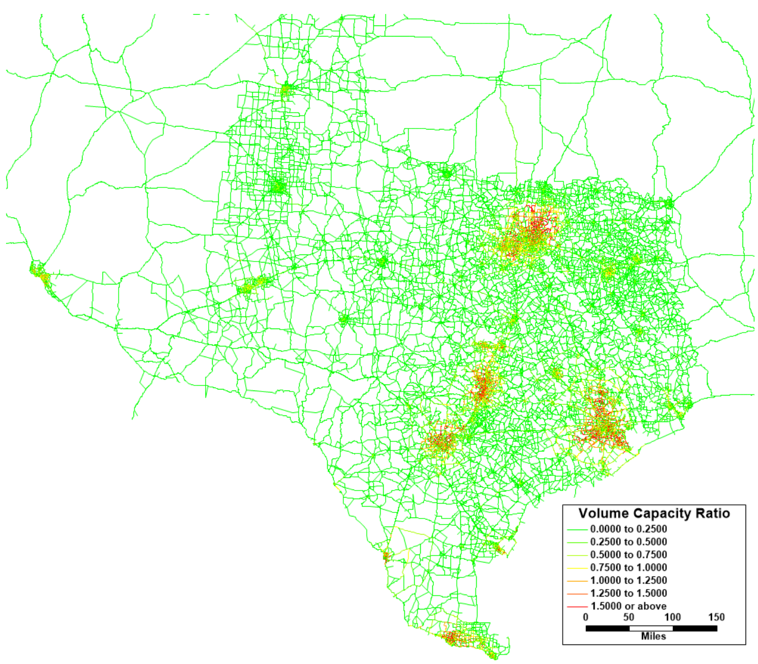


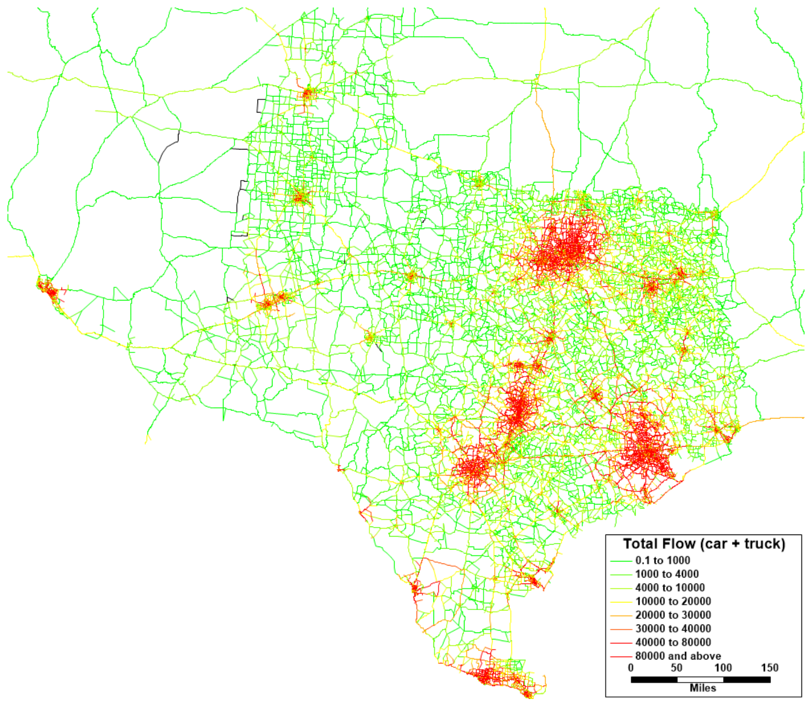
Figure 57. Comparing predicted trip distance to SAM model results

7.7. Results

The following discussion looks at mode split shifts before and after AVs and ATrucks are introduced in passenger and freight transport markets across Texas. Trip length distributions and travel patterns across zone pairs are examined, and the network congestion is also presented.



a. Volume capacity ratio in Texas



b. Link flow in Texas

Figure 58. Network performance

Figure 58a shows the volume capacity ratio across the Texas region. Congestion (v/c > 1.25) occurs mostly among the major cities across Texas, especially Dallas-Fort Worth, Houston, San Antonio, and Austin. Congestion is also observed in a few other cities like El Paso, Corpus Christi, and Amarillo. In the visualization of the total flow in Figure 58b, the sample patterns merge.

Using the SAM-V4 mode choice specification, mode share for freight by industry sector can be obtained (Table 42). 1.7 billion tons of goods (60% of total goods) are transported by HTrucks and ATrucks, while the rest are transported by CL (23.5%) and IM (16.5%). All of the 15 industries modeled would witness more truck trips and fewer CL and IM trips after ATrucks are introduced. This happened even for the commodities that are oriented toward CL or IM. Across all commodities transported in ton-miles, truck transportation saw a 7.8% increase, while CL and IM dropped by 12.6% and 2.3%, respectively. The increase in truck trips varies by mode but most transportation by CL and IM decreases. Coal commodity truck ton-miles see a massive increase (140.2%), mainly shifted from CL, which dominates coal transportation prior to ATruck implementation.

Figure 59a presents the trip distribution of trucks before and after ATrucks become available. Since the freight mode choice is doubly constrained, truck share after AV introduction shows the same trend as the "before" scenario. Truck trips increase slightly for all trip distances, with the conventional HTruck gaining a greater share of ton-miles at all distances than the ATruck in this megaregion, since the ATruck costs more, especially for these intermediate (under 5 hour) travel times. In the future, as the cost of ATrucks decreases, their market share would be expected to increase. The jump at 230 miles can be seen as the distance between Dallas-Fort Worth and Houston, San Antonio and Houston, or Austin and Houston. It is evident that Houston is the main freight center in the megaregion.

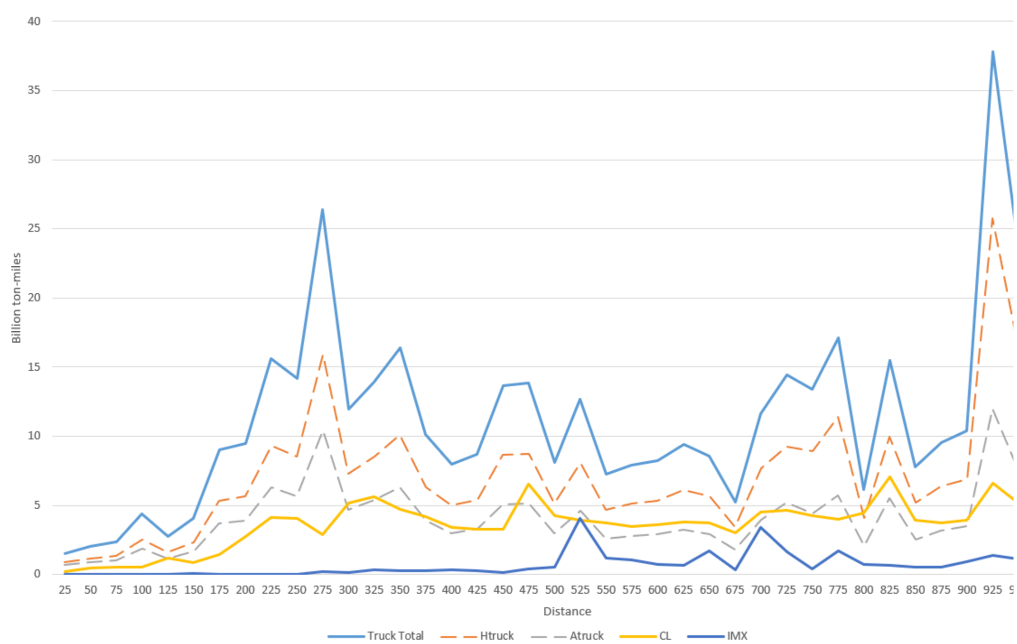
Figure 59b shows the percentage change in ton-miles transported by each mode after ATruck introduction. The increase in truck ton-miles is quite stable, with a roughly 5% increase across all distances. The decreasing trend for IM is stable as well, at around 3%, except for a greater drop at distances longer than 800 miles. Interestingly, CL mode choice decreases the most for short distances, due to a shift to truck mode, but this decrease diminishes at longer transportation distances.

The ATruck base case assumes that ATruck cost is 1.5 times (on a per-mile basis) HTruck cost, due to the high initial cost of automation technology (e.g., cameras and radars). However, ATruck costs may vary in the future due to different stages of automation implementation and advances in automation technology. Thus two additional scenarios were tested, one assuming that ATruck cost is 1.25 times HTruck cost, and another in which they cost the same. Figure 60 shows the change in mode share (measured in ton-miles) by trip distance when ATruck cost drops from 1.5 times HTruck cost to its equal. ATruck mode share increases across all distances when its cost is lower, with the most dramatic effect on longer trip distances, reaching a 50% increase in mode share at about a thousand miles. Accordingly, the other modes' shares decrease across all distances, especially HTruck's, which drops 20% at a thousand miles.

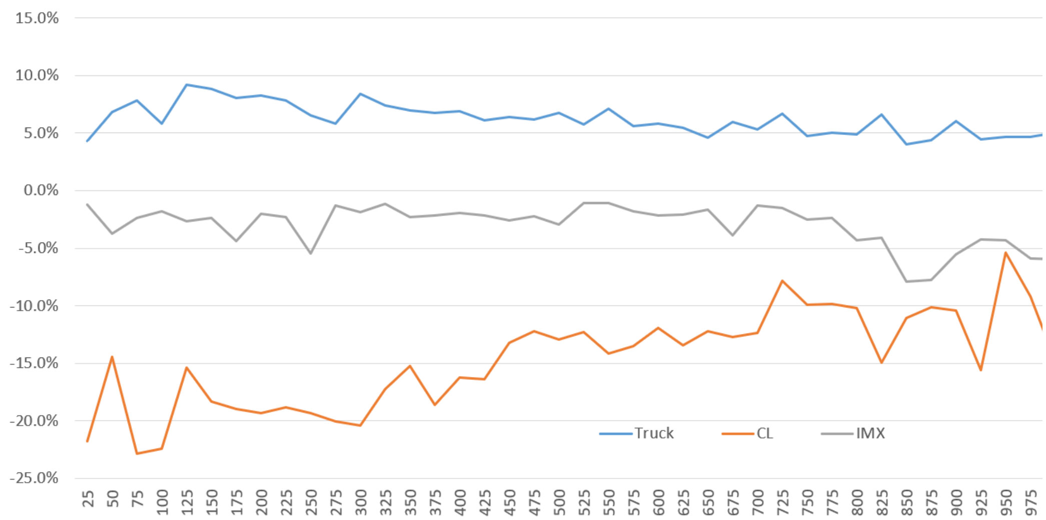
Table 42. Mode shares of freight ton-miles moved within Texas

Commodity	Billion Ton-Miles Transported after ATrucks Introduced					Mode Share after ATrucks Introduced					Change in Tons Transported from Base Case		
	HTruck	ATruck	Truck	Rail	IM	HTruck	ATruck	Truck	Rail	IM	Truck	Rail	IM
Agriculture	0.1	0.2	0.2	119.3	0.3	0.1%	0.1%	0.2%	99.6%	0.2%	33.8%	−0.1%	−0.1%
Mining	22.6	6.5	29.1	15.5	0.6	50.0%	20.1%	64.5%	34.2%	1.3%	77.4%	−46.3%	−47.2%
Coal	9.8	4.2	14.0	63.1	8.1	11.5%	5.4%	16.5%	74.1%	9.5%	140.2%	−9.4%	−9.3%
Nonmetallic Minerals	39.0	73.6	112.6	87.6	0.2	19.4%	39.5%	56.2%	43.7%	0.1%	14.2%	−15.6%	−21.0%
Food	25.3	68.2	93.5	10.0	17.2	21.0%	58.5%	77.5%	8.3%	14.3%	4.6%	−15.7%	−15.9%

Commodity	Billion Ton-Miles Transported after ATrucks Introduced					Mode Share after ATrucks Introduced					Change in Tons Transported from Base Case		
	HTruck	ATruck	Truck	Rail	IM	HTruck	ATruck	Truck	Rail	IM	Truck	Rail	IM
Consumer Manufacturing	2.1	0.7	2.8	0.1	2.2	40.8%	18.7%	55.2%	1.4%	43.4%	68.9%	−41.9%	−37.7%
Non-Durable Manufacturing	4.3	13.4	17.7	1.7	0.1	22.2%	69.8%	91.0%	8.5%	0.5%	1.6%	−14.1%	−15.2%
Lumber	21.3	17.5	38.8	0.2	0.0	54.7%	45.0%	99.4%	0.5%	0.0%	0.7%	−47.4%	−46.3%
Durable Manufacturing	34.1	56.8	91.0	5.6	0.2	35.3%	60.1%	94.0%	5.8%	0.2%	2.5%	−29.3%	−29.5%
Paper	1.0	3.3	4.3	11.6	0.8	6.3%	20.2%	25.9%	69.3%	4.8%	13.5%	−4.5%	−5.5%
Chemicals	59.3	148.1	207.4	48.2	0.9	23.1%	59.9%	80.9%	18.8%	0.4%	4.8%	−17.2%	−18.1%
Petroleum	30.5	67.1	97.6	12.7	0.3	27.6%	61.9%	88.2%	11.5%	0.3%	2.3%	−15.9%	−20.8%
Clay, Concrete, Glass	21.5	49.1	70.5	5.9	0.2	28.0%	65.2%	92.1%	7.7%	0.2%	1.9%	−19.0%	−21.2%
Primary Metal	6.2	17.8	24.0	12.3	0.7	16.8%	50.0%	64.8%	33.3%	1.9%	6.9%	−11.7%	−12.8%
Secondary & Misc. Mixed	0.3	0.7	1.0	0.1	263.2	0.1%	0.3%	0.4%	0.0%	99.6%	31.0%	−0.2%	−0.2%
Total	277.6	527.0	804.6	393.9	294.9	18.6%	36.7%	53.9%	26.4%	19.7%	7.8%	−12.6%	−2.3%



a. Ton-miles transported by distance for each mode



b. Percentage change in ton-miles transported by distance for truck, CL, and IM modes after ATruck introduction

Figure 59. Freight trip distance before and after ATrucks

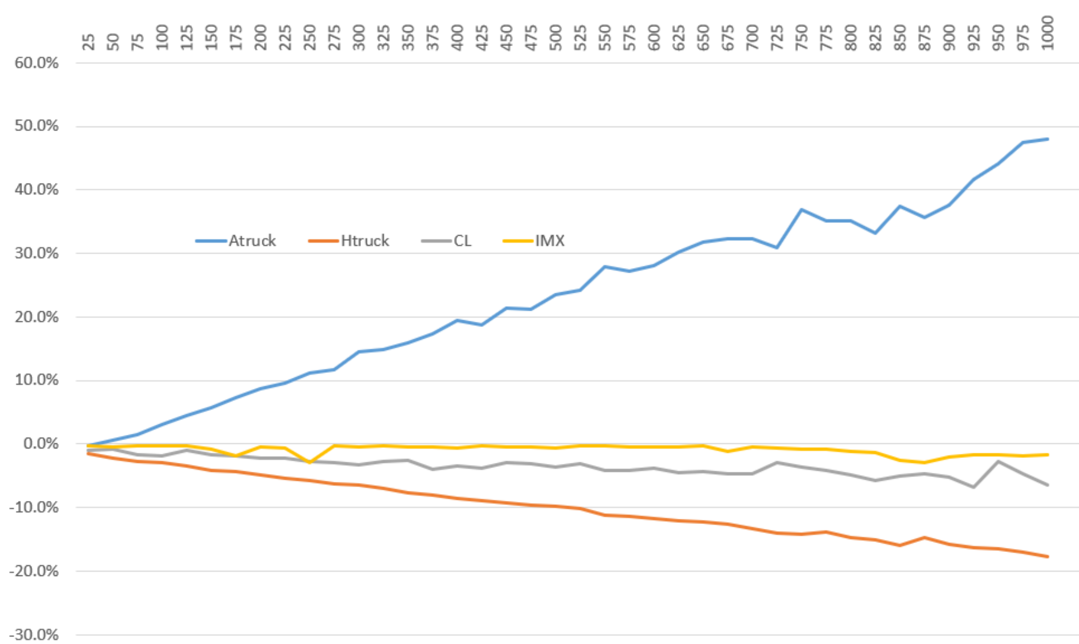


Figure 60. Mode share change (in ton-miles) by trip distance if ATruck cost drops from 1.5 times HTruck cost to the same as HTruck cost

Table 43. Mode splits in freight ton-miles under different ATruck cost assumptions

Commodity	Billion ton-miles	Mode share (ATruck cost = 1.5 HTruck cost)		% change in mode share (ATruck cost = 1.25 HTruck cost)			% change in mode share (ATruck cost = HTruck cost)		
	Truck	HTruck	ATruck	Truck	HTruck	ATruck	Truck	HTruck	ATruck
Agriculture	0.2	69.5%	30.5%	+7.7%	−1.9%	+29.4%	+7.7%	−4.2%	+69.6%
Mining	29.1	22.4%	77.6%	+4.7%	−12.2%	+9.5%	+4.7%	−23.5%	+18.6%
Coal	14.0	29.9%	70.1%	+0.0%	0.0%	+0.0%	+0.0%	0.0%	+0.0%
Nonmetallic minerals	112.6	65.4%	34.6%	+2.5%	−4.3%	+15.2%	+2.5%	−9.2%	+33.6%
Food	93.5	72.9%	27.1%	+1.8%	−8.9%	+30.6%	+1.8%	−19.7%	+68.7%
Consumer manufacturing	2.8	26.2%	73.8%	+3.5%	−7.4%	+7.3%	+3.5%	−14.6%	+14.6%
Non-durable manufacturing	17.7	75.6%	24.4%	+0.9%	−10.6%	+36.4%	+0.9%	−23.9%	+83.5%
Lumber	38.8	45.0%	55.0%	+0.1%	−6.6%	+5.5%	+0.1%	−13.0%	+10.8%
Durable manufacturing	91.0	62.5%	37.5%	+0.3%	−4.7%	+8.5%	+0.3%	−9.5%	+17.3%
Paper	4.3	75.8%	24.2%	+6.2%	−5.5%	+43.0%	+6.2%	−13.3%	+111.4%
Chemicals	207.4	71.4%	28.6%	+1.6%	−8.4%	+26.5%	+1.6%	−18.3%	+58.7%
Petroleum	97.6	68.7%	31.3%	+1.0%	−7.2%	+19.0%	+1.0%	−16.1%	+43.6%
Clay, concrete, glass	70.5	69.6%	30.4%	+0.8%	−8.1%	+21.1%	+0.8%	−17.7%	+46.6%
Primary metal	24.0	74.0%	26.0%	+3.0%	−8.0%	+34.5%	+3.0%	−18.3%	+81.1%
Secondary & misc. mixed	1.0	71.2%	28.8%	+7.8%	−2.0%	+32.2%	+7.8%	−4.9%	+82.7%
Total	804593.7	65.5%	34.5%	+1.5%	−7.2%	+18.2%	+3.5%	−15.8%	+40.2%

Table 43 shows the change in mode share (measured in ton-miles transported) for different commodities as the relationship between ATruck and HTruck costs changes. Cost impacts paper mode choice the most, with an 111.4% increase in the mode choice of ATruck when its cost drops to the same as that of an HTruck, followed by impacts on non-durable manufacturing, secondary and miscellaneous mixed goods, and primary metal mode choices. Overall, the ATruck mode share increases by 40.2% when ATruck and HTruck cost the same, and by 18.2% when ATruck cost is 1.25 times HTruck cost, as compared to when it’s 1.5 times the cost. Interestingly, the mode share for trucks overall does not change much across the scenarios, and therefore the change brought by the reduced cost of ATrucks is mainly shifting truck use from HTrucks to ATrucks. Coal does not show a change in mode share when ATruck cost changes, which is due to the utility function of this commodity, which does not involve a cost term. Coal is more sensitive to transportation time than costs, which could be the reason for this unchanged mode split.

In the long term, the cost of ATrucks will decrease, especially when drivers are no longer needed to operate trucks. Employees will manage and control the truck fleet remotely, without attending to each vehicle, or they may perform other work activities while traveling onboard. Therefore, a sensitivity analysis was conducted on different cost reductions based on current HTruck cost, assuming the cost-ratio factor as ATruck cost divided by HTruck cost, ranging from 1 to 0.5. When the ATruck cost drops from equal to HTruck cost to half of it, this will have little impact on those commodities for which trucks already dominate the mode share. However, for commodities like food, paper, and primary metal, the mode share of trucks increases by over 10%. Tons of agriculture, paper, and secondary and miscellaneous mixed goods transported by truck will increase over 60%, shifted from CL and IM. Overall, truck mode share increases by 4.2% when the cost of ATrucks is reduced to half the HTruck cost, and tons transported by truck increases 7.3% (Table 44).

Table 44. Mode share of ATrucks by commodity with different cost assumptions

ATruck cost/HTruck cost	1	0.9	0.8	0.7	0.6	0.5	% change in mode share from 1 to 0.5	% change in tons from 1 to 0.5
Agriculture	0.2%	0.2%	0.2%	0.2%	0.2%	0.3%	0.1%	64.1%
Mining	68.3%	69.4%	70.5%	71.6%	72.6%	73.6%	5.3%	7.7%
Coal	10.4%	10.4%	10.4%	10.4%	10.4%	10.4%	0.0%	0.0%
Nonmetallic minerals	58.6%	59.6%	60.6%	61.7%	62.7%	63.7%	5.2%	8.8%
Food	75.8%	78.5%	81.0%	83.2%	85.2%	87.0%	11.2%	14.8%
Consumer manufacturing	70.9%	72.0%	73.0%	73.9%	74.9%	75.7%	4.8%	6.8%
Non-durable manufacturing	89.0%	90.8%	92.3%	93.6%	94.7%	95.6%	6.6%	7.4%
Lumber	99.3%	99.4%	99.4%	99.4%	99.4%	99.4%	0.1%	0.1%
Durable manufacturing	91.8%	92.2%	92.5%	92.9%	93.2%	93.5%	1.7%	1.8%
Paper	25.0%	27.6%	30.7%	34.3%	38.6%	43.5%	18.5%	74.0%
Chemicals	80.6%	82.3%	83.8%	85.3%	86.6%	87.8%	7.2%	8.9%
Petroleum	92.4%	93.0%	93.5%	93.9%	94.4%	94.7%	2.3%	2.5%
Clay, concrete, glass	93.5%	94.1%	94.5%	95.0%	95.4%	95.7%	2.2%	2.3%
Primary metal	66.1%	68.9%	71.7%	74.4%	77.0%	79.5%	13.4%	20.3%
Secondary & misc. mixed	0.4%	0.5%	0.5%	0.6%	0.7%	0.8%	0.4%	80.6%
Total	57.0%	57.9%	58.8%	59.7%	60.5%	61.2%	4.2%	7.3%

7.8. Conclusions

This report applies long-distance travel demand models for passengers and freight across the US and Texas, with and without the inclusion of AVs and ATrucks, which are expected on Texas roadways in the upcoming decades. We microsimulate long-distance (over 75 miles one-way) travel choices and vehicle ownership choices for a 10% sample of the US population, and for freight simulations, we use TxDOT's SAM-V4 as a four-step, traditional application. Both applications are for domestic travel only. The passenger travel model indicated that in 2019, there were expected to be 0.85 vehicles per capita in the US, which is consistent with the estimate of 0.83 vehicles per capita in 2020 based on US census data. AV ownership per capita is likely to be 0.33 vehicles after their introduction, assuming a $3,500 technology premium in 2040. Two long-distance trips per month per capita were estimated for the 10% synthetic population, which matches the NHTS data. Mode splits for long-distance, domestic trips prior to AV access were estimated as 64.10% by private automobile, 30.42% by rental car, and 5.49% by airplane. After AVs become available at $0.70 per mile operation cost, mode splits shift to 31.67% by conventional HV, 23.02% by conventional rental car, 23.54% by AV, 18.24% by SAVs, and 3.53% by air. Assuming a $3,500 technology cost premium (e.g., in year 2040), total PMT per capita in long-distance trips is estimated to raise 35%, from 280 to 379 per month.

In the short term, assuming that ATrucks cost 1.5 times as much as HTrucks but save the latter's dwelling time, 1.7 trillion tons of goods are anticipated to be transported by HTrucks and ATrucks across Texas in 2045, while the remaining goods will be transported by CL (23.5%) and IM (16.5%). All 15 modeled industries would witness increases in truck trips and decreases in CL and IM trips due to the use of ATrucks. Across all commodities transported in tons, ton-miles transported by trucks will increase 7.8%, while ton-miles transported by CL and IM will drop by 12.6% and 2.3%, respectively. In the long term, ATruck cost may drop below HTruck cost when on-board truck drivers are no longer required. When ATruck cost drops to half of HTruck cost, truck mode share increases by 4.2% and the number of tons transported by truck increases an additional 7.3%. Such cost decreases will also bring an over 10% increase in trucks' mode share for transporting certain commodities like food, paper, and primary metal. They may also lead to an over 60% increase in truck mode choice for a few selected commodities that do not have a high share of truck use to start with, but this does not significantly impact the truck mode share.

Value of Research (VoR): LD-AV

Quantitative valuation of research projects is difficult since agency implementation of the research results is not yet known. Two assessment areas, value of travel time savings and safety benefits, are examined here to arrive at benefit-cost ratios (BCRs) for TxDOT Research and Technology Innovation Project 0-7081, "Understanding the Impact of Autonomous Vehicles on Long-Distance Travel Mode and Destination Choice in Texas." The project delivers model equations and parameters, as well as direct estimates of AVs' impacts on long-distance passenger and freight travel in Texas. If these models and estimates are used in TxDOT's transportation planning process and those of regional MPOs, border towns, freight carriers, airlines, and ports, Texans can expect a more efficient, less congested, and higher-welfare future. This is because this project's results will help planners, engineers, and businesses prepare for a future with AVs—including SAVs and ATrucks. In addition to the benefits quantified here of lowered driving burdens for AV travelers and benefits of improved infrastructure that supports long-distance AV journeys, the project has other quantifiable benefits. Examples include the following:

- better understanding of future travel demand that can help inform appropriate roadway user fees that are fair and supportive of infrastructure maintenance (to complement motor fuel taxes) for freight and passenger travel in smart vehicles.
- economic benefits (like lower-cost shipping, retail development that caters to more roadway users, and avoidance of overbuilt parking lots and unnecessary airport expansions) from anticipating, designing for, and essentially welcoming AVs onto Texas roadways; and
- better planning for added pavement maintenance needs, which will result from higher roadway use rates, thanks to AVs making "driving" easier.

Perceived Travel Time Benefits (via Lower Driving Burdens)

By facilitating planning for and responding to AV use of Texas roadways, this project ultimately enables lower-cost travel. It does so in two ways: (1) reducing the perceived burden of travel for those previously driving cars and trucks by enabling an infrastructure that supports better use of travelers' time, and (2) opening up new modes of travel so that many travelers and some goods avoid the costs and challenges of air travel or rail use by staying on the roadways.

Various assumptions are used to perform this assessment. All monetary values are expressed in year 2022 dollars, unless otherwise noted. Although travel demand and congestion are expected to rise over time, this analysis relies on network traffic levels from 2019, to avoid COVID-19 impacts.

Annual congestion costs: Roadway delay costs were estimated to be as follows in Texas's major regions (Schrank et al., 2021):

Metro Area	Delay Cost (2019, in billions)	Hours Wasted (2019, in millions)
Houston	\$5.66	263
Dallas-Fort Worth	\$4.81	220
Austin	\$1.73	81.1
San Antonio	\$1.58	71.9
TOTAL	*\$13.8*	*636*

Value of travel time (VOTT): Without AVs, the value of travel time for commuter drivers is estimated to be \$20 per hour (Schrank et al., 2021). Since those previously driving can pursue more productive activities while in AVs (like sleeping, reading emails, eating, or making phone calls), Zhong et al. (2020) and others have estimated that "driver" VOTT will fall by about 32% and 24% in suburban and urban settings, respectively. ("Non-driving" passengers' VOTT is assumed not to change). Using the midpoint of these percentages, 28%, drivers' VOTT while riding in an AV (along with those of passengers) is likely to average around \$14.52 per hour (pivoting off of Schrank et al.'s 2021 estimate).

Long-distance AV mode share on roadways: Assuming a readily available supply of AVs in the marketplace, this research project forecasts that for long-distance passenger travel in Texas, conventional passenger vehicles will have a mode share of just 31.67% in 2040, down from 64.10% today, with AVs making up a 41% share. (Air travel and rental vehicles account for the remaining). This means that roughly 43% (41/(41 + 32+ 23) = 0.33) of ground travelers will be using AVs for their long-distance travels by 2040.

Congestion cost savings: The difference in congestion costs in 2040 between a scenario with AVs and a business-as-usual scenario without AVs can be computed as follows:

$$c_1 = c_0 + h(v_1 - v_0)(1 - m)$$

where:

c_0 = Total cost of congestion prior to AV introduction
c_1 = Total cost of congestion in 2040, after AV introduction
h = Number of hours wasted due to congestion (using today's congestion as a proxy)

v_0 = VOTT for those in conventional vehicles
v_1 = VOTT for those riding in AVs
m = Mode share predicted for personal commuter transport in 2040

The total congestion cost in the scenario with AVs is $11.4 billion in 2040, with a perceived savings of $2.41 billion per year (assuming 2019 traffic volumes persist year after year, which is conservative).

Contribution share of this research: While the researchers believe that the work performed in this project is important for TxDOT to successfully integrate AVs into their long-distance travel demand forecasting by 2040, they acknowledge that many efforts funded by TxDOT will contribute to this. These include construction, policy changes, taxation changes, economic incentives, and more. Researchers also expect much of the time savings benefit will be on short-distance trips. If we hypothetically and conservatively estimate that this project enables just 0.01% of the future improvement in time quality attributed to AVs, the final economic impact of this project with respect to reduction in VOTT lost due to congestion is $241,000 for a single year.

Safety Benefits

Another project contribution would be Texas's faster realization of AVs' safety benefits. This value is estimated by computing fatal crash costs in a future where AVs do not exist, and then adjusting for expected fatal crash reductions due to faster adoption of AVs in Texas thanks to better planning, investment, and policy decisions by TxDOT and other users of this project's work.

Today's crash costs: Texas's 4,489 fatalities caused by crashes in 2021 cost about $51.4 billion (TxDOT, 2021), using figures from the National Safety Council (2020) and the Texas Peace Officer's Crash Reports (CR-3). This analysis just looks at fatalities, since non-fatal injuries typically have much smaller costs than fatal crashes (e.g., just 2.1% in Harmon et al. [2008]). It is important to note that inclusion of all types of crashes would result in a significantly higher total cost.

Future crash costs: A linear extrapolation of TxDOT's fatal crash costs (TxDOT, 2021) puts 2040 crash costs at approximately $40 billion (in today's dollars).

Estimate of crash reductions caused by AVs: AVs' crash reduction benefits are expected to be sizable. The Insurance Institute for Highway Safety (IIHS, 2020) suggests that a 33% crash reduction relative to the rates experienced today with HVs alone is a realistic estimate, assuming agencies maintain today's transportation system's objectives of speed and convenience.

Contribution share of this research: If we hypothetically assume that this project enables an additional 0.01% reduction, attributable to AVs, in Texas roadway fatalities, the final annual economic impact of this project with respect to fatality reduction can be estimated as:

$40,000,000,000 * 33% * 0.01% = $1,720,000

Benefit-Cost Ratio (BCR)

The results of these two analyses are then compared to this project's cost and projected into the future on the Value of Research spreadsheet to arrive at an estimated BCR for this project. The method utilized in the spreadsheet treats the cost savings according to a 20-year decay as a mechanism to assess long-term effects of the project.

TxDOT project cost: $366,199
Conservative benefits estimate: $1,961,000 **(**$241,000 + $1,720,000)
Example (conservative) single-year BCR: 5.3:1

In summary, despite the challenges in predicting the future, this rough quantitative analysis conservatively suggests at least a fourfold benefit from this project's work in 2040, and much more over the course of the following 20 years, while the project also broadly contributes to TxDOT's other efforts to prepare for a future where AVs are readily available in the marketplace.

Appendix A: Qualtrics Survey

The following Qualtrics-software-based survey was presented to respondents.

Introduction & Long-Distance Trips Definitions

Impacts of Self-Driving Vehicles and COVID-19 on Long-Distance Travel Behavior

Dear Respondent,

The **University of Texas at Austin's** Center for Transportation Research is conducting a research study to explore long-distance travel habits under different self-driving vehicle options as well as the influence of the COVID-19 pandemic.

Self-driving vehicles are a new technology with the potential to improve the safety and mobility of long-distance travel. Meanwhile, the COVID-19 pandemic also brings uncertainty to future long-distance travel behavior. Policymakers and transportation planners need to understand how long-distance travel choices vary, in order to **anticipate how traffic conditions will evolve and identify optimal transport policies and network design strategies.**

The survey will take approximately **20-25 minutes** to complete. The survey will ask questions about you, your **recent long-distance travels**, and your travel **preferences**. Your individual responses are **CONFIDENTIAL**. The information in this survey will be used only for research purposes and in ways that will not reveal who you are.

There are **no known risks** involved in your participation in this study and no direct benefits. You are not obligated to participate in the survey and you can stop at any time. But your assistance will facilitate **better transportation planning** for the Austin area and the other U.S. regions.

Your input is VERY IMPORTANT since it is critical that this survey's responses represent all perspectives and types of residents.

If you have any questions or comments about this study, please feel free to contact me directly at **kkockelm@mail.utexas.edu or (512) 471-0210**. If you have any questions about your rights as a research participant, please contact the Office of Research Support by phone at (512) 471-8871 or email at orsc@uts.cc.utexas.edu. Your completion of the survey indicates your willingness to participate in the study.

Thank you very much for your time and cooperation.

Dr. Kara Kockelman
Professor of Transportation Engineering & Faculty Sponsor
www.caee.utexas.edu/prof/kockelman

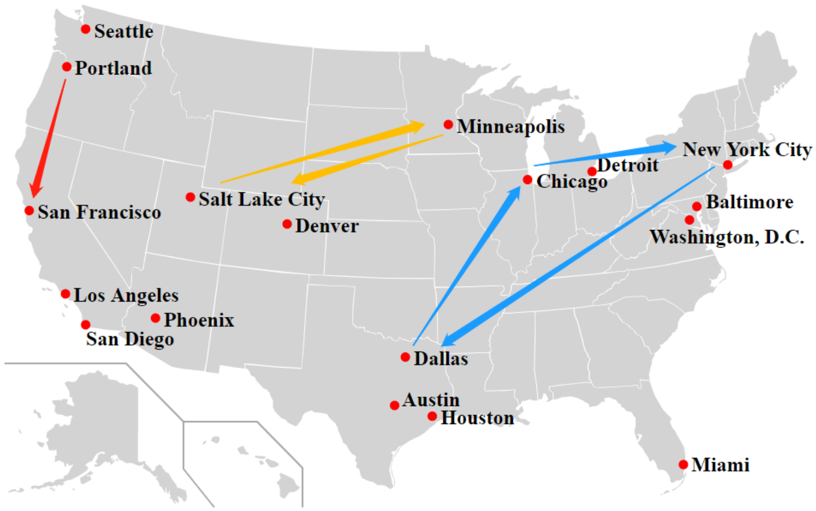
Section 1: Introduction & Long-Distance Trip Definitions

Please **TAKE YOUR TIME** on this survey. There are many unusual questions that require **careful** reading and thoughtful answers. Those completing the survey in less than 10 minutes are unlikely to have completely read all questions. Following are some definitions of terms used in the questions. ***Please read carefully before moving on:***

SELF-DRIVING VEHICLE:
The vehicle is designed to perform **all driving functions** for the **entire trip**. This design anticipates that the driver will provide the destination or navigation input, but the **driver is NOT expected to be available for vehicle control at any time during the trip**.

ONE-WAY TRIPS, ROUND TRIPS, AND CHAINED TRIPS:

- A **round trip (journey)** is defined as a trip to a **destination and directly back again** (e.g., travel from Salt Lake City to Minneapolis and then back to Salt Lake City). A round trip is counted as one trip.
- A **one-way trip (trip segment)** is defined as a **trip from one place to another but not back again** (e.g., travel from Portland to San Francisco).
- A **chained trip** is one in which you made at least one **intermediate stop** while traveling from the origin to the destination (e.g., travel from Dallas to Chicago and then to New York before going back to Dallas).



<u>LONG-DISTANCE TRAVEL:</u>

We define **long-distance travel** as a **one-way trip** with a **distance greater than 75 miles** from the **origin to the destination (or a round trip** with a **distance greater than 150 miles in total)**. For example, the distance between central Austin and central San Antonio is 80 miles, Los Angeles to San Francisco is 382 miles, and Las Vegas to Los Angeles is 270 miles. Here is a link to help you determine the trip distance: https://www.mapdevelopers.com/distance_from_to.php.

Long-Distance Trips Frequency & Trip Purpose

Section 2: Long-Distance Trip Frequency & Trip Purpose

This section will ask you about **long-distance trips** you took in **2019 and 2020** (before and during the COVID-19 pandemic).

*It may be helpful to **grab a calendar and a piece of paper** to write down the trips you've taken with dates, locations, and other relevant details because we will be asking more questions regarding these trips later in the survey.*

Recall that for this survey, a **long-distance trip** is any trip at least **75 miles one-way** from the origin to the destination (or a **round trip** that is more than **150 miles in total**). If you are not sure of the trip distance, you can use this link to verify: https://www.mapdevelopers.com/distance_from_to.php.

When providing the number of trips you have taken:

- Count a **round trip as a single trip.**
- Count a **one-way trip as a single trip.**
- Count **each part of a chained trip (one-way or round trip) as a single trip.**

Note: For questions related to **specific dates**, please regard the **first day** of departure as your **date of the trip**.

Please take **a few minutes** to consider this question.

Did you make any **long-distance trips in calendar year 2019**?

Yes

No

How many **long-distance trips** did you take **in calendar year 2019**?

Please enter **whole numbers** (*integers) only.*

Number of **non-business** long-distance (75 miles one-way) trips	0
Number of **business** long-distance (75 miles one-way) trips	0
Total	0

Did you make any **long-distance trips in calendar year 2020**?

Yes

No

How many **long-distance trips** did you take **in calendar year 2020**?

Please enter **whole numbers** (*integers*) only.

Number of **non-business** long-distance (75 miles one-way) trips	0
Number of **business** long-distance (75 miles one-way) trips	0
Total	0

In calendar year 2019 and 2020, which was your **primary way of traveling, for long-distance one-way trips** between **75 and 500 miles** long? Please select one for each category.

	Personal car	Rental car	Bus	Train	Airplane	Boat/Ship	I did not make such trips that are between 75 and 500 miles.
Non-Business (visit with family &/or friends, vacation/recreation)	○	○	○	○	○	○	○
Business (conferences or out-of-city meetings)	○	○	○	○	○	○	○

In calendar year 2019 and 2020, which was your **primary way of traveling for long-distance one-way trips longer than 500 miles**? Please select one for each category.

	Personal car	Rental car	Bus	Train	Airplane	Boat/Ship	I did not make such trips that are longer than 500 miles.
Non-Business (visit with family &/or friends, vacation/recreation)	○	○	○	○	○	○	○
Business (conferences or out-of-city meetings)	○	○	○	○	○	○	○

What is the average **frequency** of your **long-distance trips** during the following periods? Please select one for each category.

	Before the COVID-19 pandemic	During the COVID-19 pandemic	Your travel plan after COVID-19 is no longer a concern
Never	○	○	○
Every week	○	○	○
More than twice a month	○	○	○
More than once a month	○	○	○
Every month	○	○	○
Every 3 months	○	○	○
Every 6 months	○	○	○
Every 9 months	○	○	○
On average, less than 1 trip per year	○	○	○

What are the **purposes** of your **long-distance trips** during the following periods? (Multiple choices).

	Before the COVID-19 pandemic	During the COVID-19 pandemic	Your travel plan after COVID-19 is no longer a concern
To/From Work	☐	☐	☐

	Before the COVID-19 pandemic	During the COVID-19 pandemic	Your travel plan after COVID-19 is no longer a concern
Work-Related Business	☐	☐	☐
Shopping	☐	☐	☐
Other Family/Personal Business	☐	☐	☐
School/Church	☐	☐	☐
Medical/Dental	☐	☐	☐
Visit Friends/Relatives	☐	☐	☐
Other Social/Recreational	☐	☐	☐
Other	☐	☐	☐
Refused/Don't Know	☐	☐	☐

Self-Driving and Shared Vehicle Technology Definition

Section 3: Self-Driving & Shared Vehicle Technology Definition

Please read carefully before moving forward.

While multiple levels of automation may become available in the future, **full self-driving automation** is the highest level of automation, and is anticipated to become commonly available in the near future —which may have significant impacts on transportation. **In this survey, "self-driving" or "automated" will refer to full self-driving automation**, as defined below, and the vehicle type can be flexible, such as **regular size sedan**, **pick-up truck**, **van**, among others.

Self-Driving Automation: At this level of automation, the vehicle is designed to perform all driving functions for the entire trip. This design anticipates that the driver will provide the destination or navigation input, but the **driver is not expected to be available for vehicle control at any time during the trip.**

Self-Driving Vehicle

Based on the above definitions, how would you describe "**self-driving automation**" or "**autonomous vehicles**"?

... vehicles that **require a human-driver** for their use.

... vehicles that can **assist a human-driver** during their use.

... vehicles that **can perform all driving functions** for the entire trip, but **need the driver's attention**.

... vehicles that **can perform all driving functions** for the entire trip, **regardless** of the availability of a **driver**.

Suppose you are traveling in a **self-driving car**, how likely are you to do the following **activities**?

	Unlikely	Less likely	Neutral	More likely	Absolutely
Talking to other passengers	○	○	○	○	○
Making phone calls or messaging	○	○	○	○	○
Emailing or browsing internet	○	○	○	○	○
Using social media	○	○	○	○	○
Reading for leisure	○	○	○	○	○
Watching videos or playing games	○	○	○	○	○
Listening to music	○	○	○	○	○
Eating or drinking	○	○	○	○	○
Sleeping or resting	○	○	○	○	○

	Unlikely	Less likely	Neutral	More likely	Absolutely
Working or studying	◯	◯	◯	◯	◯
Watching the roadway or landscape	◯	◯	◯	◯	◯

Please read carefully before continuing.

The average price of a new car in the U.S. is **$33,000** and a used car is less than **$20,000**. In the near future, the added cost when purchasing a car with self-driving technology may be unaffordable for many individuals.

However, a fleet operator (such as a business or government agency) can offset this high initial cost with long-term profits. **This fleet of vehicles could be shared among a group of people who will avoid such up-front capital costs while still benefiting from their daily use.** For example, imagine a **taxi fleet of self-driving cars**, in which autonomous vehicle technology will increase each car's usability by sending the car to designated locations without the need for a human driver. For this survey, such an arrangement is termed as "**shared autonomous vehicle fleet**."

Shared Autonomous Vehicle (SAV) Fleets: A fleet of self-driving vehicles, **shared among people in a city or region** that can be shared by people **paying on a per-mile basis**.

Self-Driving Vehicle

Based on the definition provided above, do you think that **shared self-driving cars or SAVs** will allow people to...

... **reduce** upfront vehicle **ownership costs**.

... travel as usual, but **pay** on a **per-mile basis**.

Both of the above.

None of the above.

Long-Distance Trip Revealed & Stated Preferences

Section 4: Long-Distance Trip Revealed & Stated Preferences

In this section, we will ask you a **series of questions** regarding **a long-distance trip** you made during the **pre-COVID-19 period** (*** from January 2019 to March 2020 ***).

Remember that one-way long-distance trips are **more than 75 miles** (or a **round trip of more than 150 miles in total**) from the **origin to the destination**.

*It may be helpful to **grab a calendar and a piece of paper** to write down the trips you've taken with dates, locations, and other relevant details.*

How **long ago** was this trip? If you know the **exact date** of the trip, please answer in **MM/DD/YYYY** form. **Or**, if you are **unsure** of the exact date, please **estimate the date** as best you can.

Please make sure that this **a long-distance trip** that was made during the **pre-COVID-19 period** (*** from January 2019 to March 2020 ***). If you did not make long-distance trips in 2019 and 2020, please recall any trip before **2019.**

Note: please regard the **day of departure** as the **date of the trip**.

Exact date **(MM/DD/YYYY)**:

Or estimated date **(MM/DD/YYYY)**:

Is this a **one-way trip** or a **round trip**? Or is it a part of multiple **long-distance trips chained together**?

Remember that:

- A **round trip** is defined as a trip to a destination and directly back again.
- A **one-way trip** is defined as a trip from one place to another but not back again.
- A **chained trip** is one in which you made at least one intermediate stop while traveling from the origin to the destination.

A round trip.

A one-way trip.

A round trip as part of a chained trip.

A one-way trip as part of a chained trip.

What was the **purpose** of **this trip**? Please select **all** that apply.

Business

To **visit customers**.

To **meet** other **colleagues**.

To **interview** or to **be interviewed**.

For **marketing** and **advertising** purposes.

For **consulting** and **advising** purposes.

To attend a **convention**, **conference**, and/or **seminar**.

For **personal business** (please specify. Example: a legal matter.)

For **other business**-related reasons (please specify):

Non-business

To **visit friends**.

To **visit family** and/or to **attend family**-related events (Examples: weddings, funerals).

For **leisure** and/or **vacation** purposes.

For **recreational** purposes (Examples: sports, hunting, fishing, boating, camping).

For **entertainment** purposes (Examples: attending theaters, sporting events).

For **other non-business**-related reasons (please specify):

What was the **origin** (Examples: home address, work address) of the trip and the **destination** (Examples: hotel, friend's home, meeting site) of the trip?

For **round trips**, please give the destination as the place you went to and came back from. For example, for a round trip from **Chicago** to **Miami** and back to **Chicago**, the origin would be **Chicago, Illinois** and the destination would be **Miami, Florida**.

Note: If you traveled **outside of the country**, please enter the name of the city, select **outside of the United States** in the State dropdown menu, and select the **country**.

	City	State	Country
Origin			
Final Destination			

What were the **origins** and **destinations** of the **long-distance trips** in your **chained trip**?
Please provide **all 75+ miles** trips within your **chained trip** in **chronological order.**

Consider this example of a **chained trip** taken from **Los Angeles** to **San Francisco**, then from **San Francisco** to **Seattle**, and finally from **Seattle** to **Los Angeles**. Your first long-distance trip origin would be **Los Angeles, CA** and your destination would be **San Francisco, CA**. And your second long-distance trip origin would be **San Francisco, CA** and your destination to be **Seattle, WA**. And your third long-distance trip origin would be **Seattle, WA** and your destination would be **Los Angeles, CA**.

Note: If you are still traveling on a round trip (planning to return to your home in the near future), please answer **as if you have completed your trip**.
If your trip involves a location **outside of the country,** please provide the name of the country in selecting from the State drop-down menu.

	Origin		Destination		Was this trip **business**, **non-business**, or related to **both**?	
	City	State	City	State	Business	Non-Business
1st long-distance trip					☐	☐
2nd long-distance trip					☐	☐
3rd long-distance trip					☐	☐
4th long-distance trip					☐	☐
5th long-distance trip					☐	☐
6th long-distance trip					☐	☐
7th long-distance trip					☐	☐
8th long-distance trip					☐	☐

	Origin		Destination		Was this trip **business**, **non-business**, or related to **both**?	
	City	State	City	State	Business	Non-Business
9th long-distance trip					☐	☐
10th long-distance trip					☐	☐

What was the approximate **distance** from the **origin to the furthest destination** of these trips?

Here is a link to help you determine the trip distance: https://www.mapdevelopers.com/distance_from_to.php.

Between 75 and 500 miles.

Between 500 and 1000 miles.

More than 1000 miles.

What was your **primary way of traveling** for this trip (most time spent or longest travel distance)? Please select one.

Personal Car

Rental Car

Bus

Train

Airplane

Boat/Ship

Other (please specify):

What is the estimated **time spent** in the following **activities** for your **one-way** trip to your destination?

Please provide a time for **all** that apply and leave zeros if the option does not apply.

Time scheduling the trip to the airport (e.g., reserving a van or calling Uber/Lyft, renting a car) 0 minut

Time traveling to the airport (driving or being driven by someone else)	0	minut
Time parking at the airport	0	minut
Time spent going through airport security	0	minut
Time waiting at the airport	0	minut
Airplane time	0	minut
Time scheduling the trip from the airport (e.g., reserving a van, calling Uber/Lyft, renting a car)	0	minut
Time traveling from the airport to your destination (driving or being driven by someone else)	0	minut
Time parking at your destination	0	minut
Total	0	minut

Are you willing to **use a self-driving car** to make this **same trip,** if your **travel time is the same** as before?

Yes, I am willing to use a self-driving car for this trip if the travel time is the same.

I may be willing to use a self-driving car for this trip if the travel time is the same.

No, I am not willing to use a self-driving car for this trip if the travel time is the same.

I do not know.

You answered previously that you are **willing to use a self-driving car** to make this **same trip** if your travel time **is the same**. What is the **cost per mile** ($/mile) that you are willing to pay to use a **self-driving car** under this condition?

For reference:
- A trip by **taxi** costs approximately **$2.60 per mile**.
- A trip by **e-hailing service like Uber/Lyft** costs approximately **$2.00 per mile**.
- A trip by **rental car** costs approximately **$0.60 per mile**.

$/mile

0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3

	$/mile
	0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3
Cost per mile ($/mile)	

Are you willing to **use a self-driving car** to make this **same trip**, if your **travel time increases by 10%**?

Yes, I am willing to use a self-driving car for this trip if the travel time increases by 10%.

I may be willing to use a self-driving car for this trip if the travel time increases by 10%.

No, I am not willing to use a self-driving car for this trip if the travel time increases by 10%.

I do not know.

You answered previously that you are **willing to use a self-driving car** to make this **same trip even** if your travel time **increases by 10%**. What is the **cost per mile** ($/mile) that you are willing to pay to use a **self-driving car** under this condition?

For reference:
- A trip by **taxi** costs approximately **$2.60 per mile**.
- A trip by **e-hailing service like Uber/Lyft** costs approximately **$2.00 per mile**.
- A trip by **rental car** costs approximately **$0.60 per mile**.

	$/mile
	0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3
Cost per mile ($/mile)	

Are you willing to **use a self-driving car** to make this **same trip**, if your **travel time increases by 25%**?

Yes, I am willing to use a self-driving car for this trip if the travel time increases by 25%.

I may be willing to use a self-driving car for this trip if the travel time increases by 25%.

No, I am not willing to use a self-driving car for this trip if the travel time increases by 25%.

I do not know.

You answered previously that you are **willing to use a self-driving car** to make this **same trip even** if your travel time **increases by 25%**. What is the **cost per mile** ($/mile) that you are willing to pay to use a **self-driving car** under this condition?

For reference:
- A trip by **taxi** costs approximately **$2.60 per mile**.
- A trip by **e-hailing service like Uber/Lyft** costs approximately **$2.00 per mile**.
- A trip by **rental car** costs approximately **$0.60 per mile**.

	$/mile
	0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3
Cost per mile ($/mile)	

Are you willing to **use a self-driving car** to make this **same trip**, if your **travel time increases by 50%**?

Yes, I am willing to use a self-driving car for this trip if the travel time increases by 50%.
I may be willing to use a self-driving car for this trip if the travel time increases by 50%.
No, I am not willing to use a self-driving car for this trip if the travel time increases by 50%.
I do not know.

You answered previously that you are **willing to use a self-driving car** to make this **same trip even** if your travel time **increases by 50%**. What is the **cost per mile** ($/mile) that you are willing to pay to use a **self-driving car** under this condition?

For reference:
- A trip by **taxi** costs approximately **$2.60 per mile**.
- A trip by **e-hailing service like Uber/Lyft** costs approximately **$2.00 per mile**.
- A trip by **rental car** costs approximately **$0.60 per mile**.

	$/mile
	0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3
Cost per mile ($/mile)	

During your trip, did you **stop** for any reason, such as layovers, highway rest areas, restaurants, or accommodations (including overnight stays during travel) **before** you reached your **final destination**?

Yes, I stopped on my way to my destination.

No, I did not make any stops.

Self-driving cars can **free drivers' hands** and may make it possible to **skip unnecessary stops,** such as layovers. However, traveling using **self-driving cars may cost more**. How likely are you to **choose self-driving cars for your trip**, depending on the following cost assumptions?

	Unlikely	Less likely	Neutral	More likely	Absolutely
The trip costs 50% less	○	○	○	○	○
The trip costs 25% less	○	○	○	○	○
The trip has the same cost as your trip	○	○	○	○	○
The trip costs 25% more	○	○	○	○	○
The trip costs 50% more	○	○	○	○	○

Would you consider making **additional stops during the trip**, either for leisure or business purposes, if you were traveling using a **self-driving car instead**?

Yes, I would include more stops on my way to my destination.

I may consider including more stops.

No, I would not include more stops.

I would have made fewer stops instead.

Would you have **changed your destination to a new place** if you were making this same trip using a **self-driving car instead**?

Yes, I would have chosen a destination with a further distance.

Yes, I would have chosen a destination with a shorter distance.

I may have changed my destination choice.

No, I would not have changed my destination.

During your trip, how long did you **stay** at your destination?
This **excludes** nights spent during transit.

0 days

0 nights

Would you be **willing** to either **extend** or **shorten** your **stay** if you were traveling using a **self-driving car instead**? (Leave zeros if a choice does not apply.)

0 more days

0 fewer days

How many people did you travel with and **who** were they?
(If you traveled alone, please leave this section filled with zeros.)

Family members	0
Friends	0
Colleagues and/or associates	0
Other (please specify):	0
Total	0

How many of your children (if any) did you travel with?

0
1
2+
Prefer not to answer

If **self-driving cars were available** for this trip, would you have traveled with more people?
Please indicate **how many more** people **would have traveled** with you and **who** they are.
(If you wouldn't have taken anyone else, please leave this section filled with zeros.)

______ more family members	0
______ more friends	0
______ more colleagues and/or associates	0
Other (please specify):	0
Total	0

If **self-driving cars were available** for this trip, would you have taken your children with you?

Yes, I would have taken my children with me.

I may have taken my children with me.

No, I wouldn't have taken my children with me.

What were your estimated **expenses** during this trip?
Please include any expenses that you **paid for others** as well.

If you are unsure of the cost of your trip, this website may help in estimating costs: http://www.budgetyourtrip.com/.

Note: If you are still traveling on a round trip (planning to return to your home in the near future), please answer **as if you have completed your trip**.

Accommodation (hotel, Airbnb, cabin rental, etc.)	$ 0
Fuel	$ 0
Other transportation costs (Examples: airplane tickets, train tickets, taxi or bus fares, tolls, etc.)	$ 0
Total	$ 0

How much of the total cost was **covered by someone other than yourself?**
Please give us a **percentage** estimate.

	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
Percentage of cost covered by someone else																					

How much **more** are you **willing to pay** to **use a self-driving car instead** for this long-distance trip?

	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
Percentage of cost increment																					

What are the important **considerations** for you when **deciding** to make a **long-distance** trip using a **self-driving vehicle**? (Multiple choices)

The safety offered by a self-driving car.

The convenience offered by a self-driving car so I can use my travel for other activities instead of driving.

The convenience offered by a self-driving car so I can have a good rest during long-distance driving.

The reliability of the self-driving vehicle.

The self-driving car's ability to self-park.

Other (please specify):

What are important **considerations** for you when **deciding** to **not** make a **long-distance trip** using a **self-driving car**? (Multiple Choices)

I enjoy driving by myself.

I'm not yet confident in the safety benefits provided by self-driving cars.

The potential for faulty software in a self-driving car concerns me.

Confusion arising from the mixture of human-driven and self-driven cars concerns me.

The possibility of being tracked in a self-driving car concerns me.

Other (please specify):

Long-Distance Future Scenario Questions

Section 5: Future Scenarios about Long-distance Travel

In this section, we will ask you a series of questions related to a **hypothetical future** of **affordable self-driving cars**.

In the **next 10 to 15 years**, assume that **self-driving cars** are **on the market** and are **affordable** for you. Companies like FedEx and UPS use self-driving trucks to deliver mail and parcels. Your neighborhood pizza store delivers to your home **without a human in the car**. Companies like **Uber and Lyft use self-driving cars** to lower their spending on driver wages.

Self-Driving Vehicle

In this **hypothetical future** of affordable self-driving cars, what would your **way of travel** be for **one-way trips** between **75 and 500 miles**? Please select one for each category.

	Personal self-driving car	Self-driving rental car	Conventional car	Conventional rental car	Bus	Train	Airplane	Boat/:

	Personal self-driving car	Self-driving rental car	Conventional car	Conventional rental car	Bus	Train	Airplane	Boat/
Non-Business (visit with family &/or friends, vacation/recreation)	○	○	○	○	○	○	○	○
Business (conferences or out-of-city meetings)	○	○	○	○	○	○	○	○

In this **hypothetical future** of affordable self-driving cars, what would your **way of travel** be for **one-way trips** **longer than 500 miles**? Please select one for each category.

	Personal self-driving car	Self-driving rental car	Conventional car	Conventional rental car	Bus	Train	Airplane	Boat/S
Non-Business (visit with family &/or friends, vacation/recreation)	○	○	○	○	○	○	○	○
Business (conferences or out-of-city meetings)	○	○	○	○	○	○	○	○

How do you think your household would **change its travel** **<u>frequency</u>** (trips per year) if you could take **long-distance trips** using a **self-driving car** instead of a human-driven car?

I would make fewer trips per year.

I would make about the same number of trips per year.

I would make 50% more trips per year.

I would make twice as many trips per year.

I would make three times as many trips or more.

I do not know.

How do you think your household would **change the total** **<u>duration</u>** of a **leisure trip** (days per trip) if you could take **long-distance trips** using a **self-driving car** instead of a human-driven car?

I would travel for fewer days.

I would travel for about the same number of days.

I would travel for about 50% more days.

I would travel for about twice as many days.

I would travel for about three times as many days or more.

I do not know.

How do you think your household would **change the travel <u>distance</u>** (miles per year) if you could take **long-distance trips** using a **self-driving car** instead of a human-driven car?

The travel distance would be shorter.

No change in travel distance.

The travel distance would be about 50% more.

The travel distance would be doubled.

The travel distance would be more than doubled.

What would your **departure time preference** be if you could take a **long-distance trip** using a **self-driving car** instead of a human-driven car? (Multiple choices)

Morning

Midday

Afternoon

Night

Other (please specify):

Which of the following factors **would apply** if **self-driving cars** were **available** for **long-distance trips**? (Multiple choices)

I would choose a different destination.

I would travel with more people.

Other (please specify):

If **self-diving cars** were available for **long-distance trips** and you could **rest while riding** the car, would you consider staying in a hotel along the way of your destination? Assume the trip includes overnight hours.

Yes, I would stay in a hotel.

Maybe I would stay in a hotel.

No, I would prefer to stay in the self-driving car.

Other (please specify):

In this section, we will ask you a series of questions related to a **hypothetical future** of **rental self-driving cars**.

If you could take a **long-distance vacation trip** using a **rental self-driving car**, which city in the United States would you travel to? Please select a destination located **more than 200 miles and less than 700 miles** away from your **home**.

	City	State
Desired destination		

Please also give a rough estimation of the total **one-way trip** distance.

Here is a link to help you determine the trip distance: https://www.mapdevelopers.com/distance_from_to.php.

For example:

- A trip from **Houston, TX** to **New Orleans, LA** is approximately **350 miles one-way.**
- A trip from **Portland, OR** to **San Francisco, CA** is approximately **650 miles one-way.**

	200	250	300	350	400	450	500	550	600	650	700
One-way distance (miles)											

You indicated that you will be traveling to a destination location approximately **miles** away from your home.

Assume that you are going to take a **round-trip** (from your **home and back**) **to this destination for vacation,** with **1 companion** (a friend or family member) who has a **driver's license**.

You **and your companion** have **3 travel options** for making this round-trip:

1. You can take your or your companion's **personal car**. The car trip costs about **$0** in total, but you or your companion have to control the wheel all the time. This car trip takes about **{Invalid Expression} / 65 * 2, 1)} hours** (total for the round-trip). You could bring more company in the car and carry your luggage on it, and you have the option to make stops along the way. Note that you may plan to spend **extra time and money to stay in a hotel overnight** if the trip is longer than **500 miles one-way**.

2. You can rent a **self-driving car**. The self-driving car can drive itself to your home to start your journey and you do not have to worry about taking the wheel. This trip by self-driving car costs about **$0** in total, and **{Invalid Expression} / 65 * 2, 1)} hours** along the way. You could also bring more company in the car and carry your luggage on it, and you have the option to make stops along the way. You can choose to **rest in the car** or spend **extra time and money to stay in a hotel overnight** if the trip is longer than **500 miles one-way**. You can choose to **extend the rental** for the duration of your stay at a cost of **$70 per day**.

3. You can also take an **airplane** and fly to the destination and then fly back. You will spend about **$100** in tickets, **$80** in total traveling to/from the airport, and **$50** in baggage fee. Travel time using an airline is about **{Invalid Expression} / 400 + 1.25 * 4, 1)} hours** in total for the whole trip, but a distance greater than **500 miles** may require **layovers** that would extend the trip by **1-2 hours more**.

Based on the previous description, **which option would you and your companion choose**?

Personal car
Self-driving car
Airplane

Now assume that you travel **alone**, but you are offered a **self-driving car** with a **reduced price** while **riding with others** (someone you do not already know).

You **only** have **2 travel options** for making this trip:

1. You can rent a **self-driving car**. The self-driving car can drive itself to your home to start your journey and you do not have to worry about taking the wheel. This trip by self-driving car costs about **$0** in total, and **{Invalid Expression} / 65 * 2, 1)} hours** along the way.

2. You can rent a **self-driving car** and **share the ride** with other users for a reduced price. The self-driving car can drive itself to your home to start your journey and you do not have to worry about taking the wheel. You would **share part of the trip or the whole trip** with other users with **similar destinations**. This trip by self-driving car in a **shared ride** costs about **$0** in total, and **{Invalid Expression} / 65 * 1.1 * 2, 1)} hours** along the way.

Which option would you choose?

Self-driving car

Self-driving car riding with others

Considering the **social-distancing policy** during **COVID-19**, how likely are you to ride a **self-driving** car **with others** (someone you do not already know) for a **reduced price**?

Unlikely

Less likely

Neutral

More likely

Absolutely

Demographics

Section 6: Demographics

Including yourself, **how many people** live in your **household**? (Please do not include anyone who usually lives somewhere else or is just visiting, such as a college student away at school; do not include your roommates if you do not support each other financially.)

1

2

3

4

5 or more (please specify):

Including yourself, **how many workers** usually live in your **household**? (Please include all the persons in your household who get paid for working full-time, part-time, or are self-employed.)

0

1

2

3

4

5 or more (please specify):

What is your **age**?

18 to 24 years

25 to 34 years

35 to 44 years

45 to 54 years

55 to 64 years

65 or more years

What **gender** do you identify with?

Male

Female

Other

Prefer not to say

Which of the following best describes your **ethnicity**?

Hispanic/Latino/Mexican American

Asian/Asian American

Black/African American

American Indian/Native American

White/European White/Caucasian

Mixed/Multiracial

Prefer not to say

Other (please specify):

How many children (those under the age of 16 years) usually live in your home?

0

1

2

3

4 or more (please specify):

Which of the following best describes your **household's total annual income** from all sources, before taxes, for all members of your household?

Less than $10,000

$10,000 to $19,999

$20,000 to $29,999

$30,000 to $39,999

$40,000 to $49,999

$50,000 to $59,999

$60,000 to $74,999

$75,000 to $99,999

$100,000 to $124,999

$125,000 to $149,999

$150,000 to $199,999

$200,000 or more

What is the **highest level of education** you have completed?

I **did not** complete high school.

I completed **high school** (or equivalent).

I completed **some college,** but **no degree**.

I obtained an **associate's or technical degree** (or equivalent).

I obtained a **Bachelor's degree**.

I obtained a **Master's degree**.

I obtained a **PhD**.

Which of the following best describes your **employment status**?

Employed, working 40 or more hours per week (including self-employed).

Employed, working less than 40 hours per week.

Student, working part time.

Student, not working.

Not employed, looking for work.

Not employed, not looking for work.

Retired.

Disabled, not able to work.

What kind of **business or industry** do you work in, have worked in previously, or are planning to work in?

What is your **marital status**?

Single

Married

Divorced

Widowed

What is your **9-digit** **home zip code**? You can find your zip code here: https://m.usps.com/m/ZipLookupAction?search=address

Ex: 76248-9760

Do you live in **Texas**?

Yes

No

Which **U.S. state** do you live in?

	State
State	

What is the **Airport Code** for your **closest airport**?
Example: The **Airport Code** for Austin-Bergstrom International airport is **AUS**.

You can find the airport code here: https://www.faa.gov/nextgen/cip/airport_facility/

If unknown, please add the **full name** of the airport.

Do you have any **high-risk factors** for severe **COVID-19** (such as advanced age, health conditions, or pregnancy)?

Yes

No

Do you have any household members who **have high-risk factors for** **severe COVID-19** (such as advanced age, health conditions, or pregnancy)?

Yes

No

Do you have any household members who **require special attention** while you are away from your household, such as **young children** or those with **disabilities**?

Yes

No

Do any of your household members over the age of 16 have disabilities that may limit their **driving ability**?

Yes

No

Not sure

Do you have at least one **pet** that would need someone other than yourself to care for them while you are gone?

Yes, I have a pet and need someone (such as another household member) to care for them while I am gone.

Yes, I have a pet but do not need anyone to care for them while I am gone.

No, I do not have any pets.

How many **automobiles, vans, and trucks** does your household own?

My household does not own a vehicle.

1

2

3

4

5 or more vehicles

How would you describe your level of **happiness** over the past **year**?

Very Unhappy	Moderately Unhappy	Neutral	Moderately Happy	Very Happy

End of the Survey

If there are **any other long-distance trips** that you would like to tell us about, **please enter them below and include any relevant details** (origin & destination, number of days you were on your trip, the form of transportation you used to get there, costs, reason for trip, whether you could have participated remotely, and anything else you think may be important).

When answering this survey, did you use **a calendar, journal, or other documentation** to help you remember details about your trips?

Yes

No

Other (please specify)

THANK YOU FOR COMPLETING OUR SURVEY!

Do you have any **comments or suggestions** for us?

Appendix B: R Script for Domestic Long-distance Model Application

```
library(tidyverse)
library(dplyr)
library(R.matlab)
library(tidyverse)
library(data.table)
library(MASS)


rm(list=ls())
set.seed(123)


# Upload PopGen, PUMS, EPA SMART Location, and vehicle ownership data and
match them wit each other in a file named Person

##################################
## Trip Frequency

# find trip frequency for each person
# first use the zero inflated to find the probability of having a trip -> use a random # to
assign 0 or 1 to each person
# use negative binomial to find the # of trips and round it to find an integer #

INTERCEPT_ZM = 7.12499
log_HHFAMINC_ZM = -0.40993 # log of HHincome
HHVEH_ZM =  -0.04275


#COUNT MODEL
Theta_Trpfreq =  exp(15.450170)
INTERCEPT_CM = 0.799302
MALE_CM = 0.172263
AGE_CM = -0.002365
WORKER_CM = -0.069870
EDUC_CM = 0.190799 # college or associate degree
NUMADULT_CM = -0.227657
HHVEHCNT_CM = 0.141216
log_HHFAMINC = -0.079772  # log of HHincome

#Zero-Inflated
zinf = NA
zinforg = NA
Person$count = NA
znboutput = NA
RandomNum <-array(runif(nrow(Person)), c(nrow(Person),1))
```

```
zinforg = 1/(1+exp(INTERCEPT_ZM  +  HHVEH_ZM *Person$NUM_VEH_OWNED
+
                    log_HHFAMINC_ZM * log(Person$hhinctype)))

Person$zinf = ifelse(zinforg>RandomNum, 1, 0)

countMu <- exp( INTERCEPT_CM  + MALE_CM *Person$P_Male +
                       AGE_CM *Person$pagetype + WORKER_CM *Person$Employed
+
                       EDUC_CM *Person$P_College_Educated +
                       NUMADULT_CM*(Person$hhsizetype-Person$hhchtype+1)+
                       log_HHFAMINC *log(Person$hhinctype)+
                       HHVEHCNT_CM *Person$NUM_VEH_OWNED)

Person$count <- rnegbin(countMu, theta = Theta_Trpfreq)



## Trip Frequency After AVs
Vmax_Trip = NA
ASC_fewertrip = -1.3134059
ASC_moretrip = -0.6425161

Diability_fewer = 1.0466447
Female_fewer = -0.4734072
AssociateDegree_fewer = -0.5268476

Disability_more = 0.3603576
Female_more = 0.1737178
AssociateDegree_more = 0.2106390

V_fewertrip = c(0,nrow(Person))
V_moretrip = c(0,nrow(Person))

V_sametripfreq = 0

V_fewertrip = ASC_fewertrip + Diability_fewer*Person$Disability
+Female_fewer*Person$P_Female
+AssociateDegree_fewer*Person$P_College_Educated
V_moretrip = ASC_moretrip + Disability_more*Person$Disability
+Female_more*Person$P_Female +
AssociateDegree_more*Person$P_College_Educated
Vmax_Trip <- apply(data.frame(V_sametripfreq, V_fewertrip, V_moretrip), 1, max)


SameTripFreq = ifelse(Vmax_Trip == V_sametripfreq, 1,0)
FewerTrips = ifelse(Vmax_Trip == V_fewertrip, 1,0)
```

```
MoreTrips = ifelse(Vmax_Trip == V_moretrip, 1,0)


multiplier = ifelse(MoreTrips==1, 2 , 1)
multiplier = ifelse(FewerTrips==1, 0.5,multiplier)
Person$count_AV = Person$count*multiplier


Person <- Person %>% filter(Person$zinf==1)
RandomNum_NBfreq <-array(runif(nrow(Person)), c(nrow(Person),1))
Person$count = ifelse((Person$count-floor(Person$count))>RandomNum_NBfreq,
ceiling(Person$count), floor(Person$count))
Person$count_AV = ifelse((Person$count_AV-
floor(Person$count_AV))>RandomNum_NBfreq, ceiling(Person$count_AV),
floor(Person$count_AV))
Person <- Person %>% filter(Person$count>0)

## Trip Frequency Ends
##################################

##################################
## Trip Season

# 1: summer
# 2: winter
# 3: spring
# 4: fall

ASC_sum = 0

ASC_win = -0.62990
b_Male_win =  0.27019
b_Edu_win = 0.21694 # college or associate degree
b_HHsize_win = -0.09684
b_vehct_win = 0.09096

ASC_spr = -0.82755
b_Age_spr = 0.01031
b_Edu_spr =  0.11676
b_Male_spr = 0.27019
b_HHsize_spr = -0.09684
b_vehct_spr = 0.09096
b_Emp_spr =   -0.25027
b_Adlt_spr =   0.08437

ASC_fall =   0.03354
b_Male_fall = 0.27019
b_Edu_fall =  0.16675
b_Inc_fall = 5.9856e-04 # HHincome in $1000
b_HHsize_fall = -0.09684
```

```
b_vehct_fall = 0.09096
b_Emp_fall =   -0.25027
b_Adlt_fall =  -0.11332


Person$Vmax = NA

V_summer = c(0,nrow(Person))
V_winer = c(0,nrow(Person))
V_spring = c(0,nrow(Person))
V_fall = c(0,nrow(Person))

  V_summer = ASC_sum

  V_winter = ASC_win +
    b_Male_win * Person$P_Male+
    b_Edu_win    * Person$P_College_Educated +
    b_HHsize_win *  Person$hhsizetype+
    b_vehct_win * Person$NUM_VEH_OWNED


  V_spring = ASC_spr  +
    b_Age_spr  * Person$pagetype    +
    b_Edu_spr    * Person$P_College_Educated +
    b_Male_spr * Person$P_Male+
    b_HHsize_spr * Person$hhsizetype +
    b_vehct_spr * Person$NUM_VEH_OWNED +
    b_Emp_spr *  Person$Employed +
    b_Adlt_spr* (Person$hhsizetype-Person$hhchtype+1)

  V_fall = ASC_fall + b_Male_fall  *Person$P_Male    +
    b_Edu_fall    * Person$P_College_Educated +
    b_Inc_fall * Person$hhinctype/1000 +
    b_HHsize_fall * Person$hhsizetype +
    b_vehct_fall * Person$NUM_VEH_OWNED +
    b_Emp_fall *   Person$Employed +
    b_Adlt_fall * (Person$hhsizetype-Person$hhchtype+1)



  Person$Vmax <- apply(data.frame(V_summer, V_winter, V_spring, V_fall), 1, max)

  vmax = Person$Vmax


  Person$summer = ifelse(vmax == V_summer, 1,0)
  Person$winter = ifelse(vmax == V_winter, 1,0)
  Person$spring = ifelse(vmax == V_spring, 1,0)
  Person$fall = ifelse(vmax == V_fall, 1,0)
```

```
  sum(Person$summer==1)/nrow(Person)
  sum(Person$winter==1)/nrow(Person)
  sum(Person$spring==1)/nrow(Person)
  sum(Person$fall==1)/nrow(Person)

# Trip Season Ends
#################################


#################################
  ## Trip Purpose

  # Trip_Purposes:
  # 1: Commute
  # 2: Work Trip
  # 3: Shopping
  # 4: Personal
  # 5: School
  # 6: Religious
  # 7: Medical
  # 8: Visit Friends or Relatives
  # 9: Other Social/Recreational
  # 10: Other


  ASC_WRT = 0
  ASC_WBT = -0.543233
  ASC_SHT = 2.916234
  ASC_PBT = 2.498416
  ASC_SCT = 2.050929
  ASC_RLT = -1.665406
  ASC_MDT = -0.15632
  ASC_VFT = 2.805805
  ASC_SOT = 2.537149
  ASC_ORT = -11.122502

  b_AGE_WBT = 0.012155
  b_AGE_SHT = 0.006691
  b_AGE_PBT = 0.012714
  b_AGE_SCT = -0.129569
  b_AGE_RLT = 0.020371
  b_AGE_MDT = 0.040912
  b_AGE_VFT = 0.005476
  b_AGE_ORT = 0.093359

  b_MALE_SHT = -0.498716
  b_MALE_PBT = -0.657965
  b_MALE_MDT = -0.197307
  b_MALE_VFT = -0.730896
```

b_MALE_SOT = -0.621629

b_EDU_WBT = 0.42181
b_EDU_SHT = 0.279359
b_EDU_SCT = 1.980288
b_EDU_VFT = 0.357579
b_EDU_SOT = 0.391097

b_WHITE_SHT = 0.273256
b_WHITE_SCT = -0.548281
b_WHITE_SOT = 0.395993

b_INCOME_WBT = 0.013596
b_INCOME_SHT = 0.007107
b_INCOME_PBT = 0.005563
b_INCOME_SCT = 0.015688
b_INCOME_RLT = 0.00916
b_INCOME_MDT = -0.017628
b_INCOME_VFT = 0.007356
b_INCOME_SOT = 0.008704
b_INCOME_ORT = 0.022453

b_FALL_WBT = 0.738291
b_FALL_PBT = -0.247305
b_FALL_MDT = 0.202146
b_FALL_VFT = 0.337261

b_SPRING_WBT = 0.682746
b_SPRING_SHT = -0.374397
b_SPRING_PBT = -0.67928
b_SPRING_SOT = -0.662715
b_SPRING_ORT = 3.172233

b_hhsize_WBT = -0.061071
b_hhsize_SHT = -0.128132
b_hhsize_PBT = -0.102816
b_hhsize_RLT = -0.106177
b_hhsize_VFT = -0.205214

b_hhveh_WBT = -0.101324
b_hhveh_PBT = -0.115326
b_hhveh_SCT = -0.254812
b_hhveh_VFT = -0.202286
b_hhveh_ORT = -0.989605

b_numadlt_WBT = -0.857942
b_numadlt_SHT = -0.436381
b_numadlt_PBT = -0.188189
b_numadlt_SOT = -0.419459

```
b_wrkr_SHT = -2.177625
b_wrkr_PBT = -1.870397
b_wrkr_SCT = -3.996799
b_wrkr_RLT = -2.012919
b_wrkr_MDT = -3.244297
b_wrkr_VFT = -2.131344
b_wrkr_SOT = -2.391524


Person$Vmax = NA

Person$Trip_Purpose = NA

V_WRT = c(0,nrow(Person))
V_WBT = c(0,nrow(Person))
V_SHT = c(0,nrow(Person))
V_PBT = c(0,nrow(Person))
V_SCT = c(0,nrow(Person))
V_RLT = c(0,nrow(Person))
V_MDT = c(0,nrow(Person))
V_VFT = c(0,nrow(Person))
V_SOT = c(0,nrow(Person))
V_ORT = c(0,nrow(Person))

V_WRT = ASC_WRT

V_WBT = ASC_WBT + b_AGE_WBT  * Person$pagetype   +
  b_EDU_WBT    * Person$P_College_Educated +
  b_INCOME_WBT * Person$hhinctype/1000 +
  b_FALL_WBT * (Person$fall) +
  b_SPRING_WBT * (Person$spring) +
  b_hhsize_WBT * Person$hhsizetype +
  b_hhveh_WBT * Person$NUM_VEH_OWNED +
  b_numadlt_WBT * (Person$hhsizetype-Person$hhchtype+1)

V_SHT = ASC_SHT + b_AGE_SHT  * Person$pagetype    +
  b_MALE_SHT   * Person$P_Male            +
  b_EDU_SHT    * Person$P_College_Educated +
  b_WHITE_SHT  * Person$P_White   +
  b_INCOME_SHT* Person$hhinctype/1000 +
  b_SPRING_SHT * (Person$spring) +
  b_hhsize_SHT * Person$hhsizetype +
  b_numadlt_SHT * (Person$hhsizetype-Person$hhchtype+1) +
  b_wrkr_SHT * Person$Employed

V_PBT = ASC_PBT + b_AGE_PBT  * Person$pagetype    +
  b_MALE_PBT   * Person$P_Male            +
```

```
   b_INCOME_PBT * Person$hhinctype/1000 +
   b_FALL_PBT *(Person$fall) + b_SPRING_PBT * (Person$spring) +
   b_hhsize_PBT * Person$hhsizetype + b_hhveh_PBT * Person$NUM_VEH_OWNED
+
   b_numadlt_PBT * (Person$hhsizetype-Person$hhchtype+1) + b_wrkr_PBT *
Person$Employed

  V_SCT = ASC_SCT + b_AGE_SCT  * Person$pagetype  +
   b_EDU_SCT    * Person$P_College_Educated +
   b_WHITE_SCT  * Person$P_White          +
   b_INCOME_SCT * Person$hhinctype/1000 +
   b_hhveh_SCT * Person$NUM_VEH_OWNED+ b_wrkr_SCT * Person$Employed


  V_RLT = ASC_RLT + b_AGE_RLT  * Person$pagetype    +
   b_INCOME_RLT * Person$hhinctype/1000 +
   b_wrkr_RLT  * Person$Employed

  V_MDT = ASC_MDT + b_AGE_MDT  * Person$pagetype   +
   b_MALE_MDT   * Person$P_Male           +
   b_INCOME_MDT * Person$hhinctype/1000 +
   b_FALL_MDT * (Person$fall) +
   b_wrkr_MDT * Person$Employed

  V_VFT = ASC_VFT + b_AGE_VFT  * Person$pagetype   +
   b_MALE_VFT   * Person$P_Male           +
   b_EDU_VFT    * Person$P_College_Educated +
   b_INCOME_VFT * Person$hhinctype/1000 +
   b_FALL_VFT * (Person$fall) +
   b_hhsize_VFT *  Person$hhsizetype + b_hhveh_VFT * Person$NUM_VEH_OWNED
+
   b_wrkr_VFT * Person$Employed

  V_SOT = ASC_SOT + b_MALE_SOT * Person$P_Male   +
   b_EDU_SOT    * Person$P_College_Educated +
   b_WHITE_SOT  * Person$P_White          +
   b_INCOME_SOT * Person$hhinctype/1000 +
   b_SPRING_SOT * (Person$spring) +
   b_numadlt_SOT * (Person$hhsizetype-Person$hhchtype+1) +
   b_wrkr_SOT * Person$Employed

  V_ORT = ASC_ORT + b_AGE_ORT  * Person$pagetype    +
   b_INCOME_ORT * Person$hhinctype/1000 +
   b_SPRING_ORT * (Person$spring) +
   b_hhveh_ORT * Person$NUM_VEH_OWNED


  Person$Vmax2 <- apply(data.frame(V_WRT, V_WBT, V_SHT, V_PBT, V_ORT,
V_RLT, V_MDT, V_VFT,V_SOT, V_ORT ), 1, max)
  vmax2 = Person$Vmax2
```

```
  Commute = ifelse(vmax2== V_WRT, 1, 0)
  Business= ifelse(vmax2== V_WBT, 1, 0)
  Shopping= ifelse(vmax2== V_SHT, 1, 0)
  Personal= ifelse(vmax2== V_PBT, 1, 0)
  School= ifelse(vmax2== V_ORT, 1, 0)
  Religious= ifelse(vmax2== V_RLT, 1, 0)
  Medical= ifelse(vmax2== V_MDT, 1, 0)
  VisitFriends= ifelse(vmax2== V_VFT, 1, 0)
  Recreational= ifelse(vmax2== V_SOT, 1, 0)
  Other= ifelse(vmax2== V_ORT, 1, 0)

  Person$Trip_Purpose = Commute*1+Business*2+Shopping*3+Personal*4+School*5+
    Religious*6+Medical*7+VisitFriends*8+Recreational*9+Other*10


  CommuteTrp = sum(Person$Trip_Purpose==1)/nrow(Person)*100
  BusinessTrp = sum(Person$Trip_Purpose==2)/nrow(Person)*100
  ShoppingTrp = sum(Person$Trip_Purpose==3)/nrow(Person)*100
  PersonalTrp = sum(Person$Trip_Purpose==4)/nrow(Person)*100
  SchoolTrp = sum(Person$Trip_Purpose==5)/nrow(Person)*100
  ReligiousTrp = sum(Person$Trip_Purpose==6)/nrow(Person)*100
  MedicalTrp = sum(Person$Trip_Purpose==7)/nrow(Person)*100
  VisitFRTrp = sum(Person$Trip_Purpose==8)/nrow(Person)*100
  RecreationalTrp = sum(Person$Trip_Purpose==9)/nrow(Person)*100
  OtherTrp = sum(Person$Trip_Purpose==10)/nrow(Person)*100
  trpPurpose = c(CommuteTrp,BusinessTrp, ShoppingTrp, PersonalTrp,SchoolTrp,
               ReligiousTrp, MedicalTrp, VisitFRTrp, RecreationalTrp,OtherTrp)

# Trip Purpose Ends
##################################

##################################
## Destination Choice before AV

# Distance from home location to all destinations (NUMAs): build dist file for all
NUMAs
Geo2NUMA <- read.csv('.../Tract_Index_NUMA.csv')
numaIDmap <- read.csv('.../numa_map4477.csv')

NUMA = rep(0,NROW(Person))
for (i in 1:NROW(Person)){
  NUMA_temp = Geo2NUMA$numa_id[Person$geo.x[i]]
  if (is.numeric(NUMA_temp)){
    NUMA[i] = numaIDmap$NUMA_map[NUMA_temp+1];
  }
}
Person$NUMA = NUMA
Person <- Person %>% filter(Person$NUMA != "0")
```

```
# Read rJourney data
# Road file (RjourneyData_Road) includes Distance, Toll, CarTime
# Air file (RjourneyData_Air) includes Time, EconomyFare, BusinessFare
# Match destination alternatives with the EPA SMART Location data
# Find Euclidean distances of the home locations to destination alternatives
(NUMA_Dist)
# Data Structure: RjourneyData_Road$Toll[i,j], i & j are NUMAs


# Build MC logsum for all NUMAS to NUMAs
ASC_car_NB = 0
ASC_rcar_NB = -0.709864235
ASC_air_NB = -1.187837823

b_tt_car_NB = -0.281153029
b_tt_rcar_NB = -0.103684715
b_tt_air_NB = -0.270527915

b_cost_NB = -0.00215975

lambda_NB = 1.0

Car_Operation = 0.5
rcar_Operation = 0.1

Num_NUMAs = nrow(SMART) # Number of NUMA zones in this study
car_cost = matrix(0, nrow = Num_NUMAs, ncol = Num_NUMAs)
rcar_cost = matrix(0, nrow = Num_NUMAs, ncol = Num_NUMAs)
air_cost = matrix(0, nrow = Num_NUMAs, ncol = Num_NUMAs)
lg1 = matrix(0, nrow = Num_NUMAs, ncol = Num_NUMAs)
MC_logsum_NB = matrix(0, nrow = Num_NUMAs, ncol = Num_NUMAs)

for (i in 1:Num_NUMAs){
  for (j in 1:Num_NUMAs){
    car_cost[i,j]  = ASC_car_NB  + b_tt_car_NB*RjourneyData_Road$CarTime[i,j]+
b_cost_NB*(RjourneyData_Road$Toll[i,j]+Car_Operation*RjourneyData_Road$Distanc
e[i,j]-(34.38*RjourneyData_Road$CarTime[i,j]/60+18.05))
    rcar_cost[i,j] = ASC_rcar_NB + b_tt_rcar_NB*RjourneyData_Road$CarTime[i,j] +
b_cost_NB*(RjourneyData_Road$Toll[i,j]+Car_Operation*RjourneyData_Road$Distanc
e[i,j]+50*(floor(RjourneyData_Road$CarTime[i,j]/(60*24))+1)-
(10.43*RjourneyData_Road$CarTime[i,j]/60+66.91))
    air_cost[i,j]  = ASC_air_NB  + b_tt_air_NB*RjourneyData_Air$Time[i,j] +
b_cost_NB*(RjourneyData_Air$EconomyFare[i,j]-
(52.65*RjourneyData_Air$Time[i,j]/60+272.98))
    lg1[i,j] = log(exp(car_cost[i,j]/lambda_NB) + exp(rcar_cost[i,j]/lambda_NB))
    explogsum =
exp(lg1[i,j])*ifelse(RjourneyData_Road$CarTime[i,j]>0,1,0)+exp(air_cost[i,j])*ifelse(Rj
ourneyData_Air$Time[i,j]>0,1,0)
    if(explogsum!=0){
```

```
    MC_logsum_NB[i,j] = log(explogsum);
    }else{
      MC_logsum_NB[i,j] = -9999
    }
  }
}

# MCLogsum for business trips
MC_logsum_B = matrix(0, nrow = Num_NUMAs, ncol = Num_NUMAs)

for (i in 1:Num_NUMAs){
  for (j in 1:Num_NUMAs){
    car_cost[i,j]  = ASC_car_NB  + b_tt_car_NB*RjourneyData_Road$CarTime[i,j]+
b_cost_NB*0.5*(RjourneyData_Road$Toll[i,j]+Car_Operation*RjourneyData_Road$Di
stance[i,j]-(34.38*RjourneyData_Road$CarTime[i,j]/60+18.05))
    rcar_cost[i,j] = ASC_rcar_NB + b_tt_rcar_NB*RjourneyData_Road$CarTime[i,j] +
b_cost_NB*0.5*(RjourneyData_Road$Toll[i,j]+Car_Operation*RjourneyData_Road$Di
stance[i,j]+50*(floor(RjourneyData_Road$CarTime[i,j]/(60*24))+1)-
(10.43*RjourneyData_Road$CarTime[i,j]/60+66.91))
    air_cost[i,j]  = ASC_air_NB  + b_tt_air_NB*RjourneyData_Air$Time[i,j] +
b_cost_NB*0.5*(RjourneyData_Air$EconomyFare[i,j]-
(52.65*RjourneyData_Air$Time[i,j]/60+272.98))
    lg1[i,j] = log(exp(car_cost[i,j]/lambda_NB) + exp(rcar_cost[i,j]/lambda_NB))
    explogsum = exp(lg1[i,j])*ifelse(RjourneyData_Road$CarTime[i,j]>0,1,0) +
exp(air_cost[i,j])*ifelse(RjourneyData_Air$Time[i,j]>0,1,0)
    if (explogsum!= 0){
    MC_logsum_B[i,j] = log(explogsum);
    }else{
      MC_logsum_B[i,j]  = -9999
    }
  }
}

# Load job counts and pop_dens for all NUMAs
E8_Ret = SMART$E8_Ret
E8_Off = SMART$E8_off
E8_Ind = SMART$E8_Ind
E8_Svc = SMART$E8_Svc
E8_Ent = SMART$E8_Ent
E8_Ed  = SMART$E8_Ed
E8_Hlth= SMART$E8_Hlth
E8_Pub = SMART$E8_Pub
JobDens = SMART$TotEmp/SMART$Ac_Total
Pop_Dens = SMART$TotPop/SMART$Ac_Total

# Destination choice model coefficients
param_NB = c(0.0172304,0.0015238, -0.0683365, 0.0268633, 0.0191227, -0.0191561)

param_B = c(0.0113599, 0.0047736, -0.0487999, 0.0206213, 0.0569246, -0.04418867)
```

```
# Destination Choice Before AVs
Z = matrix(0, nrow = 8, ncol = Num_NUMAs)
LongDist = matrix(0, nrow = 1, ncol = Num_NUMAs)
V1 = matrix(0, nrow = 1, ncol = Num_NUMAs)
V2 = matrix(0, nrow = 1, ncol = Num_NUMAs)
expv = matrix(0, nrow = 1, ncol = Num_NUMAs)

Person$Distance_D = 0
Person$CarTime_D = 0
Person$CarToll_D = 0
Person$AirTime_D = 0
Person$AirTime_D = 0


Distance_D = 0
CarTime_D = 0
CarToll_D = 0
AirTime_D = 0
AirFare_D = 0
AccessDist_D =0
CarDistance_D =0
Destination =0

start_time <- Sys.time()

for (k in 1:nrow(Person)){
  if (Person$Trip_Purpose[k] != 1|Person$Trip_Purpose[k] != 2){ # non-business trips
      Z[1,] = MC_logsum_NB[Person$NUMA[k],] # from home Numa to Numa i
      Z[2,] = Pop_Dens # Destination i
      Z[3,] = log(E8_Ret+1);# Destination i
      Z[4,] = log(E8_Ind+1);
      Z[5,] = log(E8_Svc+1);
      Z[6,] = log(E8_Pub+1);
      Z[1,] =
ifelse(MC_logsum_NB[Person$NUMA[k],]!=0,MC_logsum_NB[Person$NUMA[k],],-
99999)
      V1 = Z[1,]*param_NB[1] + Z[2,]*param_NB[2] + Z[3,]*param_NB[3]+
Z[4,]*param_NB[4]+Z[5,]*param_NB[5]+ Z[6,]*param_NB[6]
      V1 = ifelse(NUMA_Dist$Distance.eud[Person$NUMA[k],]>25,V1,-99999)
      expv = exp(V1)


    # New: Find the maximum random probability
    P_D = expv/sum(expv)
    r_D = runif(1, 0, 1)
    P_cum_D = cumsum(P_D)
    for (i in 1:length(P_D)){
      if (i==1){
        if (r_D<P_cum_D[1]){
          temp =1
```

```
      }
    }else{
      if ((r_D<P_cum_D[i])&&(r_D>=P_cum_D[i-1])){
          temp = i
      }
    }
  }
  Destination[k] = temp

  Distance_D[k] = NUMA_Dist$Distance.eud[Person$NUMA[k],Destination[k]]
  CarDistance_D[k] = RjourneyData_Road$Distance[Person$NUMA[k],Destination[k]]
  CarTime_D[k] = RjourneyData_Road$CarTime[Person$NUMA[k],Destination[k]]
  CarToll_D[k] = RjourneyData_Road$Toll[Person$NUMA[k],Destination[k]]
  AirTime_D[k] = RjourneyData_Air$Time[Person$NUMA[k],Destination[k]]
  AirFare_D[k] = RjourneyData_Air$EconomyFare[Person$NUMA[k],Destination[k]]
  AccessDist_D[k]=
RjourneyData_Air$AccessDistance[Person$NUMA[k],Destination[k]]

 }else{ # business trips

  Z[1,] = MC_logsum_B[Person$NUMA[k],] # from home Numa to Numa i
  Z[2,] = Pop_Dens # Destination i
  Z[3,] = log(E8_Ret+1)# Destination i
  Z[4,] = log(E8_Ind+1)
  Z[5,] = log(E8_Svc+1)
  Z[6,] = log(E8_Hlth+1)
  LongDist = ifelse(NUMA_Dist$Distance.eud[Person$NUMA[k],]>25,1,-99999)
  Z[1,] =
ifelse(MC_logsum_B[Person$NUMA[k],]!=0,MC_logsum_B[Person$NUMA[k],],-
99999)
  V1 = Z[1,]*param_B[1] + Z[2,]*param_B[2] + Z[3,]*param_B[3]+
Z[4,]*param_B[4]+Z[5,]*param_B[5]+ Z[6,]*param_B[6]
  V1 = ifelse(NUMA_Dist$Distance.eud[Person$NUMA[k],]>25,V1,-99999)
  expv = exp(V1)

  # Find the maximum random prob
  P_D = expv/sum(expv)
  r_D = runif(1, 0, 1)
  P_cum_D = cumsum(P_D)
  for (i in 1:length(P_D)){
    if (i==1){
      if (r_D<P_cum_D[1]){
        temp =1
      }
    }else{
      if ((r_D<P_cum_D[i])&&(r_D>=P_cum_D[i-1])){
          temp = i
      }
    }
  }
```

```
    Destination[k] = temp

    Distance_D[k] = NUMA_Dist$Distance.eud[Person$NUMA[k],Destination[k]]
    CarDistance_D[k] = RjourneyData_Road$Distance[Person$NUMA[k],Destination[k]]
    CarTime_D[k] = RjourneyData_Road$CarTime[Person$NUMA[k],Destination[k]]
    CarToll_D[k] = RjourneyData_Road$Toll[Person$NUMA[k],Destination[k]]
    AirTime_D[k] = RjourneyData_Air$Time[Person$NUMA[k],Destination[k]]
    AirFare_D[k] = RjourneyData_Air$EconomyFare[Person$NUMA[k],Destination[k]]
    AccessDist_D[k]=
RjourneyData_Air$AccessDistance[Person$NUMA[k],Destination[k]]
  }
}
End_time <- Sys.time()
End_time-start_time

DCResults_before = cbind(Destination, Distance_D,CarDistance_D, CarTime_D,
CarToll_D, AirTime_D,AirFare_D, AccessDist_D)
Person$Destination =  Destination
Person$Distance_D = Distance_D
Person$CarDistance_D = CarDistance_D
Person$CarTime_D = CarTime_D
Person$CarToll_D = CarToll_D
Person$AirTime_D =AirTime_D
Person$AirFare_D = AirFare_D
Person$AccessDist_D = AccessDist_D

#################################
# Destination Choice before AV

ASC_AV_NB = -0.090286984
b_tt_AV_NB = -0.113089708
AV_Operation = 0.70
mu_SP = 0.75157487

AV_cost = matrix(0, nrow = Num_NUMAs, ncol = Num_NUMAs)
lg1 = matrix(0, nrow = Num_NUMAs, ncol = Num_NUMAs)
MC_logsum_AVNB = matrix(0, nrow = Num_NUMAs, ncol = Num_NUMAs)

for (i in 1:Num_NUMAs){
  for (j in 1:Num_NUMAs){
    car_cost[i,j]  = ASC_car_NB  + b_tt_car_NB*RjourneyData_Road$CarTime[i,j]+
b_cost_NB*(RjourneyData_Road$Toll[i,j]+Car_Operation*RjourneyData_Road$Distanc
e[i,j]-(34.38*RjourneyData_Road$CarTime[i,j]/60+18.05))
    rcar_cost[i,j] = ASC_rcar_NB + b_tt_rcar_NB*RjourneyData_Road$CarTime[i,j] +
b_cost_NB*(RjourneyData_Road$Toll[i,j]+Car_Operation*RjourneyData_Road$Distanc
e[i,j]+50*(floor(RjourneyData_Road$CarTime[i,j]/(60*24))+1)-
(10.43*RjourneyData_Road$CarTime[i,j]/60+66.91))
    AV_cost[i,j]  = ASC_AV_NB  + b_tt_AV_NB*RjourneyData_Road$CarTime[i,j]+
b_cost_NB*(RjourneyData_Road$Toll[i,j]+AV_Operation*RjourneyData_Road$Distanc
e[i,j]-(46.97*RjourneyData_Road$CarTime[i,j]/60+14.42))
```

```
    air_cost[i,j]  = ASC_air_NB  + b_tt_air_NB*RjourneyData_Air$Time[i,j] +
b_cost_NB*(RjourneyData_Air$EconomyFare[i,j]-
(52.65*RjourneyData_Air$Time[i,j]/60+272.98))
    lg1[i,j] = log(exp(car_cost[i,j]*mu_SP) +
exp(rcar_cost[i,j]*mu_SP)+exp(AV_cost[i,j]*mu_SP))
    explogsum = exp(lg1[i,j])*ifelse(RjourneyData_Road$CarTime[i,j]>0,1,0) +
exp(air_cost[i,j]*mu_SP)*ifelse(RjourneyData_Air$Time[i,j]>0,1,0)
    if (explogsum!=0){
      MC_logsum_AVNB[i,j] = log(explogsum);
    } else{
      MC_logsum_AVNB[i,j] = -9999
    }
  }
}

MC_logsum_AVB = matrix(0, nrow = Num_NUMAs, ncol = Num_NUMAs)


for (i in 1:Num_NUMAs){
  for (j in 1:Num_NUMAs){
    car_cost[i,j]  = ASC_car_NB  + b_tt_car_NB*RjourneyData_Road$CarTime[i,j]+
b_cost_NB*0.5*(RjourneyData_Road$Toll[i,j]+Car_Operation*RjourneyData_Road$Di
stance[i,j]-(34.38*RjourneyData_Road$CarTime[i,j]/60+18.05))
    rcar_cost[i,j] = ASC_rcar_NB + b_tt_rcar_NB*RjourneyData_Road$CarTime[i,j] +
b_cost_NB*0.5*(RjourneyData_Road$Toll[i,j]+Car_Operation*RjourneyData_Road$Di
stance[i,j]+50*(floor(RjourneyData_Road$CarTime[i,j]/(60*24))+1)-
(10.43*RjourneyData_Road$CarTime[i,j]/60+66.91))
    AV_cost[i,j]  = ASC_AV_NB  + b_tt_AV_NB*RjourneyData_Road$CarTime[i,j]+
b_cost_NB*0.5*(RjourneyData_Road$Toll[i,j]+AV_Operation*RjourneyData_Road$Di
stance[i,j]-(46.97*RjourneyData_Road$CarTime[i,j]/60+14.42))
    air_cost[i,j]  = ASC_air_NB  + b_tt_air_NB*RjourneyData_Air$Time[i,j] +
b_cost_NB*0.5*(RjourneyData_Air$EconomyFare[i,j]-
(52.65*RjourneyData_Air$Time[i,j]/60+272.98))
    lg1[i,j] = log(exp(car_cost[i,j]*mu_SP) +
exp(rcar_cost[i,j]*mu_SP)+exp(AV_cost[i,j]*mu_SP))*ifelse(RjourneyData_Road$CarT
ime[i,j]>0,1,0)
    explogsum = exp(lg1[i,j])*ifelse(RjourneyData_Road$CarTime[i,j]>0,1,0) +
exp(air_cost[i,j]*mu_SP)*ifelse(RjourneyData_Air$Time[i,j]>0,1,0)
    if (explogsum!=0){
    MC_logsum_AVB[i,j] = log(explogsum);
  } else{
    MC_logsum_AVB[i,j] = -9999
  }
  }
}

# Distance Change
Vmax_Trip = NA
ASC_shorter = -0.2725579
ASC_longer = 0.7464680
```

```
age_shorter = -0.02938859
white_shorter =  -0.5637792
male_shorter = 0.5670996
LD500_shorter = -0.14343693

age_longer = -0.03477074
white_longer = -0.2190936
male_longer = 0.2751390
LD500_longer = -1.19815479

V_shorter  = c(0,nrow(Person))
V_longer = c(0,nrow(Person))

V_samedist = 0

V_shorter  = ASC_shorter + age_shorter*Person$pagetype
+white_shorter*Person$P_White+
  male_shorter*Person$P_Male +LD500_shorter*(Person$Distance_D>500)



V_longer = ASC_longer + age_longer*Person$pagetype
+white_longer*Person$P_White+
male_longer*Person$P_Male + LD500_longer*(Person$Distance_D>500)

Vmax_Dist <- apply(data.frame(V_samedist, V_shorter , V_longer), 1, max)


SameDist = ifelse(Vmax_Dist == V_samedist, 1,0)
ShorterTrip = ifelse(Vmax_Dist == V_shorter, 1,0)
LongerTrip = ifelse(Vmax_Dist == V_longer, 1,0)

Dist_multiplier = ifelse(LongerTrip==1, 1.5 , 1)
Dist_multiplier2 = ifelse(ShorterTrip==1, 0.5 , 1)
################################

## Destination Choice after AV
Destination_AV = rep(0, NROW(Person))
dim(Destination_AV) = c(NROW(Person))


Distance_DAV = 0
CarTime_DAV = 0
CarToll_DAV = 0

AirTime_DAV = 0
AirTime_DAV = 0
AirFare_DAV = 0
```

```
AccessDist_DAV = 0
CarDistance_DAV = 0

start_time <- Sys.time()
for (k in 1:nrow(Person)){
  if (Person$Trip_Purpose[k] != 1|Person$Trip_Purpose[k] != 2){ #non-business


    Z[1,] = MC_logsum_AVNB[Person$NUMA[k],] # from home Numa to Numa i
    Z[2,] = Pop_Dens # Destination i
    Z[3,] = log(E8_Ret+1);# Destination i
    Z[4,] = log(E8_Ind+1);
    Z[5,] = log(E8_Svc+1);
    Z[6,] = log(E8_Pub+1);
    Z[1,] =
ifelse(MC_logsum_AVNB[Person$NUMA[k],]!=0,MC_logsum_AVNB[Person$NUMA
[k],],-99999)

    V2 = Z[1,]*param_NB[1] + Z[2,]*param_NB[2] + Z[3,]*param_NB[3]+
Z[4,]*param_NB[4]+Z[5,]*param_NB[5]+ Z[6,]*param_NB[6]

    V2 = ifelse(NUMA_Dist$Distance.eud[Person$NUMA[k],]>25,V2,-99999)
    expv = exp(V2)

    # New: Find the maximum random prob
    P_D = expv/sum(expv)
    r_D = runif(1, 0, 1)
    P_cum_D = cumsum(P_D)
    for (i in 1:length(P_D)){
      if (i==1){
        if (r_D<P_cum_D[1]){
          temp =1
        }
      }else{
        if ((r_D<P_cum_D[i])&&(r_D>=P_cum_D[i-1])){
          temp = i
        }
      }
    }

    Destination_AV[k] = temp

    Distance_DAV[k] =
NUMA_Dist$Distance.eud[Person$NUMA[k],Destination_AV[k]]
    CarDistance_DAV[k] =
RjourneyData_Road$Distance[Person$NUMA[k],Destination_AV[k]]
    CarTime_DAV[k] =
RjourneyData_Road$CarTime[Person$NUMA[k],Destination_AV[k]]
    CarToll_DAV[k] = RjourneyData_Road$Toll[Person$NUMA[k],Destination_AV[k]]
    AirTime_DAV[k] = RjourneyData_Air$Time[Person$NUMA[k],Destination_AV[k]]
```

```
  AirFare_DAV[k] =
RjourneyData_Air$EconomyFare[Person$NUMA[k],Destination_AV[k]]
  AccessDist_DAV[k]=
RjourneyData_Air$AccessDistance[Person$NUMA[k],Destination_AV[k]]

 }else{ # business

  Z[1,] = MC_logsum_AVB[Person$NUMA[k],] # from home Numa to Numa i
  Z[2,] = Pop_Dens # Destination i
  Z[3,] = log(E8_Ret+1)# Destination i
  Z[4,] = log(E8_Ind+1)
  Z[5,] = log(E8_Svc+1)
  Z[6,] = log(E8_Hlth+1)
  Z[1,] =
ifelse(MC_logsum_AVB[Person$NUMA[k],]!=0,MC_logsum_AVB[Person$NUMA[k],
],-99999)

  V2 = Z[1,]*param_B[1] + Z[2,]*param_B[2] + Z[3,]*param_B[3]+
Z[4,]*param_B[4]+Z[5,]*param_B[5]+ Z[6,]*param_B[6]
  V2 = ifelse(NUMA_Dist$Distance.eud[Person$NUMA[k],]>25,V2,-99999)
  expv = exp(V2)
  P_D = expv/sum(expv)
  r_D = runif(1, 0, 1)
  P_cum_D = cumsum(P_D)
  for (i in 1:length(P_D)){
   if (i==1){
    if (r_D<P_cum_D[1]){
     temp =1
    }
   }else{
    if ((r_D<P_cum_D[i])&&(r_D>=P_cum_D[i-1])){
     temp = i
    }
   }
  }

  Destination_AV[k] = temp

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

  Distance_DAV[k] =
NUMA_Dist$Distance.eud[Person$NUMA[k],Destination_AV[k]]
  CarDistance_DAV[k] =
RjourneyData_Road$Distance[Person$NUMA[k],Destination_AV[k]]
  CarTime_DAV[k] =
RjourneyData_Road$CarTime[Person$NUMA[k],Destination_AV[k]]
  CarToll_DAV[k] = RjourneyData_Road$Toll[Person$NUMA[k],Destination_AV[k]]
  AirTime_DAV[k] = RjourneyData_Air$Time[Person$NUMA[k],Destination_AV[k]]
  AirFare_DAV[k] =
RjourneyData_Air$EconomyFare[Person$NUMA[k],Destination_AV[k]]
```

```
   AccessDist_DAV[k]=
RjourneyData_Air$AccessDistance[Person$NUMA[k],Destination_AV[k]]
  }
}
end_time <- Sys.time()
end_time-start_time


DCResults_AV = cbind(Destination_AV, Distance_DAV,CarDistance_DAV,
CarTime_DAV, CarToll_DAV, AirTime_DAV,AirFare_DAV, AccessDist_DAV)

Person$Destination_AV = Destination_AV
Person$Distance_DAV = Distance_DAV
Person$CarDistance_DAV = CarDistance_DAV
Person$CarTime_DAV = CarTime_DAV
Person$CarToll_DAV = CarToll_DAV
Person$AirTime_DAV =AirTime_DAV
Person$AirFare_DAV = AirFare_DAV
Person$AccessDist_DAV = AccessDist_DAV

# Destination Choice Ends
##################################

##################################
# Find party size for each individual
theta = 2.086870433
PS_ASC = 0.628116377
PS_Age = -0.003968709
PS_Female = 0.204270333
PS_Commute =  -1.305129877
PS_Business = -1.151497601
PS_Shopping = -0.22035264
PS_VisitFriends = -0.225880759


# Party size before AV
Party_Size_mu = exp(PS_ASC)* exp(PS_Age*Person$pagetype)*
  exp(PS_Female*Person$P_Female)*
  exp(PS_Commute*(Person$Trip_Purpose==1))*

exp(PS_Business*(Person$Trip_Purpose==2))*exp(PS_Shopping*(Person$Trip_Purpose
==3))*
  exp(PS_VisitFriends*(Person$Trip_Purpose==8))

Party_Size <- rnegbin(Party_Size_mu, theta = 2.081392099)

# Party size after AV
Party_Size_AV_mu = exp(PS_ASC)* exp(PS_Age*Person$pagetype)*
  exp(PS_Female*Person$P_Female)*
  exp(PS_Commute*(Person$Trip_Purpose==1))*
```

```
exp(PS_Business*(Person$Trip_Purpose==2))*exp(PS_Shopping*(Person$Trip_Purpose
==3))*
  exp(PS_VisitFriends*(Person$Trip_Purpose==8))
Party_Size_AV <- rnegbin(Party_Size_AV_mu, theta = 2.081392099)

  RandomNum_NumVeh <-array(runif(nrow(Person)), c(nrow(Person),1))
  Party_Size = ifelse((Party_Size-floor(Party_Size))>RandomNum_NumVeh,
ceiling(Party_Size), floor(Party_Size))+1
  Party_Size_AV = ifelse((Party_Size_AV-
floor(Party_Size_AV))>RandomNum_NumVeh, ceiling(Party_Size_AV),
floor(Party_Size_AV))+1

  Person$Party_Size = Party_Size
  Person$Party_Size_AV = Party_Size_AV

# Party Size Ends
##################################

##################################
# Mode Choice before AVs

Utility_car = rep(0,NROW(Person))
Utility_rcar =rep(0,NROW(Person))
Utility_air = rep(0,NROW(Person))
Utility_AV = rep(0,NROW(Person))
Utility_SAV = rep(0,NROW(Person))
Mode = rep(0,NROW(Person))
expsum = 0

b_female_car = -0.207439158
b_Party_rcar = 0.128564158
b_Access_Air = -0.027957021
b_LD_Air = 1.914413897

# Non-Business
for (k in 1:NROW(Person)){
  if (Person$Trip_Purpose[k] != 1|Person$Trip_Purpose[k] != 2){

    Utility_car = ASC_car_NB  + b_tt_car_NB*CarTime_D[k]/60+
              b_cost_NB*0.5*(CarToll_D[k]+CarDistance_D[k]*Car_Operation-
(34.38*CarTime_D[k]/60+18.05))+
              b_female_car* Person$P_Female[k]


    Utility_rcar = ASC_rcar_NB + b_tt_rcar_NB*CarTime_D[k]/60 +

b_cost_NB*0.5*(CarToll_D[k]+CarDistance_D[k]*rcar_Operation+50*ceiling(CarTime
_D[k]/24)-(10.43*CarTime_D[k]/60+66.91))+
               b_Party_rcar*Party_Size[k]
```

```
    Utility_air = ASC_air_NB  + b_tt_air_NB*AirTime_D[k]/60 +
                b_cost_NB*0.5*(AirFare_D[k]-(52.65*AirTime_D[k]/60+272.98)) +
b_Access_Air* AccessDist_D[k]+
                b_LD_Air*(Distance_D[k]>500)



    expsum = exp(Utility_car)+exp(Utility_rcar)+exp(Utility_air)
    prob_car = exp(Utility_car)/expsum
    prob_rcar = exp(Utility_rcar)/expsum
    prob_air = exp(Utility_air)/expsum

    if (CarTime_D[k] == 0){
      prob_car = 0
      prob_rcar = 0
      prob_air = 1
    }else if (AirTime_D[k] == 0){
      prob_car = prob_car
      prob_rcar = prob_rcar
      prob_air = 0
    }
    prob_mode = c(prob_car, prob_rcar, prob_air)
    r_D = runif(1, 0, 1)
    P_cum_D = cumsum(prob_mode)
    for (i in 1:length(prob_mode)){
      if (i==1){
        if (r_D<P_cum_D[1]){
          temp =1
        }
      }else{
        if ((r_D<P_cum_D[i])&&(r_D>=P_cum_D[i-1])){
          temp = i
        }
      }
    }

    Mode[k] = temp

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


  }else{

    # Business

    Utility_car = ASC_car_NB  + b_tt_car_NB*CarTime_D[k]/60+
      b_cost_NB*(CarToll_D[k]+CarDistance_D[k]*Car_Operation-
(34.38*CarTime_D[k]/60+18.05))+
      b_female_car* Person$P_Female[k]
```

```
    Utility_rcar = ASC_rcar_NB + b_tt_rcar_NB*CarTime_D[k]/60 +

b_cost_NB*(CarToll_D[k]+CarDistance_D[k]*rcar_Operation+50*ceiling(CarTime_D[
k]/24)-(10.43*CarTime_D[k]/60+66.91))+
      b_Party_rcar*Party_Size[k]

    Utility_air = ASC_air_NB  + b_tt_air_NB*AirTime_D[k]/60 +
      b_cost_NB*(AirFare_D[k]-(52.65*AirTime_D[k]/60+272.98)) + b_Access_Air*
AccessDist_D[k]+
      b_LD_Air*(Distance_D[k]>500)

    expsum = exp(Utility_car)+exp(Utility_rcar)+exp(Utility_air)
    prob_car = exp(Utility_car)/expsum
    prob_rcar = exp(Utility_rcar)/expsum
    prob_air = exp(Utility_air)/expsum

    if (CarTime_D[k] == 0){
      prob_car = 0
      prob_rcar = 0
      prob_air = 1
    }else if (AirTime_D[k] == 0){
      prob_car = prob_car
      prob_rcar = prob_rcar
      prob_air = 0
    }

    prob_mode = c(prob_car, prob_rcar, prob_air)
    r_D = runif(1, 0, 1)
    P_cum_D = cumsum(prob_mode)
    for (i in 1:length(prob_mode)){
      if (i==1){
        if (r_D<P_cum_D[1]){
          temp =1
        }
      }else{
        if ((r_D<P_cum_D[i])&&(r_D>=P_cum_D[i-1])){
          temp = i
        }
      }
    }

    Mode[k] = temp

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


  }
}
```

```
Person$Mode = Mode

Mode_AV = 0

b_age_AV = -0.02304909
b_edu_AV = 0.724960476

prob_SAV = 0
## Mode Choice after AV
for (k in 1:NROW(Person)){ # Non-business
  if (Person$Trip_Purpose[k] != 1|Person$Trip_Purpose[k] != 2){
   # Non-Business

    Utility_car = (ASC_car_NB  + b_tt_car_NB*CarTime_DAV[k]/60+
      b_cost_NB*0.5*(CarToll_DAV[k]+CarDistance_DAV[k]*Car_Operation-
(34.38*CarTime_DAV[k]/60+18.05))+
      b_female_car* Person$P_Female[k])*mu_SP


    Utility_rcar = (ASC_rcar_NB + b_tt_rcar_NB*CarTime_DAV[k]/60 +

b_cost_NB*0.5*(CarToll_DAV[k]+CarDistance_DAV[k]*rcar_Operation+50*ceiling(C
arTime_DAV[k]/24)-(10.43*CarTime_DAV[k]/60+66.91))+
      b_Party_rcar*Party_Size_AV[k])*mu_SP

    Utility_air = (ASC_air_NB  + (b_tt_air_NB-0.8)*AirTime_DAV[k]/60 +
      b_cost_NB*0.5*abs(AirFare_DAV[k]-(52.65*AirTime_DAV[k]/60+272.98)) +
b_Access_Air* AccessDist_DAV[k]+
      b_LD_Air*(Distance_DAV[k]>500))*mu_SP

    Utility_AV = (ASC_AV_NB  + b_tt_AV_NB*CarTime_DAV[k]/60+
              b_cost_NB*0.5*(CarToll_DAV[k]+CarDistance_DAV[k]*AV_Operation-
(46.97*CarTime_DAV[k]/60+14.42))+
              b_age_AV* Person$pagetype[k]+
              b_edu_AV* Person$P_College_Educated[k])*mu_SP



   expsum = exp(Utility_car)+exp(Utility_AV)+ exp(Utility_rcar)+exp(Utility_air)

   prob_car = exp(Utility_car)/expsum
   prob_AV = exp(Utility_AV)/expsum
   prob_rcar = exp(Utility_rcar)/expsum
   prob_air = exp(Utility_air)/expsum

   if (CarTime_DAV[k] == 0){
    prob_car = 0
    prob_rcar = 0
```

```
    prob_AV = 0
    prob_air = 1
  }else if (AirTime_DAV[k] == 0){
    prob_car = prob_car
    prob_rcar = prob_rcar
    prob_AV = prob_AV
    prob_air = 0
  }
  if (Person$hhcurrveh_AV[k] ==0) {
    prob_SAV = prob_AV
    prob_AV = 0
  }

  prob_mode_AV = c(prob_car, prob_rcar,prob_AV,prob_SAV, prob_air)
  r_D = runif(1, 0, 1)
  P_cum_D = cumsum(prob_mode)
  for (i in 1:length(prob_mode)){
    if (i==1){
      if (r_D<P_cum_D[1]){
        temp =1
      }
    }else{
      if ((r_D<P_cum_D[i])&&(r_D>=P_cum_D[i-1])){
        temp = i
      }
    }
  }

  Mode_AV[k] = temp

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

  }else{

  # Business

    Utility_car = (ASC_car_NB  + b_tt_car_NB*CarTime_DAV[k]/60+
                  b_cost_NB*(CarToll_DAV[k]+CarDistance_DAV[k]*Car_Operation-
(34.38*CarTime_DAV[k]/60+18.05))+
                  b_female_car* Person$P_Female[k])*mu_SP


    Utility_rcar = (ASC_rcar_NB + b_tt_rcar_NB*CarTime_DAV[k]/60 +

b_cost_NB*(CarToll_DAV[k]+CarDistance_DAV[k]*rcar_Operation+50*ceiling(CarTi
me_DAV[k]/24)-(10.43*CarTime_DAV[k]/60+66.91))+
                   b_Party_rcar*Party_Size_AV[k])*mu_SP

    Utility_air = (ASC_air_NB  + (b_tt_air_NB-0.8)*AirTime_DAV[k]/60 +
```

```
                  b_cost_NB*abs(AirFare_DAV[k]-(52.65*AirTime_DAV[k]/60+272.98))
+ b_Access_Air* AccessDist_DAV[k]+
                  b_LD_Air*(Distance_DAV[k]>500))*mu_SP

   Utility_AV = (ASC_AV_NB  + b_tt_AV_NB*CarTime_DAV[k]/60+
                 b_cost_NB*(CarToll_DAV[k]+CarDistance_DAV[k]*AV_Operation-
(46.97*CarTime_DAV[k]/60+14.42))+
                 b_age_AV* Person$pagetype[k]+
                 b_edu_AV* Person$P_College_Educated[k])*mu_SP

   expsum = exp(Utility_car)+exp(Utility_AV)+ exp(Utility_rcar)+exp(Utility_air)

   prob_car = exp(Utility_car)/expsum
   prob_AV = exp(Utility_AV)/expsum
   prob_rcar = exp(Utility_rcar)/expsum
   prob_air = exp(Utility_air)/expsum

  if (CarTime_DAV[k] == 0){
   prob_car = 0
   prob_rcar = 0
   prob_AV = 0
   prob_air = 1
  }else if (AirTime_DAV[k] == 0){
   prob_car = prob_car
   prob_rcar = prob_rcar
   prob_AV = prob_AV
   prob_air = 0
  }
  if (Person$hhcurrveh_AV[k] ==0) {
   prob_SAV = prob_AV
   prob_AV = 0
  }
  prob_mode_AV = c(prob_car, prob_rcar, prob_AV,prob_SAV, prob_air)
  r_D = runif(1, 0, 1)
  P_cum_D = cumsum(prob_mode)
  for (i in 1:length(prob_mode)){
   if (i==1){
    if (r_D<P_cum_D[1]){
     temp =1
    }
   }else{
    if ((r_D<P_cum_D[i])&&(r_D>=P_cum_D[i-1])){
     temp = i
    }
   }
  }
  Mode_AV[k] = temp
  }

}
```

```
Person$Mode_AV = Mode_AV
# Mode Choice Ends
##################################

# Report PMT, VMT, and mode share before and after AVs
```

Appendix C: Python Script for Weighting Process via Iterative Proportional Fitting

```
from ipfn import ipfn
import numpy as np
import csv
import datetime

def read_targets(filename):
    # create two sets that works for the collected survey and the ACS data
    # matrix index =  education, gender, ege, region

    acs = np.zeros((4,2,6,5), dtype=int)
    # read the target form ACS input
    acs_rd = csv.reader(open(filename, 'r'))
    gender_index = 0
    age_index = 0
    region_index = 0
    for row in acs_rd:
        if row[0].startswith("Northeast") or row[0].startswith("Midwest") \
                or row[0].startswith("South") or row[0].startswith("West"):
            acs[:, gender_index, age_index, region_index] = [int(row[1]),
                                 int(row[2]), int(row[3]), int(row[4])]
            region_index += 1
        elif row[0].startswith("Texas"):
            acs[:, gender_index, age_index, region_index] = [int(row[1]),
                                 int(row[2]), int(row[3]), int(row[4])]
            acs[0, gender_index, age_index, region_index -2 ] -= int(row[1])
            acs[1, gender_index, age_index, region_index -2 ] -= int(row[2])
            acs[2, gender_index, age_index, region_index -2 ] -= int(row[3])
            acs[3, gender_index, age_index, region_index -2 ] -= int(row[4])
            age_index += 1
            region_index = 0
        if age_index == 6:
            gender_index = 1
            age_index = 0
    acs = acs[0,:,:,:] + acs[1,:,:,:] + acs[2,:,:,:] + acs[3,:,:,:]
    print("ACS Matrix ---------")
    print(acs)

    return acs


def survey_response(filename,date):
    survey = np.zeros((4, 2, 6, 5), dtype=int)
    gender_index, age_index, region_index, education_index = set_indices()
    # read dataset from survey data
    db_rd = csv.reader(open(filename, 'r'))
    for row in db_rd:
        if row[0].startswith("5") or row[0].startswith("6") :
            if row[187] in gender_index.keys() \
                and row[186] in age_index.keys() \
                and row[199] in region_index.keys() \
                and row[193] in education_index.keys():
                if date == True:
                    if row[47] == "NA":
                        # print(row[48])
```

```
                    read_date = row[48].split('/')
                    d = datetime.datetime(2000+int(read_date[2]),
                         int(read_date[0]), int(read_date[1]))
                else:
                    # print(row[47])
                    read_date = row[47].split('/')

                    d = datetime.datetime(2000+int(read_date[2]),
                         int(read_date[0]), int(read_date[1]))
                if d < datetime.datetime(2020, 4, 1):
                    survey[education_index[row[193]],
                         gender_index[row[187]], age_index[row[186]],
                         region_index[row[199]]] += 1
            else:
                survey[education_index[row[193]], gender_index[row[187]],
                         age_index[row[186]], region_index[row[199]]] += 1

    survey = survey[0, :, :, :] + survey[1, :, :, :] + survey[2, :, :, :] +
                         survey[3, :, :, :]
    print("Survey Matrix "," ---------", date, " ---------")
    print("Total valid = ",survey[:,:,:].sum())
    print(survey)
    return survey

def set_indices():
    # define the index of each category
    gender_index = {"Male" : 0, "Female" : 1}
    age_index = {"18 to 24 years" : 0, "25 to 34 years" : 1, "35 to 44 years" :
                         2, "45 to 54 years" : 3,
                 "55 to 64 years" : 4, "65 or more years" : 5}
    region_index = {"CT" : 0, "ME" : 0, "MA" : 0, "NH" : 0, "NJ" : 0, "NY" : 0,
                         "PA" : 0, "RI" : 0, "VT" : 0,
                    "IL" : 1, "IN" : 1, "IA" : 1, "KS" : 1, "MI" : 1, "MN" : 1,
                         "MO" : 1, "NE" : 1, "ND" : 1, "OH" : 1, "SD":1,
                         "WI" : 1,
                    "AL" : 2, "AR" : 2, "DE" : 2, "FL" : 2, "GA" : 2, "KY" : 2,
                         "LA" : 2, "MD" : 2, "MS" : 2,
                    "NC" : 2, "OK" : 2, "SC" : 2, "TN" : 2, "VA" : 2, "WV" : 2,
                         "Washington D.C." : 2,
                    "AK" : 3, "AZ" : 3, "CA" : 3, "CO" : 3, "HI" : 3, "ID" : 3,
                         "MT" : 3, "NV" : 3, "NM" : 3, "OR" : 3, "UT" : 3,
                         "WA" : 3, "WY" : 3,
                    "NA" : 4}
    education_index = {"I completed high school (or equivalent)." : 0, "I did
                         not complete high school." : 0,
                       "I completed some college, but no degree." : 1, "I
                         obtained an associate's or technical degree (or
                         equivalent)." : 1,
                       "I obtained a Bachelor's degree." : 2,
                       "I obtained a Master's degree." : 3, "I obtained a PhD."
                         : 3}
    return gender_index, age_index, region_index, education_index

def IPF(acs, survey):
    #calculate the marginal from ACS data
    #education(4),
    #gender(2), ege(6), region(5)
    xipp = np.array([acs[0,:,:].sum(), acs[1,:,:].sum()])
    xpjp = np.array([acs[:,0,:].sum(), acs[:,1,:].sum(), acs[:,2,:].sum(),
                         acs[:,3,:].sum(), acs[:,4,:].sum(),
                         acs[:,5,:].sum()])
```

```
    xppk = np.array([acs[:,:,0].sum(), acs[:,:,1].sum(), acs[:,:,2].sum(),
                         acs[:,:,3].sum(), acs[:,:,4].sum()])

    xijp = np.array([[acs[0, 0, :].sum(), acs[0, 1, :].sum(), acs[0, 2,
                         :].sum(), acs[0, 3, :].sum(), acs[0, 4, :].sum(),
                         acs[0, 5, :].sum()],
                     [acs[1, 0, :].sum(), acs[1, 1, :].sum(), acs[1, 2,
                         :].sum(), acs[1, 3, :].sum(), acs[1, 4, :].sum(),
                         acs[1, 5, :].sum()]])
    xpjk = np.array([[acs[:, 0, 0].sum(), acs[:, 0, 1].sum(), acs[:, 0,
                         2].sum(), acs[:, 0, 3].sum(), acs[:, 0, 4].sum()],
                     [acs[:, 1, 0].sum(), acs[:, 1, 1].sum(), acs[:, 1,
                         2].sum(), acs[:, 1, 3].sum(), acs[:, 1, 4].sum()],
                     [acs[:, 2, 0].sum(), acs[:, 2, 1].sum(), acs[:, 2,
                         2].sum(), acs[:, 2, 3].sum(), acs[:, 2, 4].sum()],
                     [acs[:, 3, 0].sum(), acs[:, 3, 1].sum(), acs[:, 3,
                         2].sum(), acs[:, 3, 3].sum(), acs[:, 3, 4].sum()],
                     [acs[:, 4, 0].sum(), acs[:, 4, 1].sum(), acs[:, 4,
                         2].sum(), acs[:, 4, 3].sum(), acs[:, 4, 4].sum()],
                     [acs[:, 5, 0].sum(), acs[:, 5, 1].sum(), acs[:, 5,
                         2].sum(), acs[:, 5, 3].sum(), acs[:, 5,
                         4].sum()]])
    xipk = np.array([[acs[0, :, 0].sum(), acs[0, :, 1].sum(), acs[0, :,
                         2].sum(), acs[0, :, 3].sum(), acs[0, :, 4].sum()],
                     [acs[1, :, 0].sum(), acs[1, :, 1].sum(), acs[1, :,
                         2].sum(), acs[1, :, 3].sum(), acs[1, :,
                         4].sum()]])

    aggregates = [xipp, xpjp, xppk, xijp, xpjk, xipk]
    dimensions = [[0], [1], [2], [0, 1], [1, 2], [0,2]]

    IPF = ipfn.ipfn(survey, aggregates, dimensions)
    m = IPF.iteration()
    print("Weighted Matrix ---------")
    print(m.astype(int))

    # calculate weights for each category = total weights in that category /
                         number of collected samples in that category
    weights = np.zeros((2, 6, 5), dtype=int)
    for i in range (2):
        for j in range (6):
            for k in range (5):
                weights[i,j,k] = m[i,j,k] / survey[i,j,k]
    print("Weights Matrix ---------")
    print(weights)

    return weights


def write_db(weights,filename, out_filename):
    # Read and calculate the average of the weights
    gender_index, age_index, region_index, education_index = set_indices()
    total = []
    total_tx = []
    with open(filename, 'r') as csvinput:
        for row in csv.reader(csvinput):
            if row[0].startswith("5") or row[0].startswith("6"):
                if row[187] in gender_index.keys() \
                        and row[186] in age_index.keys() \
                        and row[199] in region_index.keys() \
                        and row[193] in education_index.keys():
```

```
                total += [weights[gender_index[row[187]],
                         age_index[row[186]], region_index[row[199]]]]
            if row[187] in gender_index.keys() \
                    and row[186] in age_index.keys() \
                    and row[199] == "NA" \
                    and row[193] in education_index.keys():
                total_tx += [weights[gender_index[row[187]],
                         age_index[row[186]], region_index[row[199]]]]
        avg = sum(total) / len(total)
        avg_tx = sum(total_tx) / len(total_tx)
    csvinput.close()

    # write weights to database for each category
    with open(filename, 'r') as csvinput:
        with open(out_filename, 'w', newline='') as csvoutput:
            writer = csv.writer(csvoutput)
            for row in csv.reader(csvinput):
                if row[0].startswith("5") or row[0].startswith("6"):
                    if row[187] in gender_index.keys() \
                            and row[186] in age_index.keys() \
                            and row[199] in region_index.keys() \
                            and row[193] in education_index.keys():
                        wg = weights[gender_index[row[187]],
                             age_index[row[186]], region_index[row[199]]]
                        if row[199] == "NA":
                            wg_tx = weights[gender_index[row[187]],
                             age_index[row[186]], region_index[row[199]]]
                            writer.writerow(row + [wg, wg / avg, wg_tx /
                             avg_tx])
                        else:
                            writer.writerow(row + [wg, wg / avg, 0])
                    else:
                        writer.writerow(row + [0, 0, 0])
                else:
                    writer.writerow(row + ['Weights', 'Normalized Weights',
                         'Normalized TX Weights'])
        csvoutput.close()
    csvinput.close()

if __name__ == "__main__":

    acs = read_targets("ACS_export.csv")

    survey = survey_response("June_15_filtered2_precovid.csv", False)
    survey_datefilter = survey_response("June_15_filtered2_precovid.csv", True)

    weights = IPF(acs, survey)
    weights_datefilter = IPF(acs, survey_datefilter)


                         write_db(weights,"June_15_filtered2_precovid.csv",
                         "June_15_filtered2_precovid_weights.csv")
    write_db(weights_datefilter, "June_15_filtered2_precovid.csv",
                         "June_15_precovid_datefiltered2_weights.csv")
```

Appendix D: Python Script for Data Summary Statistics

```
#!/usr/bin/env python
# coding: utf-8

# In[1]:


import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
from matplotlib.ticker import FuncFormatter
from pandas.tseries.holiday import USFederalHolidayCalendar as calendar
import math
import csv

get_ipython().run_line_magic('matplotlib', 'inline')
plt.style.use('seaborn-paper')
plt.rcParams['font.family'] = 'sans-serif'


# In[2]:


df_raw = pd.read_csv('June_15_filtered2_precovid_weights.csv')
pd.set_option('display.max_columns', 200)
pd.set_option('display.max_rows', 200)


# In[3]:


#df.head()


# In[4]:


#df.info(verbose=True)


# In[5]:


#df.loc[0]


# In[6]:


#region filter
filt = df_raw['RegionTX'] == 'Yes'
tx_db = 1
if tx_db == 1:
    outkey = 'Texas'
```

```
    df = df_raw[filt].copy()
else:
    outkey = 'US'
    df = df_raw.copy()


# In[7]:


#raw data processing
df['2019_nb_Weights'] = df['2019_nb'] * df['Weights']
df['2019_b_Weights'] = df['2019_b'] * df['Weights']
df['2020_nb_Weights'] = df['2020_nb'] * df['Weights']
df['2020_b_Weights'] = df['2020_b'] * df['Weights']

df['schedule1_w'] = df['schedule1'] * df['Weights']
df['access_w'] = df['access'] * df['Weights']
df['parking1_w'] = df['parking1'] * df['Weights']
df['security_w'] = df['security'] * df['Weights']
df['wait_w'] = df['wait'] * df['Weights']
df['onboard_w'] = df['onboard'] * df['Weights']
df['schedule2_w'] = df['schedule2'] * df['Weights']
df['egress_w'] = df['egress'] * df['Weights']
df['parking2_w'] = df['parking2'] * df['Weights']

df['av_days_p_w'] = df['av_days_p'] * df['Weights']
df['av_days_n_w'] = df['av_days_n'] * df['Weights']

df['family_w'] = df['family'] * df['Weights']
df['friends_w'] = df['friends'] * df['Weights']
df['colleagues_w'] = df['colleagues'] * df['Weights']

df['family_more_w'] = df['family_more'] * df['Weights']
df['friends_more_w'] = df['friends_more'] * df['Weights']
df['colleagues_more_w'] = df['colleagues_more'] * df['Weights']

if tx_db != 1:
    region_index = {"CT" : 0, "ME" : 0, "MA" : 0, "NH" : 0, "NJ" : 0, "NY" : 0,
                             "PA" : 0, "RI" : 0, "VT" : 0,
                    "IL" : 1, "IN" : 1, "IA" : 1, "KS" : 1, "MI" : 1, "MN" : 1,
                             "MO" : 1, "NE" : 1, "ND" : 1,
                    "OH" : 1, "SD" : 1, "WI" : 1,
                    "AL" : 2, "AR" : 2, "DE" : 2, "FL" : 2, "GA" : 2, "KY" : 2,
                             "LA" : 2, "MD" : 2, "MS" : 2,
                    "NC" : 2, "OK" : 2, "SC" : 2, "TN" : 2, "VA" : 2, "WV" : 2,
                             "Washington D.C." : 2,
                    "AK" : 3, "AZ" : 3, "CA" : 3, "CO" : 3, "HI" : 3, "ID" : 3,
                             "MT" : 3, "NV" : 3, "NM" : 3,
                    "OR" : 3, "UT" : 3, "WA" : 3, "WY" : 3,
                    "TX" : 4}
    #df['Region']
    List = []
    for i in range(df.shape[0]):
        if df.loc[i]['RegionTX'] == "Yes":
            region = 4
        else:
            region = region_index[df.loc[i]['RegionState']]
        List += [region]
    df.insert(df.shape[1],'Region', List, allow_duplicates = True)

TotalWeightSum = df['Weights'].sum()
```

```
# # Long-Distance Trips Frequency & Trip Purpose
#

# LD Trip Frequency

# In[8]:


Output1 = [["","Non-Business","Non-Business","Business","Business"]]
Output1 += [["", "2019","2020","2019","2020"]]

filt1 = df['2019'] == 'Yes'
filt2 = df['2020'] == 'Yes'
df_2019 = df.loc[filt1]
df_2020 = df.loc[filt2]

b_19 = df['2019_b_Weights'].sum()/(df['2019_nb_Weights'].sum() +
                             df['2019_b_Weights'].sum())
nb_19 = df['2019_nb_Weights'].sum()/(df['2019_nb_Weights'].sum() +
                             df['2019_b_Weights'].sum())
b_20 = df['2020_b_Weights'].sum()/(df['2020_nb_Weights'].sum() +
                             df['2020_b_Weights'].sum())
nb_20 = df['2020_nb_Weights'].sum()/(df['2020_nb_Weights'].sum() +
                             df['2020_b_Weights'].sum())
Output1 += [["",nb_19,nb_20, b_19,b_20]]


# In[9]:


Output1 += [['Gender']]

filt19F = df_2019['Gender'] == 'Female'
filt19M = df_2019['Gender'] == 'Male'
filt20F = df_2020['Gender'] == 'Female'
filt20M = df_2020['Gender'] == 'Male'


Gender19_grp = df_2019.groupby(['Gender'])
Gender20_grp = df_2020.groupby(['Gender'])

G_19nb =
                             Gender19_grp['2019_nb_Weights'].sum()/sum(Gender19
                             _grp['2019_nb_Weights'].sum())
G_20nb =
                             Gender20_grp['2020_nb_Weights'].sum()/sum(Gender20
                             _grp['2020_nb_Weights'].sum())
G_19b =
                             Gender19_grp['2019_b_Weights'].sum()/sum(Gender19_
                             grp['2019_b_Weights'].sum())
G_20b =
                             Gender20_grp['2020_b_Weights'].sum()/sum(Gender20_
                             grp['2020_b_Weights'].sum())

Output1 += [["Female", G_19nb[0], G_20nb[0], G_19b[0], G_20b[0]]]
Output1 += [["Male", G_19nb[1], G_20nb[1], G_19b[1], G_20b[1]]]


# In[10]:
```

```
Output1 += [["Age"]]

Age19_grp = df_2019.groupby(['Age'])
Age20_grp = df_2020.groupby(['Age'])
A_19nb =
                             Age19_grp['2019_nb_Weights'].sum()/sum(Age19_grp['
                             2019_nb_Weights'].sum())
A_20nb =
                             Age20_grp['2020_nb_Weights'].sum()/sum(Age20_grp['
                             2020_nb_Weights'].sum())
A_19b = Age19_grp['2019_b_Weights'].sum()/sum(Age19_grp['2019_b_Weights'].sum())
A_20b = Age20_grp['2020_b_Weights'].sum()/sum(Age20_grp['2020_b_Weights'].sum())

Output1 += [["18 to 24 years", A_19nb[0], A_20nb[0], A_19b[0], A_20b[0]]]
Output1 += [["25 to 34 years", A_19nb[1], A_20nb[1], A_19b[1], A_20b[1]]]
Output1 += [["35 to 44 years", A_19nb[2], A_20nb[2], A_19b[2], A_20b[2]]]
Output1 += [["45 to 54 years", A_19nb[3], A_20nb[3], A_19b[3], A_20b[3]]]
Output1 += [["55 to 64 years", A_19nb[4], A_20nb[4], A_19b[4], A_20b[4]]]
Output1 += [["65 or more years", A_19nb[5], A_20nb[5], A_19b[5], A_20b[5]]]


# In[11]:


Output1 += [["Location"]]

if tx_db != 1:
    #0:Northeast; 1:Midwest; 2:South, 3:West

    Region19_grp = df_2019.groupby(['Region'])
    Region20_grp = df_2020.groupby(['Region'])


    R_19nb =
                             Region19_grp['2019_nb_Weights'].sum()/sum(Region19
                             _grp['2019_nb_Weights'].sum())
    R_20nb =
                             Region20_grp['2020_nb_Weights'].sum()/sum(Region20
                             _grp['2020_nb_Weights'].sum())
    R_19b =
                             Region19_grp['2019_b_Weights'].sum()/sum(Region19_
                             grp['2019_b_Weights'].sum())
    R_20b =
                             Region20_grp['2020_b_Weights'].sum()/sum(Region20_
                             grp['2020_b_Weights'].sum())

    Region19_grp['Region'].value_counts()

    Output1 += [["Northeast", R_19nb[0], R_20nb[0], R_19b[0], R_20b[0]]]
    Output1 += [["Midwest", R_19nb[1], R_20nb[1], R_19b[1], R_20b[1]]]
    Output1 += [["South", R_19nb[2] + R_19nb[4], R_20nb[2] + R_20nb[4], R_19b[2]
                             + R_19b[4], R_20b[2] + R_20b[4]]]
    Output1 += [["West", R_19nb[3], R_20nb[3], R_19b[3], R_20b[3]]]
    Output1 += [["Texas", R_19nb[4], R_20nb[4], R_19b[4], R_20b[4]]]


# Long-Distance Trip Frequency & COVID-19

# In[12]:
```

```
Output2 = [["",'Long-Distance Trip Frequency & COVID-19']]
Output2 += [["", outkey]]

covid_before_grp = df.groupby(['before_c'])
covid_during_grp = df.groupby(['during_c'])
covid_after_grp = df.groupby(['after_c'])

C_before =
                             covid_before_grp['Weights'].sum()/sum(covid_before
                             _grp['Weights'].sum())
C_during =
                             covid_during_grp['Weights'].sum()/sum(covid_during
                             _grp['Weights'].sum())
C_after = covid_after_grp['Weights'].sum()/sum(covid_after_grp['Weights'].sum())

covid_before_grp['Weights'].sum()

Output2 += [["","Before the COVID-19 pandemic", "During the COVID-19 pandemic",
                             "Plan after COVID-19 is no longer a concern"]]
Output2 += [["Every week", C_before[4],C_during[4],C_after[4]]]
Output2 += [["More than twice a month", C_before[6],C_during[6],C_after[6]]]
Output2 += [["More than once a month", C_before[5],C_during[5],C_after[5]]]
Output2 += [["Every month", C_before[3],C_during[3],C_after[3]]]
Output2 += [["Every 3 months", C_before[0],C_during[0],C_after[0]]]
Output2 += [["Every 6 months", C_before[1],C_during[1],C_after[1]]]
Output2 += [["Every 9 months", C_before[2],C_during[2],C_after[2]]]
Output2 += [["On average, less than 1 trip per year",
                             C_before[8],C_during[8],C_after[8]]]
Output2 += [["Never", C_before[7],C_during[7],C_after[7]]]

#Output2


# Long-Distance Trip Purposes

# In[13]:


Output3 = [["","Long-Distance Trip Purposes"]]
Output3 += [["", outkey]]

ListofStrings = ["To/From Work", "Work-Related Business", "Shopping", "Other
                             Family/Personal Business",
                 "School/Church", "Medical/Dental", "Visit Friends/Relatives",
                             "Other Social/Recreational",
                 "Other"]
ListColumns = ['before_c_p', 'during_c_p', 'after_c_p']
df_purpose_covid = pd.DataFrame(columns = ListColumns, index = ListofStrings)


for i in ListColumns:
    for j in ListofStrings:
        df_purpose_covid.loc[j][i] =
                             df[df[i].str.contains(j,na=False)]['Weights'].sum(
                             )
df_purpose_covid = df_purpose_covid/[df_purpose_covid['before_c_p'].sum(),
                             df_purpose_covid['during_c_p'].sum(),
                             df_purpose_covid['after_c_p'].sum()]
```

```
Output3 += [["","Before the COVID-19 pandemic", "During the COVID-19 pandemic",
                              "Plan after COVID-19 is no longer a concern"]]
Output3 += df_purpose_covid.reset_index().values.tolist()

#Output3


# Long-Distance Mode Choice

# In[14]:


Output4 = [["",'Revealed Long-Distance Mode Choice']]
Output4 += [["", outkey]]

short_nb_grp = df.groupby(['short_nb'])
short_b_grp = df.groupby(['short_b'])
long_nb_grp = df.groupby(['long_nb'])
long_b_grp = df.groupby(['long_b'])
#short_nb_grp['short_nb'].value_counts()

M_short_nb =
                              (short_nb_grp['Weights'].sum()/sum(short_nb_grp['W
                              eights'].sum())).rename('M_short_nb')
M_short_b =
                              (short_b_grp['Weights'].sum()/sum(short_b_grp['Wei
                              ghts'].sum())).rename('M_short_b')
M_long_nb =
                              (long_nb_grp['Weights'].sum()/sum(long_nb_grp['Wei
                              ghts'].sum())).rename('M_long_nb')
M_long_b =
                              (long_b_grp['Weights'].sum()/sum(long_b_grp['Weigh
                              ts'].sum())).rename('M_long_b')

df_mc = pd.concat([M_short_nb, M_short_b, M_long_nb, M_long_b], axis=1,
                              sort=True)
df_mc_ind = df_mc.reindex(['Airplane', 'Boat/Ship', 'Bus', 'Personal
                              car','Rental car','Train'])
df_mc_ind.sum(axis=0)
df_mc_ind/df_mc_ind.sum(axis=0)

Output4 += [["","Non-Business", "Non-Business", "Business", "Business"],
            ["between 75 and 500 miles", "longer than 500 miles", "between 75
                              and 500 miles", "longer than 500 miles"]]
Output4 += df_mc_ind.reset_index().values.tolist()
#Output4

#Output4 += [df.columns.tolist()] + df.reset_index().values.tolist()


# # Activities in AVs

# In[15]:


Output5 = [["",'Activities in AVs']]
Output5 += [["", outkey]]

Act1 = (df.groupby('Q3.03_1')['Weights'].sum()).rename('Talking to other
                              passengers')
```

```
Act2 = (df.groupby('Q3.03_2')['Weights'].sum()).rename('Making phone calls or
                              messaging')
Act3 = (df.groupby('Q3.03_3')['Weights'].sum()).rename('Emailing or browsing
                              internet')
Act4 = (df.groupby('Q3.03_4')['Weights'].sum()).rename('Using social media')
Act5 = (df.groupby('Q3.03_5')['Weights'].sum()).rename('Reading for leisure')
Act6 = (df.groupby('Q3.03_6')['Weights'].sum()).rename('Watching videos or
                              playing games')
Act7 = (df.groupby('Q3.03_7')['Weights'].sum()).rename('Listening to music')
Act8 = (df.groupby('Q3.03_8')['Weights'].sum()).rename('Eating or drinking')
Act9 = (df.groupby('Q3.03_9')['Weights'].sum()).rename('Sleeping or resting')
Act10 = (df.groupby('Q3.03_10')['Weights'].sum()).rename('Working or studying')
Act11 = (df.groupby('Q3.03_11')['Weights'].sum()).rename('Watching the roadway
                              or landscape')

df_act_con = pd.concat([Act1, Act2, Act3, Act4, Act5, Act6, Act7, Act8, Act9,
                              Act10, Act11], axis=1, sort=True)
df_act_ind = df_act_con.reindex(['Absolutely', 'More likely', 'Neutral', 'Less
                              likely','Unlikely'])
df_act = df_act_ind/df_act_ind.sum(axis=0)

Output5 += [["Likelihood"] + df_act.columns.tolist()] +
                              df_act.reset_index().values.tolist()
#Output5


# # Long-Distance Trip Revealed & Stated Preferences

# Trip Type

# In[16]:


Output6 = [["",'Trip Type']]
Output6 += [["", outkey]]

ttype =
                              df.groupby('trip_type')['Weights'].sum()/sum(df.gr
                              oupby('trip_type')['Weights'].sum())

df_ttype = ttype.to_frame().reset_index().values.tolist()

Output6 += df_ttype

#Output6


# Long-Distance Revealed Trip Purposes

# In[17]:


Output7 = [["","Long-Distance Revealed Trip Purposes"]]
Output7 += [["", outkey]]

ListofStrings = ["To visit customers", "To meet other colleagues", "To interview
                              or to be interviewed",
                 "For marketing and advertising purposes", "For consulting and
                              advising purposes",
                 "To attend a convention, conference, and/or seminar",
                 "For personal business",
```

```
                "For other business-related reasons",
                "To visit friends", "To visit family and/or to attend family-
                            related events",
                "For leisure and/or vacation purposes",
                "For recreational purposes",
                "For entertainment purposes",
                "For other non-business-related reasons"]

ListColumns = ['Q4.04']
df_tpurpose = pd.DataFrame(columns = ListColumns, index = ListofStrings)

for j in ListofStrings:
    df_tpurpose.loc[j]['Q4.04'] =
                            df[df['Q4.04'].str.contains(j,na=False)]['Weights'
                            ].sum()

df_tpurpose = df_tpurpose/TotalWeightSum

Output7 += [["","Trip Purpose"]]
Output7 += df_tpurpose.reset_index().values.tolist()

#Output7


# Travel Distance

# In[18]:


Output8 = [["","Travel Distance"]]
Output8 += [["", outkey]]

tdist =
                            df.groupby('Distance')['Weights'].sum()/sum(df.gro
                            upby('Distance')['Weights'].sum())

df_tdist = tdist.to_frame().reindex(["Between 75 and 500 miles.", "Between 500
                            and 1000 miles.", "More than 1000
                            miles."]).reset_index().values.tolist()

Output8 += df_tdist
#Output8


# Primary way of traveling

# In[19]:


Output9 = [["","Primary Way of Traveling"]]
Output9 += [["", outkey]]

twot =
                            df.groupby('wot')['Weights'].sum()/sum(df.groupby(
                            'wot')['Weights'].sum())

df_twot = twot.to_frame().reset_index().values.tolist()

Output9 += df_twot
#Output9
```

```
# Time Spent on Air Travel

# In[20]:


Output10 = [["",'Time Spent on Air Travel']]
Output10 += [["", outkey]]
ListofStrings = ["Time scheduling the trip to the airport (e.g., reserving a van
                        or calling Uber/Lyft, renting a car)",
                 "Time traveling to the airport (driving or being driven by
                        someone else)", "Time parking at the airport",
                 "Time spent going through airport security", "Time waiting at
                        the airport", "Airplane time",
                 "Time scheduling the trip from the airport (e.g., reserving a
                        van, calling Uber/Lyft, renting a car)",
                 "Time traveling from the airport to your destination (driving
                        or being driven by someone else)",
                 "Time parking at your destination", "Total"]
ListInputColumns =
                        ['schedule1','access','parking1','security','wait'
                        ,'onboard','schedule2','egress','parking2']
ListInputWColumns =
                        ['schedule1_w','access_w','parking1_w','security_w
                        ','wait_w','onboard_w','schedule2_w','egress_w','p
                        arking2_w']
ListColumns = ['Max','Mean']

df_airtime = pd.DataFrame(columns = ListColumns, index = ListofStrings)

for i in range(len(ListInputColumns)):
    df_airtime.loc[ListofStrings[i]]['Max'] = df[ListInputColumns[i]].max()
    df_airtime.loc[ListofStrings[i]]['Mean'] =
                        sum(df.groupby(ListInputColumns[i])[ListInputWColu
                        mns[i]].sum())/sum(df.groupby(ListInputColumns[i])
                        ['Weights'].sum())

df_airtime.loc['Total']['Max'] = sum(df_airtime[:-2]['Max'])
df_airtime.loc['Total']['Mean'] = sum(df_airtime[:-2]['Mean'])

Output10 += [["Minutes"] + df_airtime.columns.tolist()] +
                        df_airtime.reset_index().values.tolist()

#Output10


# Willingness to use AVs (travel time)

# In[21]:


Output11 = [["",'Willingness to use AVs (travel time)']]
Output11 += [["", outkey]]

ListofStrings = ['I do not know','Maybe','No','Yes']
ListColumns = ['No Change','10% increase','25% increase', '50% increase']
df_AV_p = pd.DataFrame(columns = ListColumns, index = ListofStrings)

AV_p_t0 =
                        (df.groupby('pre_t0')['Weights'].sum()/sum(df.grou
```

```
                             pby('pre_t0')['Weights'].sum())).rename('No
                             Change')
AV_p_t10 =
                             (df.groupby('pre_t10')['Weights'].sum()/sum(df.gro
                             upby('pre_t10')['Weights'].sum())).rename('10%
                             increase')
AV_p_t25 =
                             (df.groupby('pre_t25')['Weights'].sum()/sum(df.gro
                             upby('pre_t25')['Weights'].sum())).rename('25%
                             increase')
AV_p_t50 =
                             (df.groupby('pre_t50')['Weights'].sum()/sum(df.gro
                             upby('pre_t50')['Weights'].sum())).rename('50%
                             increase')

df_AV_p['No Change'] = AV_p_t0.values
df_AV_p['10% increase'] = AV_p_t10.values
df_AV_p['25% increase'] = AV_p_t25.values
df_AV_p['50% increase'] = AV_p_t50.values

df_AV_p_ind = df_AV_p.reindex(['Yes','Maybe','No','I do not know'])


Output11 += [[""] + df_AV_p_ind.columns.tolist()] +
                             df_AV_p_ind.reset_index().values.tolist()
#Output11


# Willingness to use AVs (travel cost)

# In[22]:


Output12 = [["",'Willingness to use AVs (travel cost)']]
Output12 += [["", outkey]]

AV_c_t50l =
                             (df.groupby('c_50less')['Weights'].sum()/sum(df.gr
                             oupby('c_50less')['Weights'].sum())).rename('50%
                             less')
AV_c_t25l =
                             (df.groupby('c_25less')['Weights'].sum()/sum(df.gr
                             oupby('c_25less')['Weights'].sum())).rename('25%
                             less')
AV_c_t0 =
                             (df.groupby('c_same')['Weights'].sum()/sum(df.grou
                             pby('c_same')['Weights'].sum())).rename('Same
                             cost')
AV_c_t25 =
                             (df.groupby('c_25more')['Weights'].sum()/sum(df.gr
                             oupby('c_25more')['Weights'].sum())).rename('25%
                             more')
AV_c_t50 =
                             (df.groupby('c_50more')['Weights'].sum()/sum(df.gr
                             oupby('c_50more')['Weights'].sum())).rename('50%
                             more')

df_AV_c = pd.concat([AV_c_t50l, AV_c_t25l, AV_c_t0, AV_c_t25, AV_c_t50], axis=1,
                             sort=True)
```

```
df_AV_c = df_AV_c.reindex(["Absolutely","More likely","Neutral","Less
                             likely","Unlikely"])

Output12 += [["Trip cost"] + df_AV_c.columns.tolist()] +
                             df_AV_c.reset_index().values.tolist()

#Output12


# Skipping/Adding stops en route

# In[23]:


Output13 = [["",'Any layovers, highway rest areas, restaurants, or
                             accommodations before reaching destination']]
Output13 += [["", outkey]]

df_stop =
                             df.groupby('stop')['Weights'].sum()/sum(df.groupby
                             ('stop')['Weights'].sum())

Output13 += df_stop.reset_index().values.tolist()


# Whether to include more stops

# In[24]:


Output14 = [["",'Whether to include more stops']]
Output14 += [["", outkey]]

incstop =
                             df.groupby('incstop')['Weights'].sum()/sum(df.grou
                             pby('stop')['Weights'].sum())

incstops = incstop.reindex(["Yes, I would include more stops on my way to my
                             destination.",
                    "I may consider including more stops.","No, I would not
                             include more stops.",
                    "I would have made fewer stops instead."])



df_incstop = pd.DataFrame(columns = ["incstop"], index = ["Will include more
                             stops","May include more stops",
                                     "Remain the same number of stops.","Will
                             have fewer stops instead."])

df_incstop["incstop"] = incstops.values

Output14 += df_incstop.reset_index().values.tolist()

# Output14


# Whether to change destination to a new place

# In[25]:
```

```
Output15 = [["","Whether to change destination to a new place "]]
Output15 += [["", outkey]]
newdest =
                             df.groupby('newdest')['Weights'].sum()/sum(df.grou
                             pby('newdest')['Weights'].sum())

newdest_ind = newdest.reindex(["Yes, I would have chosen a destination with a
                             further distance.",
                 "Yes, I would have chosen a destination with a shorter
                             distance.",
                 "I may have changed my destination choice.",
                 "No, I would not have changed my destination."])


df_newdest = pd.DataFrame(columns = ["newdest"], index = ["Yes, a further
                             place","Yes, a nearear place",
                                     "Maybe","No"])
df_newdest["newdest"] = newdest_ind.values

Output15 += df_newdest.reset_index().values.tolist()


# Long-distance Travel Duration

# In[26]:


Output16 = [["","Long-distance Travel Duration"]]
Output16 += [["", outkey]]

tripdur =
                             df.groupby('days')['Weights'].sum()/sum(df.groupby
                             ('days')['Weights'].sum())
# print(tripdur)

df_tripdur = pd.DataFrame(columns = ['tripdur'], index = [str(i) for i in
                             range(len(tripdur))])
df_tripdur['tripdur'] = float(0)

for i in tripdur.index.values:
    df_tripdur.at[str(i),'tripdur'] = tripdur[i]
# print(df_tripdur)

df_tripdur.loc['30','tripdur'] += sum(df_tripdur.iloc[30:,0])
df_tripdur = df_tripdur.rename({'30':'30+'}, axis='index').squeeze()

df_tripdur = df_tripdur.iloc[0:31]


# In[27]:


fig, ax = plt.subplots(figsize=(16,8))
width = 0.6
ax.set_xlabel('Travel Duration (days)', fontsize=14)
ax.set_ylabel('Percent (%)', fontsize=14)
ax.tick_params(axis='both', which='major', labelsize=12)
# plt.xticks(rotation=45)

ax.bar(df_tripdur.index.values.tolist(),df_tripdur*100, width)
```

```
Output16 += df_tripdur.reset_index().values.tolist()
# Output16


# Trip duration with AV

# In[28]:


Output17 = [["","Trip duration preference with AV"]]
Output17 += [["", outkey]]

filt = (df['av_days_p'] <= 20) & (df['av_days_n'] <= 20)
df_raw_avdur = df[filt]

df_avdur_ext =
                             df_raw_avdur.groupby('av_days_p')['Weights'].sum()
                             /sum(df_raw_avdur.groupby('av_days_p')['Weights'].
                             sum())

df_avdur_sho =
                             df_raw_avdur.groupby('av_days_n')['Weights'].sum()
                             /sum(df_raw_avdur.groupby('av_days_n')['Weights'].
                             sum())

Output17 += [["","Extend Stay", "Shorten Stay"],
             ["Yes", 1-df_avdur_ext[0], 1-df_avdur_sho[0]],
             ["No", df_avdur_ext[0], df_avdur_sho[0]]]
# Output17


# In[29]:


# This analysis is not useful
Output18 = [["","Trip duration with AV"]]
Output18 += [["", outkey]]
df['per_av_days_p'] = df['av_days_p']/df['days']
filt = (df['av_days_p'] > 0) | (df['av_days_n'] > 0)

df_raw_avdur = df[filt]


df_avdur_ext =
                             df_raw_avdur.groupby('per_av_days_p')['Weights'].s
                             um()/sum(df_raw_avdur.groupby('per_av_days_p')['We
                             ights'].sum())
df_avdur_sho =
                             df_raw_avdur.groupby('av_days_n')['Weights'].sum()
                             /sum(df_raw_avdur.groupby('av_days_n')['Weights'].
                             sum())

# df_avdur_ext.group(by=(df_avdur_ext.index * 1000) // 300) * 0.3)
df_avdur_ext = df_avdur_ext.to_frame().reset_index()[0:48]

# print(df_avdur_ext)
df_avdur_extdist = df_avdur_ext.groupby(by = (np.ceil(df_avdur_ext.per_av_days_p
                             * 10)).astype(int)/10)['Weights'].sum()

# df_avdur_extdist.plot.line()
```

```
# Tralve party

# In[30]:


Output18 = [["","Numebr of people traveling together"]]
Output18 += [["", outkey]]

with_family =
                             df.groupby('family')['Weights'].sum()/sum(df.group
                             by('family')['Weights'].sum())
with_friends =
                             df.groupby('friends')['Weights'].sum()/sum(df.grou
                             pby('friends')['Weights'].sum())
with_colleagues =
                             df.groupby('colleagues')['Weights'].sum()/sum(df.g
                             roupby('colleagues')['Weights'].sum())

with_family_mean =
                             sum(df.groupby('family')['family_w'].sum())/sum(df
                             .groupby('family')['Weights'].sum())
with_friends_mean =
                             sum(df.groupby('friends')['friends_w'].sum())/sum(
                             df.groupby('friends')['Weights'].sum())
with_colleagues_mean =
                             sum(df.groupby('colleagues')['colleagues_w'].sum()
                             )/sum(df.groupby('colleagues')['Weights'].sum())


Output18 += [["Family members", "Friends", "Colleagues and/or associates"],
             [with_family_mean, with_friends_mean, with_colleagues_mean],
            ["Without", with_family[0], with_friends[0], with_colleagues[0]],
             ["With", 1 - with_family[0], 1 - with_friends[0], 1 -
                             with_colleagues[0]]]
# Output18


# In[31]:


Output19 = [["","Numebr of people traveling together (size)"]]
Output19 += [["", outkey]]

df['partysize'] = df['family'] + df['friends'] + df['colleagues']
partysize =
                             df.groupby('partysize')['Weights'].sum()/sum(df.gr
                             oupby('partysize')['Weights'].sum())
partysize_ind = partysize.reset_index()
# combine 11+

partysize_ind.loc[11,'Weights'] = partysize_ind.loc[10:,'Weights'].sum()
partysize_new = partysize_ind[:12].copy()

partysize_new.loc[:,'partysize'] = (partysize_new.partysize.astype(int) +
                             1).astype(str)

partysize_new.loc[11,'partysize'] = '12+'
partysize_plot = partysize_new.set_index('partysize').squeeze()
```

```
# plot the figure
fig, ax = plt.subplots(figsize=(12,8))
width = 0.6
ax.set_xlabel('Travel Party Size', fontsize=14)
ax.set_ylabel('Percent (%)', fontsize=14)
ax.tick_params(axis='both', which='major', labelsize=12)

ax.bar(partysize_plot.index.values.tolist(),partysize_plot*100, width)

Output19 += partysize_plot.reset_index().values.tolist()

# Output19


# In[32]:


Output20 = [["","Numebr of children traveling together"]]
Output20 += [["", outkey]]


with_children =
                             df.groupby('childrenn')['Weights'].sum()/sum(df.gr
                             oupby('childrenn')['Weights'].sum())

Output20 += with_children.reset_index().values.tolist()

# Output20


# In[33]:


Output21 = [["",'Number of additional people when using AVs']]
Output21 += [["", outkey]]

df['partysize_more'] = df['family_more'] + df['friends_more'] +
                             df['colleagues_more']
partysize_more =
                             df.groupby('partysize_more')['Weights'].sum()/sum(
                             df.groupby('partysize_more')['Weights'].sum())
partysize_more_ind = partysize_more.reset_index()
# print(partysize_more_ind)

partysize_more_ind.loc[1,'Weights'] = sum(partysize_more_ind.loc[1:,'Weights'])

partysize_more_new = partysize_more_ind[:2].copy()
# print(partysize_more_new)
partysize_more_new.loc[:,'partysize_more'] =
                             (partysize_more_new.partysize_more.astype(int)).as
                             type(str)

partysize_more_new.loc[1,'partysize_more'] = '1+'
partysize_more_plot = partysize_more_new.set_index('partysize_more').squeeze()

# plot the figure
fig, ax = plt.subplots(figsize=(12,8))
width = 0.6
ax.set_xlabel('Additional Travel Party Size', fontsize=14)
ax.set_ylabel('Percent (%)', fontsize=14)
ax.tick_params(axis='both', which='major', labelsize=12)
```

```
ax.bar(partysize_more_plot.index.values.tolist(),partysize_more_plot*100, width)

Output21 += [["Number of people","Percentage"]]
Output21 += partysize_more_plot.reset_index().values.tolist()
# Output21


# In[34]:


Output22 = [["",'Whether to bring additional children when using AVs']]
Output22 += [["", outkey]]

with_add_children =
                             df.groupby('AddChildren')['Weights'].sum()/sum(df.
                             groupby('AddChildren')['Weights'].sum())


Output22 += [["","Percentage"],
            ["Yes", with_add_children["Yes, I would have taken my children with
                             me."]],
             ["Maybe", with_add_children["I may have taken my children with
                             me."]],
             ["No", with_add_children["No, I wouldn't have taken my children
                             with me."]]]

# Output22


# In[35]:


Output23 = [["","Reasons for using AVs to make long-distance trips"]]
Output23 += [["", outkey]]

ListofStrings = ["The safety offered by a self-driving car.",
                 "The convenience offered by a self-driving car so I can use my
                             travel for other activities instead of driving.",
                 "The convenience offered by a self-driving car so I can have a
                             good rest during long-distance driving.",
                 "The reliability of the self-driving vehicle.",
                 "The self-driving car’s ability to self-park.",
                 "Other"]

ListColumns = ['WithAV']
df_WithAV = pd.DataFrame(columns = ListColumns, index = ListofStrings)

for j in ListofStrings:
    df_WithAV.loc[j]['WithAV'] =
                             df[df['WithAV'].str.contains(j,na=False)]['Weights
                             '].sum()

df_WithAV = df_WithAV/TotalWeightSum

Output23 += df_WithAV.reset_index().values.tolist()
#Output23


# In[36]:
```

```
Output24 = [["","Reasons for not using AVs to make long-distance trips"]]
Output24 += [["", outkey]]

ListofStrings = ["I enjoy driving by myself.",
                 "I'm not yet confident in the safety benefits provided by self-
                            driving cars.",
                 "The potential for faulty software in a self-driving car
                            concerns me.",
                 "Confusion arising from the mixture of human-driven and self-
                            driven cars concerns me.",
                 "The possibility of being tracked in a self-driving car
                            concerns me.",
                 "Other"]

ListColumns = ['WithoutAV']
df_WithoutAV = pd.DataFrame(columns = ListColumns, index = ListofStrings)

for j in ListofStrings:
    df_WithoutAV.loc[j]['WithoutAV'] =
                            df[df['WithoutAV'].str.contains(j,na=False)]['Weig
                            hts'].sum()

df_WithoutAV = df_WithoutAV/TotalWeightSum

Output24 += df_WithoutAV.reset_index().values.tolist()
#Output24


# # Long-Distance Future Scenario Questions

# In[37]:


Output25 = [["","Stated Long-distance Mode Choice"]]
Output25 += [["", outkey]]

short_nb_stated_grp = df.groupby(['short_nb_stated'])
short_b_stated_grp = df.groupby(['short_b_stated'])
long_nb_stated_grp = df.groupby(['long_nb_stated'])
long_b_stated_grp = df.groupby(['long_b_stated'])
#short_nb_grp['short_nb'].value_counts()

M_short_nb_stated =
                            (short_nb_stated_grp['Weights'].sum()/sum(short_nb
                            _stated_grp['Weights'].sum())).rename('M_short_nb_
                            stated')
M_short_b_stated =
                            (short_b_stated_grp['Weights'].sum()/sum(short_b_s
                            tated_grp['Weights'].sum())).rename('M_short_b_sta
                            ted')
M_long_nb_stated =
                            (long_nb_stated_grp['Weights'].sum()/sum(long_nb_s
                            tated_grp['Weights'].sum())).rename('M_long_nb_sta
                            ted')
M_long_b_stated =
                            (long_b_stated_grp['Weights'].sum()/sum(long_b_sta
                            ted_grp['Weights'].sum())).rename('M_long_b_stated
                            ')
```

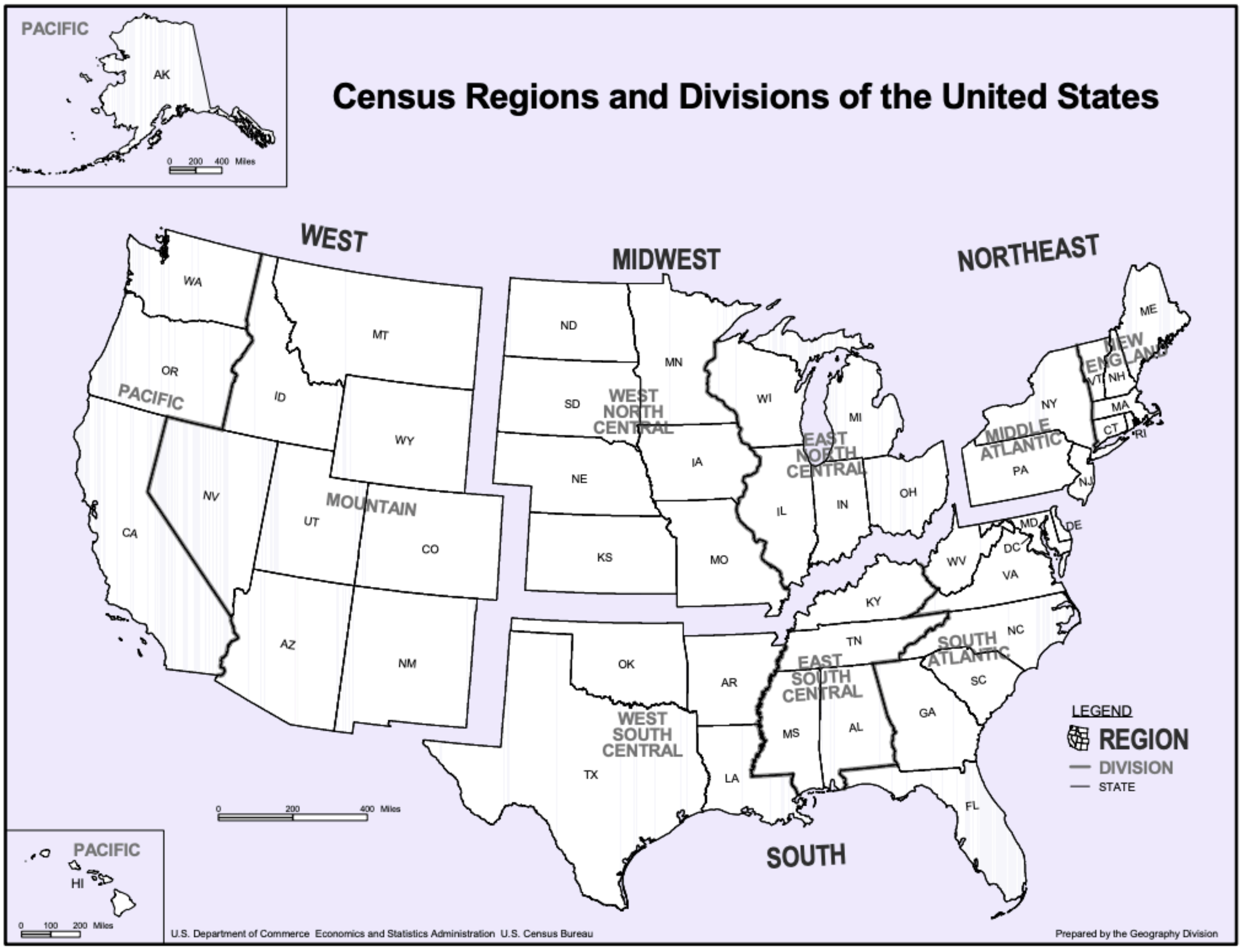
```
df_mc_stated = pd.concat([M_short_nb_stated, M_short_b_stated, M_long_nb_stated,
                             M_long_b_stated], axis=1, sort=True)
df_mc_stated_ind = df_mc_stated.reindex(['Personal self-driving car', 'Self-
                             driving rental car', 'Conventional car',
                                         'Conventional rental
                             car','Bus','Train','Airplane','Boat/Ship','Not
                             applicable'])
df_mc_stated_ind.sum(axis=0)
df_mc_stated_ind/df_mc_ind.sum(axis=0)

Output25 += [["","Non-Business", "Non-Business", "Business", "Business"],
            ["","between 75 and 500 miles", "longer than 500 miles", "between 75
                             and 500 miles", "longer than 500 miles"]]
Output25 += df_mc_stated_ind.reset_index().values.tolist()

# Output25


# In[38]:


Output26 = [["","Long-distance Trip-making with AVs: Frequency"]]
Output26 += [["", outkey]]

ld_frequency =
                             df.groupby('ld_fre')['Weights'].sum()/sum(df.group
                             by('ld_fre')['Weights'].sum())

# ld_frequency

Output26 += [["","Percentage"],
             ["Three times as many trips or more.", ld_frequency["I would make
                             three times as many trips or more."]],
             ["Twice as many trips per year.", ld_frequency["I would make twice
                             as many trips per year."]],
             ["50% more trips per year.", ld_frequency["I would make 50% more
                             trips per year."]],
             ["The same number of trips per year.", ld_frequency["I would make
                             about the same number of trips per year."]],
             ["Fewer trips per year.", ld_frequency["I would make fewer trips
                             per year."]],
             ["I do not know.", ld_frequency["I do not know."]]]

# Output26


# In[39]:


Output27 = [["","Long-distance Trip-making with AVs: Duration"]]
Output27 += [["", outkey]]

ld_duration =
                             df.groupby('ld_dur')['Weights'].sum()/sum(df.group
                             by('ld_dur')['Weights'].sum())

# print(ld_duration)

Output27 += [["","Percentage"],
             ["Three times as many days or more.", ld_duration["I would travel
                             for about three times as many days or more."]],
```

```
            ["Twice as many days.", ld_duration["I would travel for about twice
                            as many days."]],
            ["50% more days.", ld_duration["I would travel for about 50% more
                            days."]],
            ["The same number of days.", ld_duration["I would travel for about
                            the same number of days."]],
            ["Fewer days.", ld_duration["I would travel for fewer days."]],
            ["I do not know.", ld_duration["I do not know."]]]

# Output27


# In[40]:


Output28 = [["","Long-distance Trip-making with AVs: Distance"]]
Output28 += [["", outkey]]

ld_distance =
                            df.groupby('ld_dist')['Weights'].sum()/sum(df.grou
                            pby('ld_dist')['Weights'].sum())

# print(ld_distance)

Output28 += [["","Percentage"],
            ["More than doubled.", ld_distance["The travel distance would be
                            more than doubled."]],
            ["Doubled.", ld_distance["The travel distance would be doubled."]],
            ["About 50% more.", ld_distance["The travel distance would be about
                            50% more."]],
            ["No change in travel distance.", ld_distance["No change in travel
                            distance."]],
            ["Shorter.", ld_distance["The travel distance would be shorter."]]]

# Output28


# In[41]:


Output29 = [["","Long-distance Trip-making with AVs: Overnight Stay"]]
Output29 += [["", outkey]]

ld_overnightstay =
                            df.groupby('ld_overnight')['Weights'].sum()/sum(df
                            .groupby('ld_overnight')['Weights'].sum())

# print(ld_overnightstay)

Output29 += [["","Percentage"],
            ["I would stay in a hotel.", ld_overnightstay["Yes, I would stay in
                            a hotel."]],
            ["Maybe I would stay in a hotel.", ld_overnightstay["Maybe I would
                            stay in a hotel."]],
            ["In the self-driving car.", ld_overnightstay["No, I would prefer
                            to stay in the self-driving car."]],
            ["Not applicable.", ld_overnightstay["Other (please specify):"]]]

# Output29
```

```
# In[42]:


Output30 = [["","Departure Time for using AVs to make long-distance trips"]]
Output30 += [["", outkey]]

ListofStrings = ["Morning",
                 "Midday",
                 "Afternoon",
                 "Night"]

ListColumns = ['depttime']
df_departtime = pd.DataFrame(columns = ListColumns, index = ListofStrings)

for j in ListofStrings:
    df_departtime.loc[j]['depttime'] =
                             df[df['depttime'].str.contains(j,na=False)]['Weigh
                             ts'].sum()

df_departtime = df_departtime/TotalWeightSum

Output30 += [["Departure Time","Precentage"]]
Output30 += df_departtime.reset_index().values.tolist()

# Output30


# In[43]:


Output31 = [["","Willingness to share a ride in an AV for long-distance trips"]]
Output31 += [["", outkey]]

ld_SAV =
                             df.groupby('SAV')['Weights'].sum()/sum(df.groupby(
                             'SAV')['Weights'].sum())

ld_SAV = ld_SAV.reindex(["Absolutely","More likely","Neutral","Less
                             likely","Unlikely"])

Output31 += ld_SAV.reset_index().values.tolist()
# Output31


# # Print Output

# In[44]:


write2csv = [Output1, Output2, Output3, Output4, Output5, Output6, Output7,
                             Output8, Output9, Output10,
             Output11, Output12, Output13, Output14, Output15, Output16,
                             Output17, Output18, Output19, Output20,
             Output21, Output22, Output23, Output24, Output25, Output26,
                             Output27, Output28, Output29, Output30,
             Output31]

#Write to file
with open('Result_Tables_' + outkey + '.csv', 'w', newline='') as tableoutput:
    writer = csv.writer(tableoutput)
    for j in range(len(write2csv)):
```

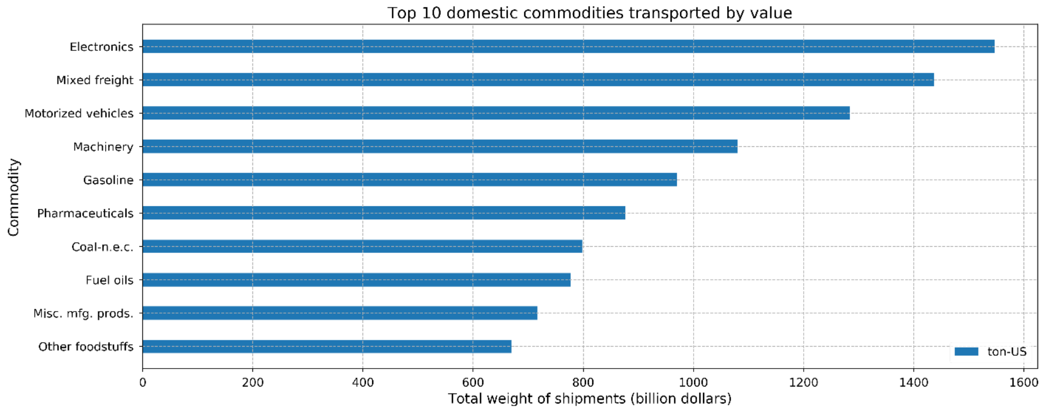
```
        writer.writerow([""])
        writer.writerow(["Table", j + 1])
        for i in range(len(write2csv[j])):
            writer.writerow(write2csv[j][i])
tableoutput.close()
```

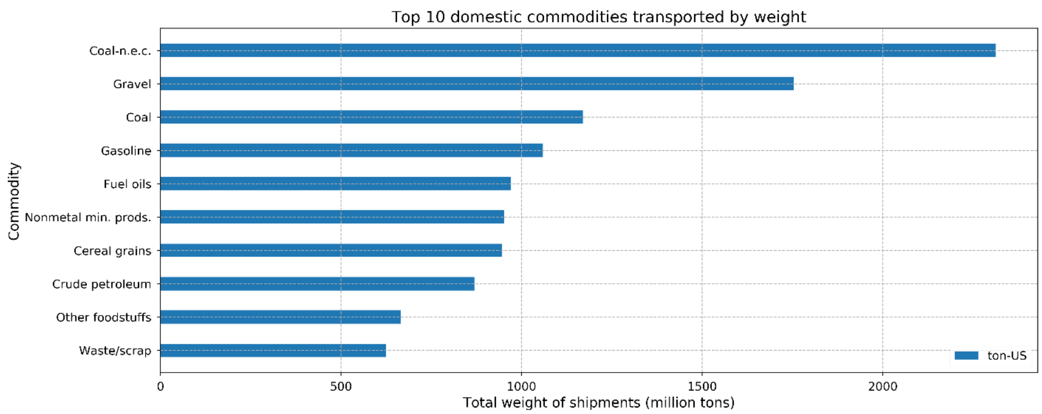
Appendix E: Census Regions and Divisions of the United States

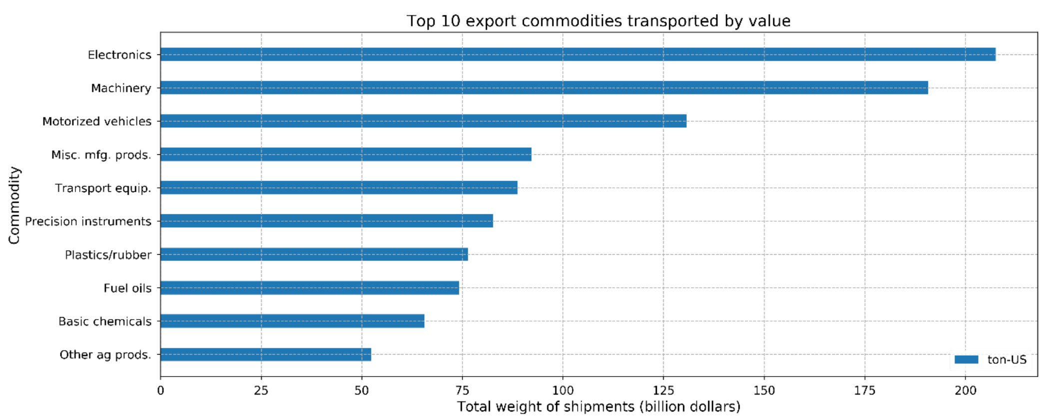


Source: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf

Appendix F: Top Freight Flow Commodity Rankings







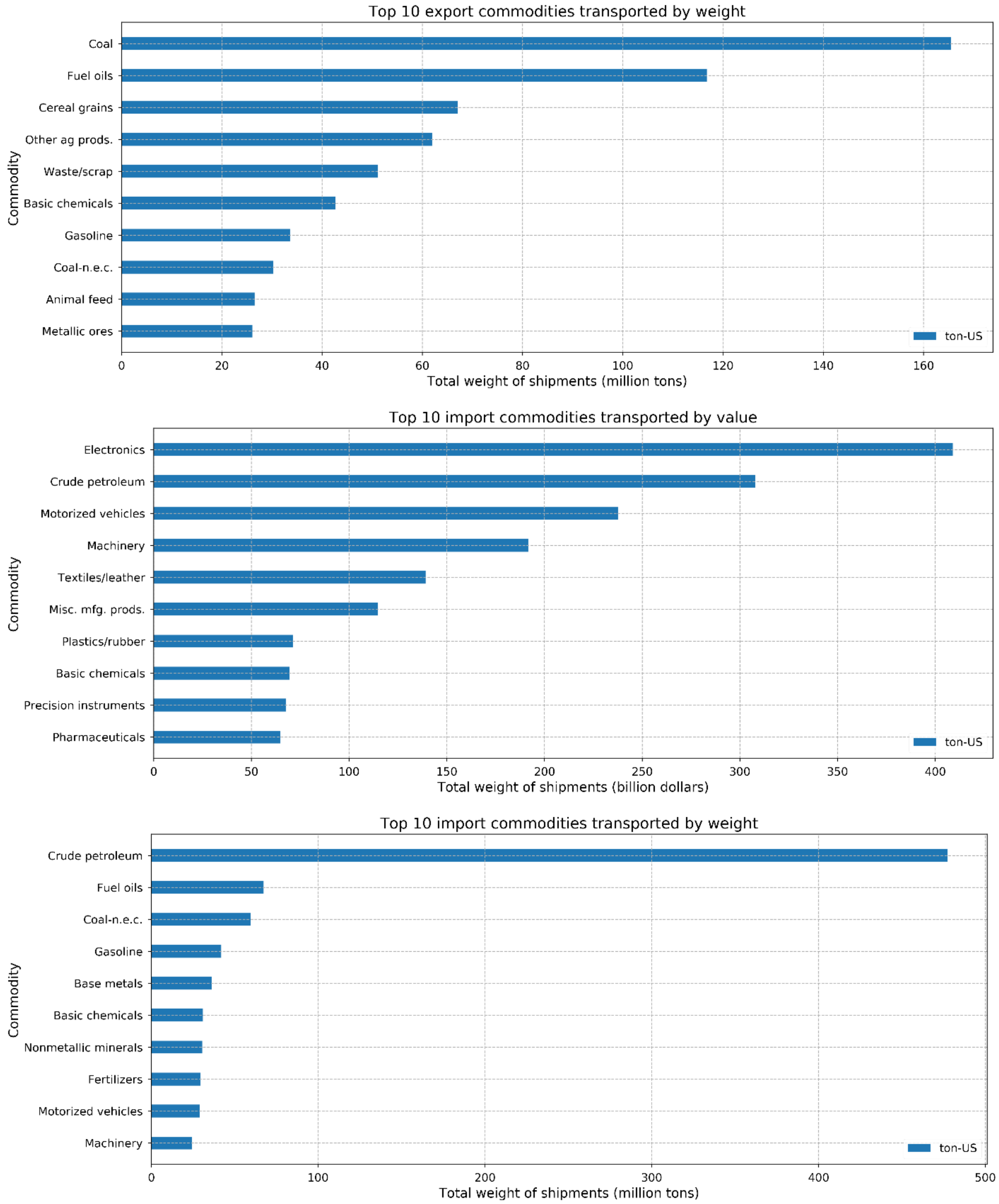
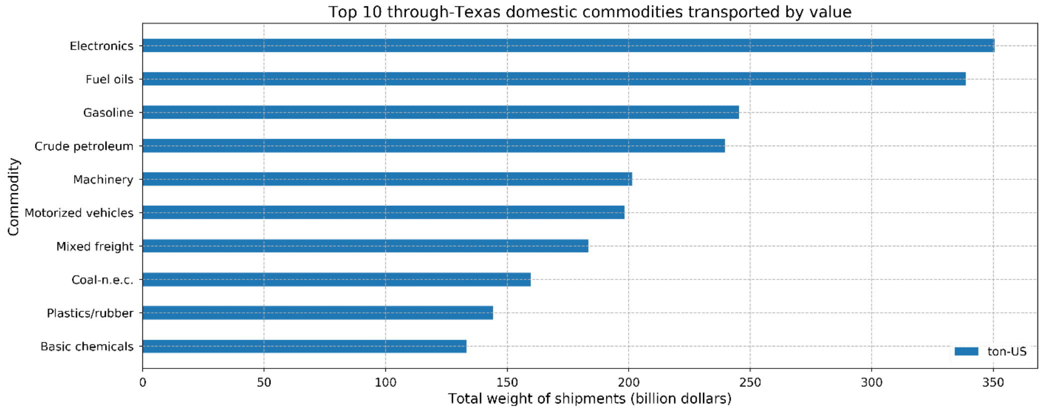
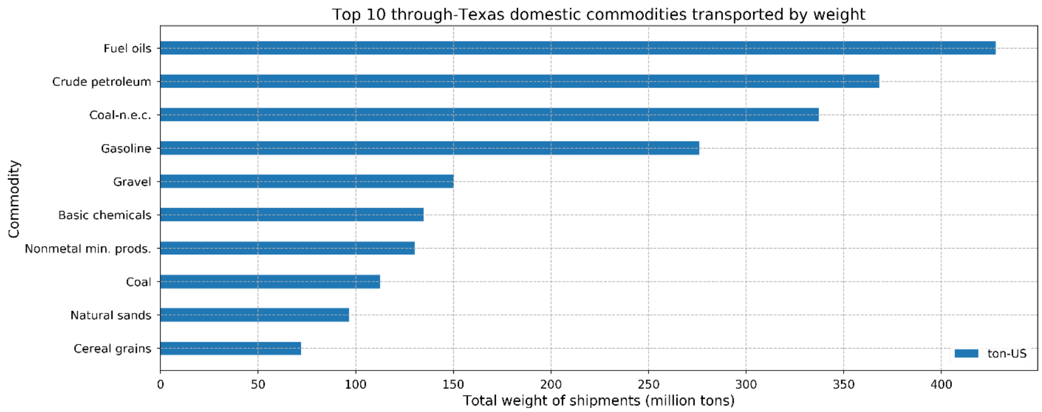
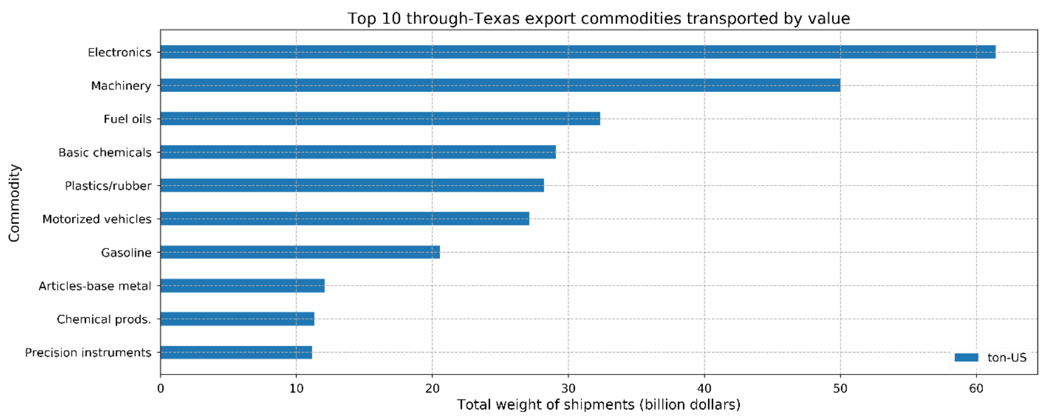
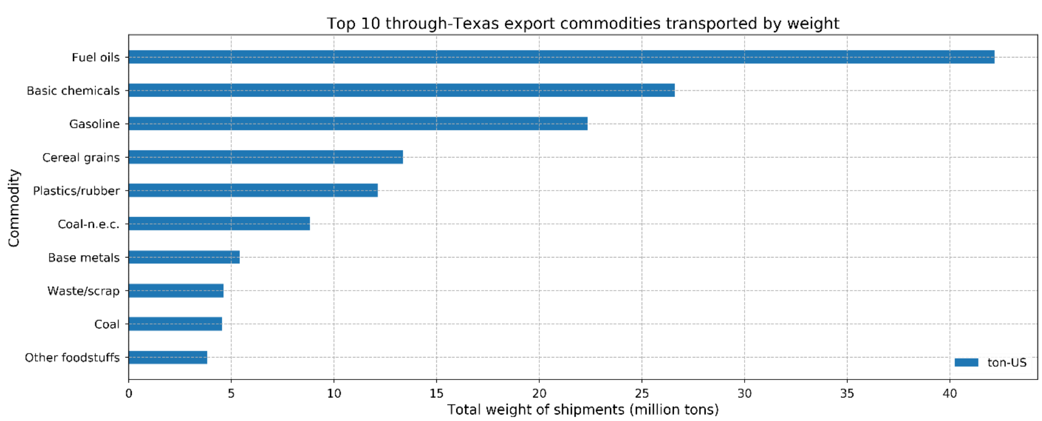


Figure E-1. US top 10 freight flow commodity rankings


Top 10 through-Texas domestic commodities transported by value
Commodity
Electronics
Fuel oils
Gasoline
Crude petroleum
Machinery
Motorized vehicles
Mixed freight
Coal-n.e.c.
Plastics/rubber
Basic chemicals
ton-US
0
50
100
150
200
250
300
350
Total weight of shipments (billion dollars)


Top 10 through-Texas domestic commodities transported by weight
Commodity
Fuel oils
Crude petroleum
Coal-n.e.c.
Gasoline
Gravel
Basic chemicals
Nonmetal min. prods.
Coal
Natural sands
Cereal grains
ton-US
0
50
100
150
200
250
300
350
400
Total weight of shipments (million tons)


Top 10 through-Texas export commodities transported by value
Commodity
Electronics
Machinery
Fuel oils
Basic chemicals
Plastics/rubber
Motorized vehicles
Gasoline
Articles-base metal
Chemical prods.
Precision instruments
ton-US
0
10
20
30
40
50
60
Total weight of shipments (billion dollars)


Top 10 through-Texas export commodities transported by weight
Commodity
Fuel oils
Basic chemicals
Gasoline
Cereal grains
Plastics/rubber
Coal-n.e.c.
Base metals
Waste/scrap
Coal
Other foodstuffs
ton-US
0
5
10
15
20
25
30
35
40
Total weight of shipments (million tons)

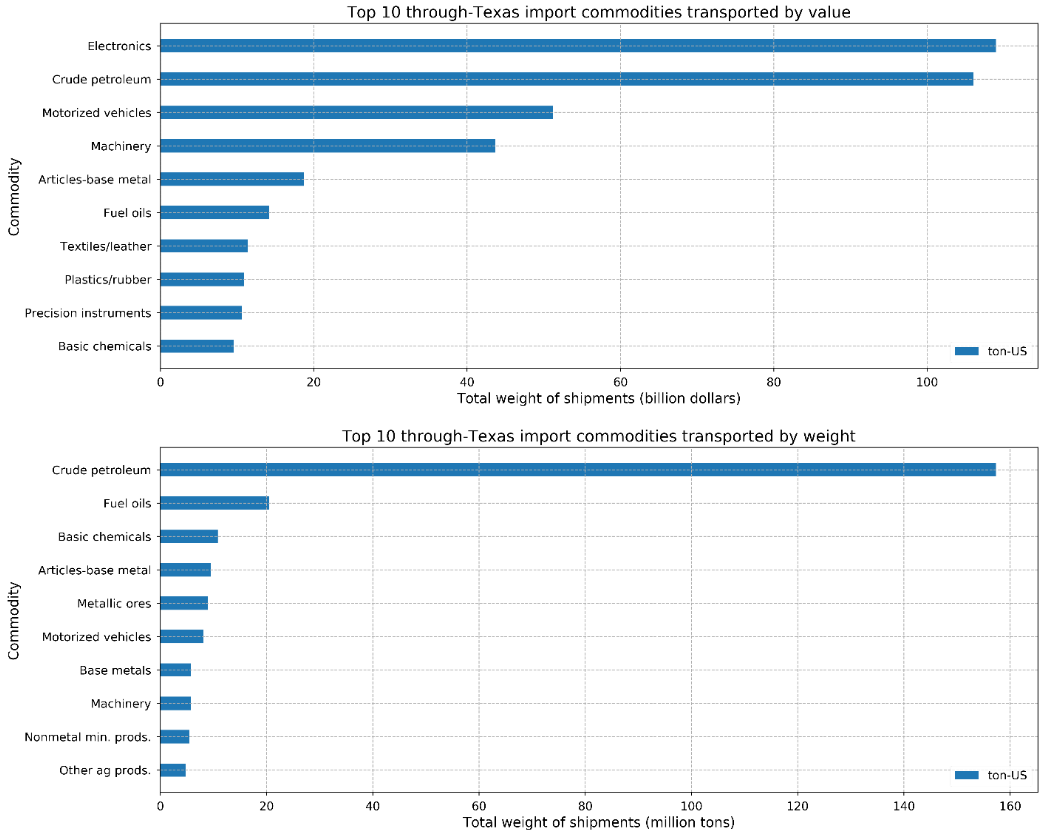
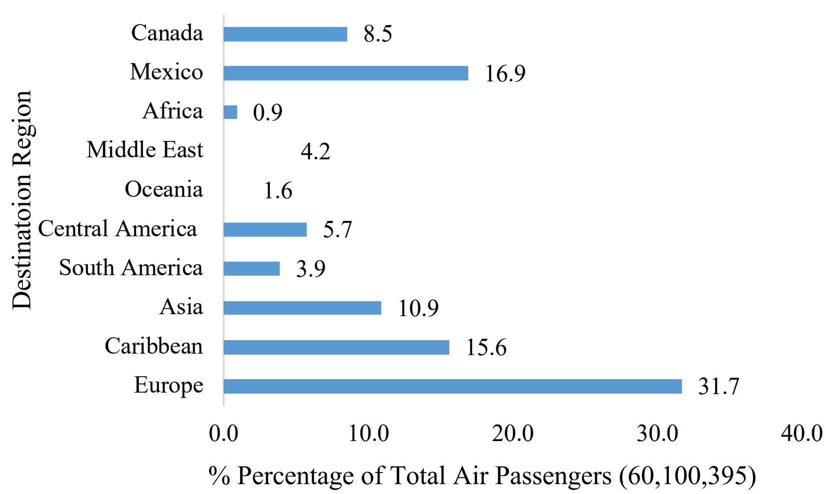


Figure E-2. Texas top 10 freight flow commodity rankings

Appendix G: International Trip Models

Data Sets Used

Using international travel datasets, this research examines the overseas destination preferences of Americans and models international travel demand to better prepare for future transportation needs. This study uses 2019 DB1B flight ticket data and the 2016/17 NHTS, as well as publicly available international travel data collected by the National Travel and Tourist Office (NTTO), Survey of International Air Travelers (SIAT), and Travel and Tourism Satellite Account (TTSA). According to the US Bureau of Transportation Statistics (BTS), US passenger-miles travelled in 2019 were 7.7 trillion, of which 4.5% were international person-miles involving US air carriers (BTS, 2022). The SIAT survey on US residents visiting overseas countries revealed that European (31.7%) and Caribbean (15.6%) countries accounted for the largest proportion of destinations after Canada and Mexico (together 25.4%) (SIAT,2019). Figure F- 1 Americans' outbound travel by air in 2019 (SIAT, 2019) show Americans' rate of air travel to different overseas regions in 2019.



F- 1 Americans' outbound travel by air in 2019 (SIAT, 2019)

The main data source in this study is the DB1B ticket data collected by BTS's Office of Airline Information. This data is a 10% random sample of US airline passenger tickets reported by US flag carriers only. It includes trip origin and destination data, yearly and quarterly indicators, number of passengers, number of legs, and distance and fare information for each itinerary. This dataset began publishing records in 1993, providing 28 years of available data. This study uses a 10% sample of the 2019 data (before the COVID-19 pandemic), which contains

2.6 million itineraries for 3.9 million passengers. Table 45 and Table 46 summarize the 2019 DB1B data's one-way itineraries to and from the US.

Table 45. Summary statistics for the 2019 DB1B round-trip air ticket data

	Mean	Median	Std dev	Max	Min
Quarter 1, $N = 246{,}168$					
Flight Fare per Itinerary ($)	953	635	1175	16427	0
Distance Flown (miles)	6669	5232	4313	26051	196
Fare per Mile ($)	0.171	0.127	0.16	2.918	0
Passengers	1.446	1	2.58	311	1
Segments	3.058	3	0.96	4	2
Quarter 2, $N = 318{,}033$					
Flight Fare per Itinerary ($)	1022	702	1151	17177	0
Distance Flown (miles)	7150	7298	4244	25870	196
Fare per Mile ($)	0.173	0.128	0.16	3.209	0
Passengers	1.414	1	2.57	427	1
Segments	3.041	3	0.96	4	2
Quarter 3, $N = 309{,}842$					
Flight Fare per Itinerary ($)	1033	733	1100	18491	0
Distance Flown (miles)	7318	7662	4167	26950	196
Fare per Mile ($)	0.171	0.128	0.15	2.883	0
Passengers	1.374	1	2.15	229	1
Segments	3.010	3	0.96	4	2
Quarter 4, $N = 174{,}532$					
Flight Fare per Itinerary ($)	1055	724	1226	17272	0
Distance Flown (miles)	6921	5331	4504	27338	196
Fare per Mile ($)	0.186	0.144	0.16	2.617	0
Passengers	1.307	1	2.37	322	1
Segments	3.327	4	0.90	4	2

Table 46. Summary statistics for the 2019 DB1B one-way air ticket data

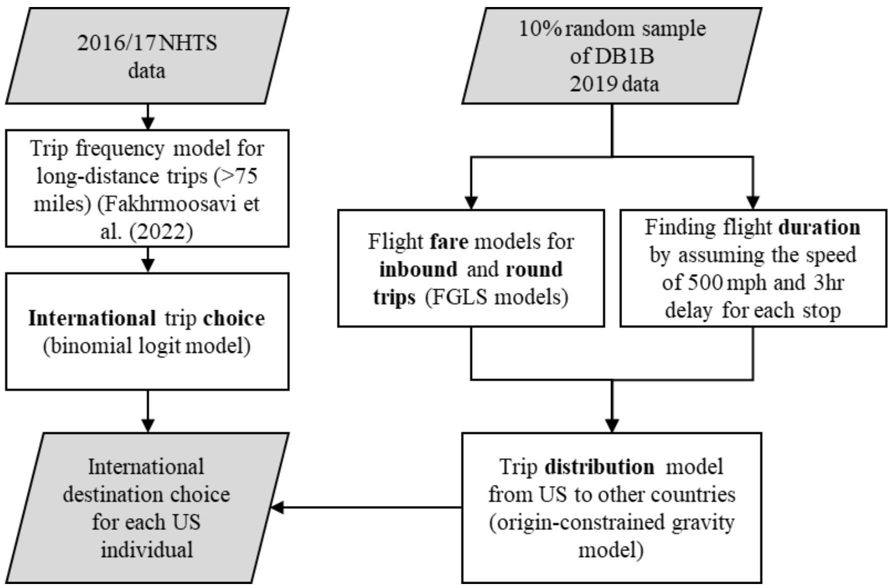
	Mean	Median	Min	Max	Std Dev
Quarter 1, $N = 371{,}334$					
Flight Fare per Itinerary ($)	494	304	0	11703	658.5
Distance Flown (miles)	3260	2129	98	21943	2621.6
Fare per Mile ($)	0.191	0.138	0	3.795	0.196
Passengers	1.539	2	1	368	3.571
Segments	1.899	2	1	4	0.648
Quarter 2, $N = 287{,}751$					
Flight Fare per Itinerary ($)	495	316	0	12743	637.7
Distance Flown (miles)	3329	2228	98	22833	2586.1
Fare per Mile ($)	0.194	0.143	0	3.867	0.196
Passengers	1.595	2	1	335	4.179
Segments	1.900	2	1	4	0.644
Quarter 3, $N = 221{,}507$					
Flight Fare per Itinerary ($)	534	342	0	11692	622.9
Distance Flown (miles)	3450	2306	98	20248	2656.4

	Mean	Median	Min	Max	Std Dev
Fare per Mile ($)	0.201	0.153	0	3.5	0.193
Passengers	1.524	2	1	440	3.714
Segments	1.913	2	1	4	0.650
Quarter 4, *N* = 167,983					
Flight Fare per Itinerary ($)	500	318	0	11477	642.4
Distance Flown (miles)	3306	2165	98	20754	2639.1
Fare per Mile ($)	0.197	0.144	0	3.469	0.191
Passengers	1.570	2	1	483	4.383
Segments	1.887	2	1	4	0.649

This study also uses the 2016/17 NHTS dataset to model Americans' trip-making choices of international versus domestic long-distance trips. The trip frequency model for long-distance trips (over 75 miles one-way) is estimated using this NHTS dataset. The 2016/17 NHTS data includes 923,572 trip records, which sum to 371 billion trips using NHTS expansion factors. In this dataset, 134.46 million expanded trips are reported as international trips, which account for only 1% of the total long-distance trips (~7 billion weighted). The population of 2019 destination nations, as well as information about the languages spoken in these countries, were collected from the United Nations website. If English is one of the languages spoken, the study considers the nation English-speaking. Additionally, each nation's major tourist attractions were obtained from the 2019 edition of Euromonitor International's city tourist arrivals research report that covers over 400 cities worldwide. (A tourist is defined as an international visitor who comes to another country for at least 24 hours and resides in paid or unpaid group or private lodging for a period not exceeding 12 months). (Top 100 City Destinations, 2019 Edition). Mexico and Canada accounted for 40% (39.9 million passengers) and 15% (14.9 million), respectively, of total outbound travel from the United States (99.7 million). The STATS Canada and Banco de Mexico websites were used to obtain data on Americans' international visits (staying one or more nights) to Canada and Mexico by land. Both records show that land travel to Mexico and Canada (39.6 million visitors total) accounts for a significant portion of overall outbound travel to these countries (54.9 million), accounting for 74.5% and 65.7% of total trips to the countries, respectively.

Models

Error! Reference source not found. illustrates the modeling framework for Americans' international trip distribution. The 2016/17 NHTS data is used to estimate Americans' international trip-making. The trip frequency model for long-distance trips is estimated using this NHTS dataset. Travelers' decision to take a long-distance international trip will be estimated using a binomial logit model. Trips are then distributed between each US origin airport and other countries' international airports using an origin-constrained gravity model and DB1B data. Flight duration and fare, English-language country indicator, tourism attraction country indicator, and population of the country are used as the inputs of this model. Flight duration is not provided in the DB1B data. Thus, it is estimated here based on an average speed and average delay for each stop. In addition, flight fares and their variation are estimated using FGLS models for US outbound and round.



F- 2 Modeling framework to predict destinations for international trips from US

Flight Fare and Duration Models

International Round Trips—US Origin

Due to the large sample size and unknown nature of heteroscedasticity, we employed feasible generalized least square models to predict the flight fare for international trips. FGLS models for 2019 round-trip itineraries (Table 47) indicate that a flight fare falls with inclusion of an intermediate stop relative to an uninterrupted trip and with an increase in the number of passengers on the itinerary. International round-trip airfares cost \$0.058 per mile on average for coach class and \$0.281 per mile for business class or higher. Trips taken from October to December are more expensive than those taken during other months of the year. Traveling to an English-speaking nation is less expensive than traveling to or from a non-English-speaking country if other variables are kept constant. Shifting all samples towards business or higher class and towards United Airlines increases the flight fare by 150% and 7.5%, respectively. Table 48 presents the model estimates when the log of linear model residuals is regressed on all dependent variables. The results show that flight prices of itineraries with more than one stop vary significantly compared to those without stops.

Table 47. FGLS model estimates for international round trips to and from US (DB1B, 2019)

Variable Name	Estimate	t-stat	P-value
(Intercept)	337.60	105.136	0.000
Distance flown (miles)	0.058	208.72	0.000
Distance flown (miles) *Business class or higher	0.281	108.32	0.000
Trip made during April through June	20.41	12.529	0.000
Trip made during July through September	18.54	11.364	0.000
Trip made during October through December	69.04	54.304	0.000
Restricted coach class	56.69	35.829	0.000
Business class or higher	−118.2	−5.649	0.000
#Passengers on the itinerary	−8.388	−63.497	0.000
ln(population of destination country)	−8.204	−36.758	0.000
Itinerary with 1 stop	−99.16	−59.481	0.000
Itinerary with 2 stops	−69.97	−67.127	0.000
Outbound trip	125.0	124.255	0.000
Destination is English speaking	−8.798	−5.588	0.000
Alaska Airlines	−53.25	−9.542	0.000
JetBlue Airlines	−15.52	−5.375	0.000
Delta Airlines	51.71	23.759	0.000
United Airlines	88.66	63.178	0.000
SkyWest Airlines	76.96	17.412	0.000
Endeavor Air	37.65	8.713	0.000
Canadian Pacific Airlines	14.90	2.92	0.004
Hawaiian Airlines	775.3	26.819	0.000
GoJet Airlines	−41.06	−11.981	0.000
Southwest Airlines	−159.9	−43.468	0.000
Spirit Airlines	143.6	47.993	0.000
Mesa Airlines	−17.45	−4.173	0.000
Republic Airlines	110.3	17.973	0.000
Eva Airlines	39.46	6.438	0.000
PSA Airlines	148.3	13.788	0.000
Frontier Airlines	−215.3	−10.231	0.000
Sun Country Airlines	−227.8	−3.188	0.001
Horizon Air	−5.249	−0.67	0.503
Distance flown (miles) *Destination is English speaking	−0.002	−6.894	0.000

Variable Name	Estimate	t-stat	P-value
Distance flown (miles) *Trip made during April through June	0.004	12.322	0.000
Distance flown (miles) *Trip made during July through September	0.006	18.248	0.000
Business class or higher *Destination is English speaking	131.2	6.995	0.000
Trip made during April through June *Business class or higher	−83.24	−4.228	0.000
Trip made during July through September *Business class or higher	−157.5	−7.571	0.000
Business class or higher *Alaska Airlines	−209.9	−3.888	0.000
Business class or higher *JetBlue Airlines	−524.0	−3.506	0.000
Business class or higher *Delta Airlines	−858.6	−38.11	0.000
Business class or higher *United Airlines	225.0	9.995	0.000
Business class or higher *SkyWest Airlines	−276.1	−4.241	0.000
Business class or higher *Endeavor Air	−578.5	−6.903	0.000
Business class or higher *Canadian Pacific Airlines	−393.6	−3.367	0.001
Business class or higher *Hawaiian Airlines	−588.2	−7.498	0.000
Business class or higher *Itinerary with 1 stop	−143.9	−4.367	0.000
Business class or higher *Itinerary with 2 stops	−318.5	−17.82	0.000
Distance flown (miles) *Alaska Airlines	−0.008	−5.609	0.000
Distance flown (miles) *JetBlue Airlines	0.009	13.158	0.000
Distance flown (miles) *Delta Airlines	0.014	39.903	0.000
Distance flown (miles) *Southwest Airlines	−0.011	−10.395	0.000
Distance flown (miles) *Spirit Airlines	−0.031	−28.813	0.000
Distance flown (miles) *SkyWest Airlines	0.017	15.968	0.000
Distance flown (miles) *Republic Airlines	0.027	26.092	0.000
Distance flown (miles) *Endeavor Air	0.029	25.92	0.000
Distance flown (miles) *Eva Airlines	0.027	15.114	0.000
Distance flown (miles) *PSA Airlines	0.013	7.713	0.000
Distance flown (miles) *Horizon Air	−0.010	−3.433	0.001
Distance flown (miles) *Hawaiian Airlines	−0.066	−18.948	0.000
Distance flown (miles) *GoJet Airlines	0.007	2.596	0.009
Distance flown (miles) *Frontier Airlines	−0.015	−2.116	0.034
Distance flown (miles) *Sun Country Airlines	0.047	2.255	0.024
Y: Fare ($) per paid Itinerary per passenger, N = 1,048,268, Adj. R^2: 0.3026			

Table 48. Variance model estimates for international round trips to and from US (DB1B, 2019)

Variable Name	Estimate	t-stat	P-value
(Intercept)	9.371	513.9	0.000
Distance flown (miles)	0.000	325.3	0.000

Variable Name	Estimate	t-stat	P-value
Trip made during April through June	−0.018	−2.97	0.003
Trip made during July through September	−0.023	−3.76	0.000
Trip made during October through November	0.253	34.45	0.000
Alaska Airlines	0.073	3.90	0.000
JetBlue Airlines	−1.017	−76.26	0.000
Delta Airlines	0.211	34.25	0.000
Southwest Airlines	−0.609	−48.24	0.000
United Airlines	0.116	17.41	0.000
Spirit Airlines	−0.970	−61.27	0.000
Mesa Airlines	0.110	5.89	0.000
SkyWest Airlines	0.156	9.90	0.000
Republic Airways	−0.081	−4.58	0.000
Endeavor Air	0.187	10.30	0.000
Canadian Pacific Air Lines	0.136	4.46	0.000
Eva Air	0.166	6.50	0.000
PSA Airlines	−0.056	−2.31	0.021
Horizon Air	0.428	11.37	0.000
Hawaiian Airlines	−0.129	−6.64	0.000
GoJet Airlines	0.262	6.05	0.000
Frontier Airlines	−0.993	−18.90	0.000
Sun Country Airline	1.336	21.93	0.000
Itinerary with 2 stops	−0.207	−23.03	0.000
Itinerary with 3 stops	−0.307	−53.60	0.000
Restricted coach class	−0.503	−63.38	0.000
Business class or higher	2.840	270.69	0.000
Outbound trip	0.351	62.39	0.000
Destination is English speaking	0.217	41.89	0.000
ln (population of destination country)	0.010	7.06	0.000
#Passengers on the itinerary	0.006	6.11	0.000
Y = log(Residuals^2) , N =1,048,268, Adj. R^2: 0.2947			

International One-Way Trips to and from US

The FGLS model results for air fares of international one-way trips to and from the US are shown in Table 49, and variance model estimates are shown in Table 50. The estimated model coefficients reveal that an average flight price per mile is \$0.078 in coach class and \$0.163 in business class or higher. The flight fare decreases as the number of passengers or number of stops on the itinerary increases. Trips made from April to June have high price variations compared to other months. Shifting the entire sample towards business or higher class trip increases the flight fare by 125%, while a shift towards Southwest Airlines decreases the cost by 58.5%.

Table 49. FGLS model estimates for international one-way trips to and from US (DB1B, 2019)

Variable Name	Estimate	t-stat	P-value
(Intercept)	320.0	154.7	0.000
Distance flown (miles)	0.078	179.0	0.000
Distance flown (miles)*Business class or higher	0.163	61.53	0.000
#Passengers on the Itinerary	−3.602	−70.87	0.000
Outbound Trip?	−34.20	−58.49	0.000
Restricted Coach Class	−7.743	−7.78	0.000
Business class or higher	−66.53	−2.40	0.016
Trip made during April to June	8.645	8.72	0.000
Trip made during July to September	1.907	1.76	0.079
Trip made during October to December	4.992	4.24	0.000
Itinerary with 1 stop	−40.23	−76.41	0.000
Itinerary with 2 stops	−23.69	−21.29	0.000
Itinerary with 3 stops	117.5	24.43	0.000
Destination is English speaking	−24.75	−27.51	0.000
Ln (Population of Destination Country)	−7.668	−53.18	0.000
Alaska Airlines	−29.55	−13.63	0.000
JetBlue Airlines	−42.76	−24.86	0.000
Delta Airlines	−11.08	−7.54	0.000
United Airlines	−16.53	−11.23	0.000
SkyWest Airlines	24.71	11.15	0.000
Canadian Pacific Airlines	−18.48	−13.53	0.000
Horizon Air	−4.890	−1.98	0.048
Hawaiian Air	253.2	16.08	0.000
SunCountry Airlines	−72.60	−12.45	0.000
Southwest Airlines	−7.897	−3.21	0.001
Spirit Airlines	−92.82	−53.99	0.000
Mesa Airlines	56.18	36.24	0.000
Republic Airlines	1.783	0.75	0.453
Endeavor Airlines	15.17	6.92	0.000
Eva Airlines	21.46	6.29	0.000
PSA Airlines	23.54	6.84	0.000
GoJet Airlines	60.15	12.90	0.000
Frontier Airlines	−118.5	−18.18	0.000
Distance flown (miles) *Trip made during April to June	−0.004	−7.81	0.000
Distance flown (miles) *Trip made during July to September	0.008	16.50	0.000

Variable Name	Estimate	t-stat	P-value
Distance flown (miles) *Trip made during October to December	−0.004	−6.92	0.000
Distance flown (miles) *Alaska Airlines	−0.008	−8.38	0.000
Distance flown (miles) *JetBlue	0.014	21.68	0.000
Distance flown (miles) *Delta Airlines	0.015	24.10	0.000
Distance flown (miles) *Southwest Airlines	−0.031	−21.67	0.000
Distance flown (miles) *United Airlines	0.018	35.57	0.000
Distance flown (miles) *Spirit Airlines	−0.032	−35.40	0.000
Distance flown (miles) *SkyWest Airlines	0.008	7.46	0.000
Distance flown (miles) *Republic Airline	0.017	12.07	0.000
Distance flown (miles) *Endeavor Airline	0.007	5.47	0.000
Distance flown (miles) *Eva Airline	0.032	15.65	0.000
Distance flown (miles) *PSA Airline	−0.009	−3.74	0.000
Distance flown (miles) *Horizon Air	−0.004	−2.54	0.011
Distance flown (miles) *Hawaiian Air	−0.039	−12.92	0.000
Distance flown (miles) *GoJet Airline	−0.005	−1.72	0.085
Distance flown (miles) *Frontier Airline	−0.012	−3.09	0.002
Business class or higher *Alaska Airlines	−123.4	−7.23	0.000
Business class or higher *JetBlue Airlines	505.0	12.18	0.000
Business class or higher *Delta Airlines	53.82	3.13	0.002
Business class or higher *United Airlines	−45.94	−4.17	0.000
Business class or higher *SkyWest Airlines	−43.62	−1.80	0.071
Business class or higher *Canadian Pacific Airlines	−95.83	−3.32	0.001
Business class or higher *Horizon Air	−68.84	−2.46	0.014
Business class or higher *Hawaiian Air	262.1	3.22	0.001
Business class or higher *SunCountry Airline	−360.0	−6.10	0.000
Business class or higher *Itinerary with 1 stop	−105.5	−11.57	0.000
Business class or higher *Itinerary with 2 stops	−353.4	−18.70	0.000
Business class or higher *Itinerary with 3 stops	−520.1	−7.41	0.000
Business class or higher *Destination is English speaking	52.91	5.56	0.000
Distance flown (miles) *Destination is English speaking	−0.010	−25.76	0.000
Business class or higher *Ln (Population of Destination Country)	11.86	4.73	0.000
Trip made during April to June *Business class or higher	−50.15	−4.93	0.000
Trip made during July to September *Business class or higher	−126.1	−9.99	0.000
Trip made during October to December *Business class or higher	−35.60	−2.69	0.007
Y: Fare ($) per paid Itinerary per passenger, N = 1,048,575, Adj. R^2: 0.2446			

Table 50. Variance model estimates for international one-way trips to and from US (DB1B, 2019)

Y = log(Residuals^2) , N = 1,048,575, Adj. R^2: 0.2896			
Variable Name	**Estimate**	**t-stat**	**P-value**

(Intercept)	9.880	628.2	0.000
Distance flown (miles)	0.000	364.8	0.000
#Passengers on the Itinerary	0.006	9.6	0.000
Itinerary with 2 stops	−0.203	−26.8	0.000
Itinerary with 3 stops	0.088	3.80	0.000
Restricted coach class	−1.037	−153.0	0.000
Business class or higher	1.959	192.8	0.000
Destination is English speaking	−0.214	−41.1	0.000
Ln (Population of Destination Country)	−0.046	−34.7	0.000
Trip made during April to June	−0.066	−13.0	0.000
Alaska Airlines	−0.174	−12.9	0.000
JetBlue Airlines	−1.324	−124.6	0.000
Delta Airlines	0.113	16.1	0.000
Southwest Airlines	−0.642	−48.4	0.000
United Airlines	−0.126	−18.7	0.000
Spirit Airlines	−1.351	−112.3	0.000
Mesa Airlines	−0.219	−13.4	0.000
SkyWest Airlines	−0.094	−6.9	0.000
Republic Airways	−0.119	−6.8	0.000
Endeavor Air	0.040	2.5	0.014
Canadian Pacific Air Lines	−0.446	−24.7	0.000
PSA Airlines	−0.162	−5.9	0.000
Horizon Air	−0.158	−7.5	0.000
Hawaiian Airlines	−0.228	−8.3	0.000
GoJet Airlines	0.164	5.1	0.000
Frontier Airlines	−1.145	−32.3	0.000

International Trip Choice

The specifications of the logistic regression model to estimate international trip choice for Americans are shown in Table 51. The model indicates that international trip frequency (per person) rises by about 16% with a 1 SD increase in the respondent's household income (i.e., $62,000). Increasing the summer trip and spring trip indicators by 1 SD also increases the frequency of international trips by 19% and 14%, respectively. International trips fall 23% when the female indicator increases by 1 SD and 31% when the full-time-employed indicator increases by 1 SD. Religious and personal business trips are also less likely to be international.

Table 51. Specifications of the logistic regression model for international versus domestic trips using the 2016/17 NHTS data

	Coefficient Estimates	t-Stat	P-Value	Practical Significance
(Intercept)	−5.594	−7.14	0.000	-
Household income ($1000)	0.006	1.63	0.103	0.161

Female	−1.067	−2.42	0.016	−0.228
Hispanic	1.424	2.67	0.008	0.148
White	1.114	2.27	0.023	0.159
Full-time employed	−1.501	−3.65	0.000	−0.315
Summer trip	0.988	1.78	0.075	0.193
Spring trip	0.907	1.68	0.094	0.140
Personal business trip	−1.066	−1.44	0.150	−0.104
Religious community trip	−14.232	−47.88	0.000	−0.869

R^2: 0.1344, N = 13,966

Trip Distribution Model

An origin-constrained gravity model was used to distribute trips among different origins and destinations. Gravity models in their traditional form consist of production, attraction (e.g., tourism attractions, population, and language of the destination), friction (i.e., travel time and/or fare), and a gravity constant term. A logarithmic operator was applied to form a log-linear gravity model, and an ordinary-least-squares (OLS) model was estimated to find the number of trips distributed between each origin and destination pair. Friction factor here is a function of impedance incorporating auto and air travel times and cost (i.e., flight fare, highway toll) normalized by value of time (VOT). VOT is assumed to be $30 per hour for air travelers and $20 per hour for auto users. Table 52 shows the specifications of this log-linear model as well as the practical significance of different statistically significant variables. This model was estimated using data from multiple sources indicating trip production for 334 major US airports and country attractions for 1,028 international airports in countries other than the US. There is a lack of data for origins and destinations of land travelers to Canada and Mexico. For the former, major airports in most touristic cities in Canadian provinces that are accessible from the US (e.g., Ontario, Quebec, British Columbia, Alberta, Nova Scotia) are considered as the destination locations. Origins are assumed to be the major airport of the closest state in the US. For Mexico, all trips are aggregated in one origin and destination pair from Texas to the state of Sinaloa. The trip distribution model indicated that trips to a foreign destination from an American origin fall 41% when the travel time increases by 7 hours or plane ticket cost rises by $210. Destinations hosting tourist attractions increase OD flow by 48%. The population and English-speaking indicators of the destination country are neither practically nor statistically significant.

Table 52. Specifications of the log-linear gravity model to estimate the number of trips between major US airports and other countries' airports

	Estimate	t-stat	P-Value	Practical Significance
(Intercept)	9.796	104.65	0.000	
Trip Production in Origin Airport	0.238	81.62	0.000	0.969
Travel Time & Cost	−1.578	124.11	0.000	0.409
Population of Destination Country	0.0013	0.50	0.616	0.0012
Tourism Indicator in Destination Country	0.907	51.60	0.000	0.136
English Speaking Country (Destination)	0.0024	0.17	0.864	0.0004

Conclusions

This research contributes to a better knowledge of Americans' overseas travel by estimating travel demand and expenses (in time and money) for trips between major US airports and international airports, as well as land trips to Mexico and Canada. The study uses 2019 DB1B aircraft ticket data, the 2016/17 NHTS, US 2019 outbound passenger travel aggregate estimates from NTTO, destination country characteristics from the UN, and major attraction data for tourists from Euromonitor's 2019 international report. The main data source of this study, 2019 DB1B provided by BTS, revealed that the flight fare for international travel falls as the number of passengers on the itinerary rises. Round trips made from October through December are more expensive than those taken during other months, while one-way trips made during April through June show high variation compared to other times of year. A round trip to an English-speaking nation is less expensive than traveling to or from a non-English-speaking country if other variables are kept constant. International round-trip airfares cost \$0.058 per mile on average for coach class and \$0.281 per mile for business class or higher. Shifting the entire sample towards business or higher class increases one-way flight fares by 125% and round-trip fares by 151%. The international trip choice model reveals that the probability of taking international trips rises 16% when household income is increased by 1 SD (i.e., \$62,000). Employment status, race, gender, trip season, and trip purpose are other significant variables affecting Americans' international trip choices. A log-linear model was used to distribute international trips among various major airports in the US and other countries. The trip distribution model indicated that travel time and cost and tourism attractions at the destination are the statistically significant variables affecting the number of trips to an international location. This model also suggested that trips to a foreign destination from an American airport fall 41% when the when the travel time element of the friction factor goes up 7 hours and increase by 48% the destinations are tourist attraction.

References

Aguilera, A., (2008). Business travel and mobile workers. *Transportation Research Part A: Policy and Practice*, 42(8), 1109–1116.

Alliance Transportation Group. (2018). *Documentation of statewide analysis model*, developed on behalf of the Texas Department of Transportation.

Arbib, J., & Seba, T. (2017). *Rethinking transportation 2020–2030: The disruption of transportation and the collapse of the internal-combustion vehicle and oil industries*. RethinkX. https://static1.squarespace.com/static/585c3439be65942f022bbf9b/t/591a2e4be6f2e1c13df930c5/1494888038959/RethinkX+Report_051517.pdf

Asgari, H., & Jin, X., (2019). Incorporating attitudinal factors to examine adoption of and willingness to pay for autonomous vehicles. *Transportation Research Record*, 2673(8), 418–429.

Ashe, A. (2022). US intermodal rail regains pricing edge against trucking. https://www.joc.com/rail-intermodal/intermodal-shipping/intermodal-rates-provide-shippers-glimmer-hope_20191008.html.

Aultman-Hall, L., (2018). Incorporating Long-Distance Travel into Transportation Planning in the United States. National Center for Sustainable Transportation, UC Davis. URL: https://escholarship.org/content/qt0ft8b3b5/qt0ft8b3b5.pdf

Aultman-Hall, L., Harvey, C., Sullivan, J., & LaMondia, J., (2016). The Impact of Long-Distance Tour Generation and Travel Attributes on Data Collection in the United States. *Transportation*, *43*(6). DOI 10.1007/s11116-016-9754-y

Aultman-Hall, L., Harvey, C., LaMondia, J. J., & Ritter, C., (2015). Design and response quality in a one-year longitudinal survey of overnight and long-distance travel. *Transportation Research Procedia*, 11, 136–153.

Axhausen, K.W., Youssefzadeh, M., Elvas, M.C., Erl, M.E., Herry, M., Lian, M.J.I., Armoogum, M.A.L., Madre, J.L., Jimmy, M., DEST, S.S., & Moritz, M.G., (1999). MEST–Methods for European Surveys of Travel Behaviour. Final Report, University of Innsbruck, Innsbruck. https://trimis.ec.europa.eu/project/methods-european-surveys-travel-behaviour

Bacon, B., & LaMondia, J.J., (2016). Typology of Travelers Based on Their Annual Intercity Travel Patterns Developed from 2013 Longitudinal Survey of Overnight Travel. *Transportation Research Record*, *2600*(1), 12–19.

Bansal, P., & Kockelman, K. M. (2017). Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies. *Transportation Research Part A: Policy and Practice, 95,* 49–63. https://doi.org/10.1016/j.tra.2016.10.013

Barth, M., Scora, G., & Younglove, T., (2004). Modal emissions model for heavy-duty diesel vehicles. *Transportation Research Record,* No. 1880, 10–20.

Beck, M.J., & Hensher, D.A., (2020). Insights into the impact of COVID-19 on household travel and activities in Australia–The early days of easing restrictions. *Transport Policy*.

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

Bernardin Jr, V. L., Ferdous, N., Sadrsadat, H., Trevino, S., & Chen, C. C., (2017). Integration of National Long-Distance Passenger Travel Demand Model with Tennessee Statewide Model and Calibration to Big Data. *Transportation Research Record*, 2653(1), 75–81.

Boerjan, P., (1995). Do we need a distinction between short and long holidays? *The Tourist Review.*

Bösch, P. M., Becker, F., Becker, H., & Axhausen, K. W. (2018). Cost-based analysis of autonomous mobility services. *Transport Policy*, *64*, 76–91. https://doi.org/10.1016/j.tranpol.2017.09.005

Burke, M., & Woolcock, G., (2013). Getting to the game: travel to sports stadia in the era of transit-oriented development. *Australian Sport* 64–83. Routledge.

Bureau of Transportation Statistics (BTS), (2003). Freedom to Travel. Publication #BTS03-08. US Department of Transportation Research and Innovative Technology Administration. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/freedom_to_travel/index.html.

Bureau of Transportation Statistics (BTS), (2017). Long Distance Transportation Patterns: Mode Choice. https://www.bts.gov/archive/publications/america_on_the_go/long_distance_transportation_patterns/entire

Bureau of Transportation Statistics (BTS), (2020). The Week in Transportation. https://www.bts.gov/covid-19/week-in-transportation

Cai, H., Fang, H., & Xu, L. C., (2011). Eat, drink, firms, government: An investigation of corruption from the entertainment and travel costs of Chinese firms. *The Journal of Law and Economics*, 54(1), 55–78.

Cambridge Systematics. (2002). *Florida intermodal statewide highway freight model: Technical memorandum task 3—Model specification*. Florida Department of Transportation.

Casley, S. V., Jardim, A. S., & Quartulli, A. M. (2013). *A study of public acceptance of autonomous cars* [Bachelor of Science thesis, Worcester Polytechnic Institute.] https://web.wpi.edu/Pubs/E-project/Available/E-project-043013-155601/unrestricted/A_Study_of_Public_Acceptance_of_Autonomous_Cars.pdf

Census Bureau. (2021). *Commodity flow survey (CFS).* https://www.census.gov/programs-surveys/cfs.html.

Childress, S., Nichols, B., Charlton, B., & Coe, S., (2015). Using an Activity-Based Model to Explore Possible Impacts of Automated Vehicles. Presented at the 94th Annual Meeting of the Transportation Research Board. Washington, DC, January 2015.

Cho, H. D. (2013). The factors that affect long-distance travel mode choice decisions and their implications for transportation policy. Ph.D. Dissertation, University of Florida.

Clements, L, & Kockelman, K., (2017). Economic Effects of Autonomous Vehicles. *Transportation Research Record.* No. 2602.

Collia, D. V., Sharp, J., & Giesbrecht, L., (2003). The 2001 national household travel survey: A look into the travel patterns of older Americans. *Journal of Safety Research*, 34(4), 461–470.

Das, S., Sekar, A., Chen, R., Kim, H. C., Wallington, T. J., Williams, E. (2017). Impacts of Autonomous Vehicles on Consumers Time-Use Patterns. *Challenges* 8, no. 2: 32. https://doi.org/10.3390/challe8020032

Erhardt, G. D., Freedman, J., Stryker, A., Fujioka, H., & Anderson, R. (2007). Ohio long-distance travel model. *Transportation Research Record*, 2003(1), 130-138.

Federal Aviation Agency Office of Environment and Energy, (2015). Aviation Emissions, Impacts, and Mitigation: A Primer. https://www.faa.gov/regulations_policies/policy_guidance/envir_policy/media/Primer_Jan2015.pdf

Federal Highway Administration. (2015). *Foundational knowledge to support a long-distance passenger travel demand modeling framework: Implementation report.* https://www.fhwa.dot.gov/policyinformation/analysisframework/docs/long-distance_model_implementation_report_final.pdf

Fagnant, D., & Kockelman, K.M., (2014). The Travel and Environmental Implications of Shared Autonomous Vehicles, Using Agent-Based Model Scenarios. *Transportation Research Part C* Vol (40): 1–13.

Fagnant, D., Kockelman, K.M., & Bansal, P., (2015). Operations of a Shared Autonomous Vehicle Fleet for the Austin, Texas Market. *Transportation Research Record* No. 2536: 98–106.

Fagnant, D. J., & Kockelman, K. M. (2016). Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas. *Transportation* 45: 1–16. https://doi.org/10.1007/s11116-016-9729-z

Fang Bao, Y., & Mckercher, B., (2008). The effect of distance on tourism in Hong Kong: A comparison of short haul and long haul visitors. *Asia Pacific Journal of Tourism Research*, 13(2), 101–111.

Federal Highway Administration (FHWA). (2015). Foundation Knowledge to Support a Long-Distance Passenger Travel Demand Modeling Framework: Implementation Report. https://www.fhwa.dot.gov/policyinformation/analysisframework/docs/long-distance_model_implementation_report_final.pdf

Federal Highway Administration, (2018). Main Study Retrieval Questionnaire. https://nhts.ornl.gov/assets/2016/NHTS_Retrieval_Instrument_20180228.pdf

Frei, A., Kuhnimhof, T.G., & Axhausen, K.W., (2010). Long distance travel in Europe today: Experiences with a new survey. *Arbeitsberichte Verkehrs-und Raumplanung*, *611*.

Geman, B., (2020). The push for post-coronavirus telework to help fight climate change. *Axios*. May 6. https://www.axios.com/telework-climate-change-post-coronavirus-a3f95c8b-5206-4f c8 -a6d9-79a963df3af8.html

Große, J., Fertner, C., & Carstensen, T. A., (2019). Compensatory leisure travel? The role of urban structure and lifestyle in weekend and holiday trips in Greater Copenhagen. *Case Studies on Transport Policy*, 7(1), 108–117.

Gurumurthy, K.M., & Kockelman, K.M., (2020). Modeling Americans' autonomous vehicle preferences: A focus on dynamic ride-sharing, privacy and long-distance mode choices. *Technological Forecasting and Social Change*, 150.

Gustafson, P., (2012). Managing business travel: Developments and dilemmas in corporate travel management. *Tourism Management*, 33(2), 276–284.

Guyot, K., & Sawhill, I. V., (2020). Telecommuting will likely continue long after the pandemic. *Brookings*. April 6. https://www.brookings.edu/blog/up-front/2020/04/06/telecommu ting-will-likely-continue-long-after-the-pandemic/

Harmon, T., Bahar, G., & Gross, F., (2018). Crash costs for highway safety analysis. FHWA-SA-17-071. https://safety.fhwa.dot.gov/hsip/docs/fhwasa17071.pdf

Harper, C.D., Hendrickson, C.T., Mangones, S., & Samaras, C., (2016). Estimating potential increases in travel with autonomous vehicles for the non-driving, elderly and people with travel-restrictive medical conditions. *Transportation Research C*, 72, 1–9.

Hess, S., Palma, D., & Hancock, T. (2021). *Apollo: Tools for choice model estimation and application*. https://cran.r-project.org/web/packages/apollo/apollo.pdf

Holz-Rau, C., Scheiner, J., & Sicks, K., (2014). Travel distances in daily travel and long-distance travel: what role is played by urban form? *Environment and Planning A*, 46(2), 488–507.

Hu, P.S., & Reuscher, T.R., (2004). Summary of travel trends: 2001 National Household Travel Survey. https://nhts.ornl.gov/2001/pub/STT.pdf

Huang, Y., Kockelman, K., & Quarles, N., (2020). How will self-driving vehicles affect US megaregion traffic? The case of the Texas triangle. *Research in Transportation Economics* 84, https://doi.org/10.1016/j.retrec.2020.101003.

Huang, Y., & Kockelman, K.M., (2019). What will autonomous trucking do to US trade flows? Application of the random-utility-based multi-regional input–output model. *Transportation*, 1–28.

Jackman, M., Lorde, T., Naitram, S., & Greenaway, T., (2020). Distance matters: the impact of physical and relative distance on pleasure tourists' length of stay in Barbados. *Annals of Tourism Research*, 80, 102794.

Insurance Institute for Highway Safety (IIHS) & Highway Loss Data Institute (HLDI). (2020, June 4). *Self-driving vehicles could struggle to eliminate most crashes.* https://www.iihs.org/news/detail/self-driving-vehicles-could-struggle-to-eliminate-most-crashes

Jiang, Y., Zhang, J., Wang, Y., & Wang, W., (2019). Capturing ownership behavior of autonomous vehicles in Japan based on a stated preference survey and a mixed logit model with repeated choices. *International Journal of Sustainable Transportation*, 13(10), 788–801.

Johnston, C., & Walker, J. (2017). *Peak car ownership: The market opportunity for electric automated mobility services*. Rocky Mountain Institute. http://rmi.org/wp-content/uploads/2017/03/Mobility_PeakCarOwnership_Report2017.pdf

Kim, S.H., Mokhtarian, P.L., & Circella, G., (2020). How, and for whom, will activity patterns be modified by self-driving cars? Expectations from the state of Georgia. *Transportation Research Part F: Traffic Psychology and Behaviour* 70, 68–80.

LaMondia, J.J., Moore, M., & Aultman-Hall, L., (2016a). Matching Household Life-Cycle Characteristics to Clustered Annual Schedules of Long-Distance and Overnight Travel. *Transportation Research Record* 2594(1), 11–17.

LaMondia, J. J., Fagnant, D. J., Qu, H., Barrett, J., & Kockelman, K., (2016b). Shifts in long-distance travel mode due to automated vehicles: Statewide mode-shift simulation experiment and travel survey analysis. *Transportation Research Record* 2566(1), 1–11.

Lavieri, P.S., Garikapati, V.M., Bhat, C.R., Pendyala, R.M., Astroza, S., & Dias, F.F., (2017). Modeling individual preferences for ownership and sharing of autonomous vehicle technologies. *Transportation Research Record*, *2665*(1), 1–10.

Lee, K.-H., Alexander, A. C., & Kim, D.-Y., (2014). A study of geographical distance groups on length of visitors' stay at local food festival destinations. *Journal of Vacation Marketing*, 20(2), 125–136.

Lee, J., & Kockelman, K., (2019). Energy and Emissions Implications of Self-Driving Vehicles. Presented at the 98th Annual Meeting of the TRB and under review for publication in *Energy Policy*.

Lee, J., Lee, D., Park, Y., Lee, S. & Ha, T., (2019). Autonomous vehicles can be shared, but a feeling of ownership is important: Examination of the influential factors for intention to use autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *107*, 411–422.

Lehto, X. Y., O'Leary, J. T., & Morrison, A. M., (2002). Do psychographics influence vacation destination choices? A comparison of British travellers to North America, Asia and Oceania. *Journal of Vacation Marketing*, 8(2), 109–125.

Le-Klähn, D.T., Zhang, X., Ponce-Lopez, R., Shaw, J., Zegras, P.C., & Ferreira Jr, J., (2019). Car-lite impacts on housing market and vehicle ownership. *Transportation Research Procedia*, 41.

Lenz, B., Fraedrich, E.M., & Cyganski, R. (2016). Riding in an autonomous car: What about the user perspective? https://www.researchgate.net/publication/306118664_Riding_in_an_Autonomous_Car_What_About_the_User_Perspective

Li, R., Kockelman, K. M., & Lee, J. (2022). Reducing Greenhouse Gas Emissions from Long-Distance Travel Business: How Far Can We Go?. *Transportation Research Record*, 2676(1), 472-486.

Litman, T. (2018). *Autonomous vehicle implementation predictions: Implications for transport planning*. Victoria Transport Policy Institute. https://www.vtpi.org/avip.pdf.

Liu, J., Kockelman, K.M., Boesch, P.M., & Ciari, F., (2017). Tracking a system of shared autonomous vehicles across the Austin, Texas network using agent-based simulation. *Transportation* 44, 1261–1278.

Lo, A., & Lam, T., (2004). Long-haul and short-haul outbound all-inclusive package tours. *Asia Pacific Journal of Tourism Research*, 9(2), 161–176.

Maeng, K., & Cho, Y., (2022). Who will want to use shared autonomous vehicle service and how much? A consumer experiment in South Korea. *Travel Behaviour and Society*, *26*, 9–17.

Martinez, L., & Crist, P., (2015). Urban Mobility System Upgrade: How shared self-driving cars could change city traffic. *International Transport Forum, Paris*.

McGuckin, N., (2018). Analysis Brief: Can We Use the NHTS to Estimate Long-Distance Travel? https://doi.org/10.13140/RG.2.2.23563.13607

McGuckin, N., Casas, J., & Wilaby, M., (2016). MI Travel Counts III Travel Characteristics Technical Report. *Michigan Department of Transportation. Lansing MI: Michigan Department of Transportation*. https://www.michigan.gov/documents/mdot/MTC_III_Travel_Characteristics_Report_554341_7.pdf

McKercher, B., Chan, A., & Lam, C., (2008). The impact of distance on international tourist movements. *Journal of Travel Research*, 47(2), 208–224.

McKercher, B., & Du Cros, H., (2003). Testing a cultural tourism typology. *International Journal of Tourism Research*, 5(1), 45–58.

Meyer, G., & Deix, S., (2014). Research and innovation for automated Driving in Germany and Europe. Part II: Industrial Research and Innovation in *Road Vehicle Automation*, 71–81. Springer: Berlin.

Mokhtarian, P., & Grossman, A., (2020). The Adoption and Travel Impacts of Teleworking: Will it be Different This Time? Presented at webinar held by Eno Center for Transportation. https://www.enotrans.org/event/webinar-the-adoption-and-travel-impacts-of-teleworking-will-it-be-different-this-time/

Molloy, J., Tchervenkov, C., Schatzmann, T., Schoeman, B., Hintermann, B., & Axhausen, K W., (2020). MOBIS-COVID19/19 Results as of 24/08/2020 (post-lockdown). https://doi.org/10.3929/ethz-b-000432671

National Academies of Sciences, Engineering, and Medicine. (2019). *Air demand in a dynamic competitive context with the automobile.* https://onlinepubs.trb.org/onlinepubs/acrp/acrp_rpt_204.pdf

National Private Truck Council. (2021). *Want to optimize your fleet? Know your average trucking cost per mile*. Paragon. https://www.paragonrouting.com/en-us/blog/post/want-optimize-your-fleet-know-your-average-trucking-cost-mile/

National Renewable Energy Laboratory (NREL), (2020). Cleansed Data from Transportation Studies and Surveys. https://www.nrel.gov/transportation/secure-transportation-data/tsdc-cleansed-data.html

National Safety Council. (2020). Guide to calculating costs: Costs of motor-vehicle injuries. https://injuryfacts.nsc.org/all-injuries/costs/guide-to-calculating-costs/data-details/

Nicolau, J. L., Zach, F. J., & Tussyadiah, I. P., (2018). Effects of distance and first-time visitation on tourists' length of stay. *Journal of Hospitality & Tourism Research*, 42(7), 1023–1038.

Nordström, M., & Engholm, A., (2021). The complexity of value of travel time for self-driving vehicles–a morphological analysis. *Transportation Planning and Technology* 44, no. 4: 400–417.

OAG Aviation Worldwide, (2020). How and When Will Aviation Recover from Covid-19? https://www.oag.com/coronavirus-airline-schedules-data

Oxford Economics, (2020, April 15). The Impact of COVID-19 on the United States Travel Economy. *Oxford Economics*. https://www.ustravel.org/sites/default/files/media_root/document/Coronavirus2020_Impacts_April15.pdf

Pawar, D.S., Yadav, A.K., Akolekar, N., & Velaga, N.R., (2020). Impact of Physical Distancing due to Novel Coronavirus (SARS-CoV-2) on Daily Travel for Work during Transition to Lockdown. *Transportation Research Interdisciplinary Perspectives*, 100203.

Pendyala, R. M., Konduri, K. C., & Christian, K. P. (2011). *PopGen 1.1 User's Manual.* Lulu.com Publishers, USA.

Perrine, K. A., Kockelman, K. M., & Huang, Y. (2020). Anticipating long-distance travel shifts due to self-driving vehicles. *Journal of Transport Geography*, 82, 102547. https://doi.org/10.1016/j.jtrangeo.2019.102547

Quarles, N., & Kockelman, K.M., (2018, January). Americans' plans for acquiring and using electric, shared and self-driving vehicles. In 97th Annual Meeting of the Transportation Research Board, Washington, DC.

Quarles, N., Kockelman, K. M., & Lee, J., (2021). America's fleet evolution in an automated future. *Research in Transportation Economics*: 101–107. https://doi.org/10.1016/j.retrec.2021.101107

Rannard, G., (2020). Coronavirus: Challenge of reshaping UK cities after lockdown. https://www.bbc.com/news/uk-52524807

Reichert, A., & Holz-Rau, C., (2015). Mode use in long-distance travel. *Journal of Transport and Land Use*, 8(2), 87–105.

Rohr, C., Fox, J., Daly, A., Patruni, B., Patil, S., & Tsang, F., 2013. Modeling long-distance travel in Great Britain. *Transportation Research Record*, 2344(1), 144–151.

Roth, S., Dai, Y., DeMatteis, J., (2017). 2017 NHTS Weighting Report. https://nhts.ornl.gov/assets/2017%20NHTS%20Weighting%20Report.pdf

Sadik-Khan, J., & Solomonow, S., (2020, June 14). Fear of Public Transit Got Ahead of the Evidence. *The Atlantic*. https://www.theatlantic.com/ideas/archive/2020/06/fear-transit-bad-cities/612979/

Saeed, T.U., Burris, M.W., Labi, S., & Sinha, K.C., (2020). An empirical discourse on forecasting the use of autonomous vehicles using consumers' preferences. *Technological Forecasting and Social Change*, *158*, 120130.

Sandow, E., & Westin, K., (2010). The persevering commuter: Duration of long-distance commuting. *Transportation Research Part A: Policy and Practice*, 44(6), 433–445.

Santos, A., McGuckin, N., Nakamoto, H.Y., Gray, D., & Liss, S. (2011). *Summary of travel trends: 2009 national household travel survey.* (No. FHWA-PL-ll-022). National Household Travel Survey. https://nhts.ornl.gov/tables09/fatcat/2009/avo_TRPTRANS_WHYTRP1S.html

Schaeffer, M. S., (2009). Putting expense fraud under the microscope. *Financial Executive*, 25(10), 42–45.

Scheiner, J., (2010). Interrelations between travel mode choice and trip distance: trends in Germany 1976–2002. *Journal of Transport Geography*, 18(1), 75–84.

Shladover, S.E., Lu, X.Y., Song, B., Dickey, S., Nowakowski, C., Howell, A., Bu, F., Marco, D., Tan, H.S., & Nelson, D., (2006). Demonstration of automated heavy-duty vehicles. *California Partners for Advanced Transit and Highways (PATH)*.

Schoettle, B., & Sivak, M. (2014). *A survey of public opinion about autonomous and self-driving vehicles in the US, the UK, and Australia*. University of Michigan Transportation Research Institute. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/108384/103024.pdf

Schrank, D., Albert, L., Eisele, B., & Lomax, T., (2021). 2021 urban mobility report, Texas A&M Transportation Institute.

Silva, D., Földes, D., & Csiszár, C., (2021). Autonomous Vehicle Use and Urban Space Transformation: A Scenario Building and Analysing Method. *Sustainability* 13, no. 6: 3008. https://doi.org/10.3390/su13063008

Simoni, M., & Kockelman, K., (2021). Vehicles that drive themselves: What to expect with autonomous vehicles. *International Encyclopedia of Transportation.* 19–25. https://doi.org/10.1016/b978-0-08-102671-7.10400-2.

Singleton, P. A. (2019). Discussing the "positive utilities" of autonomous vehicles: Will travellers really use their time productively? *Transport Reviews,* 39(1), 50–65. https://doi.org/10.1080/01441647.2018.1470584

Sullivan, J., LaMondia, J. J., Harvey, C., Garrison, C., & Aultman-Hall, L., (2016). Analysis of the Long-Distance and Overnight Travel Tour Planning Process. *Transportation Research Record*, 2594(1), 1–10.

Texas Department of Transportation (TxDOT). (2021). Comparison of motor vehicle traffic deaths, vehicle miles, death rates, and economic loss 2003–2021. https://ftp.txdot.gov/pub/txdot-info/trf/crash_statistics/2021/a.pdf

Tideswell, C., & Faulkner, B., (1999). Multidestination travel patterns of international visitors to Queensland. *Journal of Travel Research*, 37(4), 364–374.

US Census Bureau (2022). Public Use Microdata Sample (PUMS). https://www.census.gov/programs-surveys/acs/microdata.html

US Census Bureau (2022). American Community Survey Data. https://www.census.gov/programs-surveys/acs/data.html

US Environmental Protection Agency, (2018). Inventory of US Greenhouse Gas Emissions and Sinks 1990-2016 EPA 430 R-18-003. https://www.epa.gov/sites/production/files/2018-01/documents/2018_complete_report.pdf

Unger, O., Uriely, N., & Fuchs, G., (2016). The business travel experience. *Annals of Tourism Research*, 61, 142–156.

Uranga, R., (2017). *Driverless trucks: Coming soon to a road near you?* Southern California News Group. http://www.mercurynews.com/2017/03/05/driverless-trucks-coming-soon-to-a-road-near-you/

van Goeverden, K., van Arem, B., & van Nes, R., (2016). Volume and GHG emissions of long-distance travelling by Western Europeans. *Transportation Research Part D: Transport and Environment*, *45*, 28–47.

Van Wee, B., Rietveld, P., & Meurs, H., (2006). Is average daily travel time expenditure constant? In search of explanations for an increase in average travel time. *Journal of Transport Geography*, 14(2), 109–122.

Viscelli, S. (2018). Driverless? Autonomous trucks and the future of the American trucker. UC Berkeley Center for Labor Research and Education and Working Partnerships USA. http://driverlessreport.org/files/driverless.pdf

Wadud, Z., MacKenzie, D., & Leiby, P., (2016). Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles. *Transportation Research Part A: Policy and Practice*, 86, 1–18.

Wang, K., Salehin, M.F., & Habib, K.N., (2021). A discrete choice experiment on consumer's willingness-to-pay for vehicle automation in the Greater Toronto Area. *Transportation Research Part A: Policy and Practice*, *149*, 12–30.

Wang, D., He, B.Y., Gao, J., Chow, J.Y., Ozbay, K., & Iyer, S., (2020). Impact of COVID-19 Behavioral Inertia on Reopening Strategies for New York City Transit. *arXiv preprint arXiv:2006.13368*.

Wang, X., (2020). Impacts of the COVID-19 Pandemic on Travel and Tele-Activities in the US. Presented at the 2nd Bridging Transportation Researchers Conference.

Xue, L., & Zhang, Y., (2020). The effect of distance on tourist behavior: A study based on social media data. *Annals of Tourism Research*, 82, 102916.

Yankelevich, A., Rikard, R. V., Kadylak, T., Hall, M. J., Mack, E. A., Verboncoeur, J. P., & Cotton, S. R., (2018). Preparing the Workforce for Automated Vehicles. https://comartsci.msu.edu/sites/default/files/documents/MSU-TTI-Preparing-Workforce-for-AVs-and-Truck-Platooning-Reports%20.pdf

Yang, L., Shen, Q., & Li, Z., (2016). Comparing travel mode and trip chain choices between holidays and weekdays. *Transportation Research Part A: Policy and Practice*, 91, 273–285.

YASMEEN, R. (2019.). *Insights - Euromonitor.com*. Top 100 City Destinations, 2019 Edition. Retrieved September 30, 2022, from http://go.euromonitor.com/rs/805-KOK-719/images/wpTop100Cities19.pdf

Ye, X., Konduri, K., Pendyala, R. M., Sana, B., & Waddell, P. (2009). *A methodology to match distributions of both household and person attributes in the* *generation of synthetic populations*. In 88th Annual Meeting of the Transportation Research Board, Washington, DC. https://trid.trb.org/view/881554

Zhang, W., Guhathakurta, S., & Khalil, E.B., (2018). The impact of private autonomous vehicles on vehicle ownership and unoccupied VMT generation. *Transportation Research Part C: Emerging Technologies*, *90*, 156–165

Zhao, Y., & Kockelman, K.M. (2004). The random-utility-based multiregional input–output model: solution existence and uniqueness. *Transportation Research Part B: Methodological*, *38*(9), 789–807.

Zhao, Y., & Kockelman, K.M., (2018). Anticipating the regional impacts of connected and automated vehicle travel in Austin, Texas. *Journal of Urban Planning and Development*, *144*(4), 04018032.

Zhong, H., Li, W., Burris, M. W., Talebpour, A., & Sinha, K. C., (2020). Will autonomous vehicles change auto commuters' value of travel time? *Transportation Research Part D: Transport and Environment* 83.

Zumkeller, D., Chlond, B., Last, J., Manz, W., (2006). *Long-distance travel in a longitudinal perspective: the INVERMO approach in Germany* (No. 06-1131). https://trid.trb.org/view/776731