Transportation Insecurity Analysis Tool (TIAT)

Technical Documentation

December 2024

U.S. Department of Transportation
Office of the Secretary of Transportation

Bureau of Transportation Statistics

Table of Contents

Table of Contents ii

List of Tables iii

List of Figures iv

Acronyms v

Introduction 1

Modeling Approach 2

- Household Transportation Costs 4
- Household Income 22
- Housing Costs 23
- Development of Estimates for Household Types 24
- Other filters in the TIAT 26

Frequently Asked Questions (FAQ) 31

Glossary 34

Appendix A: Population Synthesis Using PopulationSim 36

- Instructions 36

Appendix B: Model Estimation Results 52

Appendix C: Use of Unweighted Survey Data for Estimating Disaggregate Models 83

Appendix D: Transportation Cost Flowchart with Sources 85

Appendix E: Rationale for Excluding Group Quarters Residents 86

Appendix F: Creating Separate 2019 and 2021 Transportation Cost Burden Estimates 88

Appendix G: Summary Comparison of 2019 and 2021 National Level Model Outputs 93

Appendix H: Estimates of Household Automotive Costs by Income Group 97

Appendix I: Use of ACS Housing Cost Data for Developing Housing Cost Estimates 104

List of Tables

Table 1: Data Sources for Household-Level Travel Models 5
Table 2: List of Travel Surveys Used for Developing the U.S. DOT Cost Burden Model 10
Table 3: List of All Independent Variables (for Household Level VMT/Miles Traveled Estimation Models), Data Sources, and Their Importance to the TIAT 13
Table 4: Control Variables Used in PopulationSim for Cost Burden Models. 16
Table 5: Data Sources for Multipliers 18
Table 6: Per Vehicle Spending by Income Level Among Households with at Least 1 Vehicle, Real 2021 dollars 20
Table 7: Per Vehicle Fixed Ownership Costs by Income-level by State, Real 2021 dollars 21
Table 8: Household Characteristics 25
Table 9: Filter Variables and Sources 27

List of Figures

Figure 1: Organization of Transportation Cost Factors by Modeling Approach 4

Acronyms

ACS	American Community Survey
BTS	Bureau of Transportation Statistics
CBSA	Core-Based Statistical Area
CEX	Consumer Expenditure Survey
COMPASS	Community Planning Association of Southwest Idaho
DOT	Department of Transportation
EPA	U.S. Environmental Protection Agency
FHWA	Federal Highway Administration
GDOT	Georgia Department of Transportation
GIS	Geographic Information Systems
GTFS	General Transit Feed Specification
HTS	Household travel survey
INRCOG	Iowa Northland Regional Council of Governments
LEHD	Longitudinal Employer-Household Dynamics
LODES	LEHD Origin-Destination Employment Statistics
MPO	Metropolitan planning organization
NCDOT	North Carolina Department of Transportation
NCTCOG	North Central Texas Council of Governments
NHTS	National Household Travel Survey
NYSDOT	New York State Department of Transportation
PSRC	Puget Sound Regional Council
PUMA	Public Use Microdata Area
PUMS	Public Use Microdata Sample
RPP	Regional Price Parity
SCDOT	South Carolina Department of Transportation
SLD	(EPA) Smart Location Database
TNC	Transportation Network Company
USDA SNAP	U.S. Department of Agriculture Supplemental Nutrition Assistance Program
U.S. DOT	U.S. Department of Transportation
USGS	U.S. Geological Survey
VMT	Vehicle miles traveled
WisDOT	Wisconsin Department of Transportation

Introduction

The U.S. Department of Transportation (U.S. DOT or the Department) defines Transportation Insecurity as a condition in which people are unable to get to where they need to go to meet the needs of their daily life regularly, reliably, affordably and safely.

This document provides details about the development process for the U.S. DOT Transportation Insecurity Analysis Tool (TIAT). U.S. DOT updated the exiting TIAT based on user engagement and feedback, expert reviews, third party reviews, and a workshop conducted in 2023. The TIAT includes updated estimates for several factors:

- Transportation Cost Burden
- Transportation & Housing Cost Burden
- Density Factors (rural vs urban, households per square mile, jobs per square mile)
- Safety and Environment
- Pedestrian Access
- Cyclist Access
- Motorist Access

The TIAT reflects a major methodological change to the transportation cost burden from the previous estimates. U.S. DOT developed these estimates of local-level transportation cost burden to better prioritize programs, policies, and investments targeting transportation affordability. Furthermore, these estimates can assist state and local agencies in prioritizing projects aimed at enhancing transportation affordability. The TIAT includes cost burden estimates for reference year 2021, which is the most recent year for which necessary data inputs were available. Additionally, the Department developed pre-pandemic estimates for reference year 2019. As compared to the related U.S. DOT Transportation Community (TC) Explorer tool, the TIAT focused on providing transportation insecurity data including cost and cost burden information in terms of their absolute dollar values or percent values, whereas the TC Explorer provides the same type of information but expressed in terms of percentiles and relative index values that better indicate how an area of interest compares to other areas of the U.S.[1]

[1] U.S. Department of Transportation. *Transportation Community (TC) Explorer v2.0 Technical Methodology*. November 18, 2024.

This document outlines the Department's development of the model approach and structure used for estimating household transportation cost burden. To be specific, it shows how the Department used Household Travel Survey (HTS) data from across the U.S. to develop models to estimate annual vehicle miles traveled (VMT), miles traveled by transit, and miles traveled by ride-hailing services (Ride-hailing and Transportation Network Companies (TNCs)), each split by work and non-work trips. It then discusses how the Department applied the models to nationwide synthetic households built for this project and finally applied cost multipliers for auto, transit and ride-hailing services to estimate travel costs per household. The subsequent sections describe the assessment of the accuracy of model estimates, frequently asked questions, and provide clear explanations for common terminology. Lastly, the documentation concludes with appendices that details the population synthesis technique, HTS data processing, and model estimation results.

This document also outlines or references the methodology updates to the density factors, the safety and environmental factors, and the pedestrian, cyclist, and motorist access factors.

Modeling Approach

The modeling approach outlined in this section aims to estimate transportation cost burden at the local level for various household types. This burden is determined by two main factors: household transportation costs and household income.

Household transportation costs encompass several key components, including automotive ownership costs, automotive operating costs, non-automotive transportation costs such as spending on regional public transit, and other transportation expenses, reflecting taxi and transportation network company (TNC) or ride-hailing services.

A simplified equation of Transportation Cost is:

Transportation Cost = Automotive Ownership Cost + Automotive Operating Cost + Regional Public Transit Cost + Taxi & Ride-hailing Services Cost

where:

Automotive Ownership Cost are the costs of owning a vehicle, including those costs that are incurred that are unrelated to vehicle mileage. This measure includes:

- Depreciation costs (also called the service flow cost of ownership)
- Finance charges
- Vehicle insurance
- Property tax on vehicles
- Registration fees

Automotive Operating Cost are the costs of driving a vehicle and keeping it in drivable condition, including:

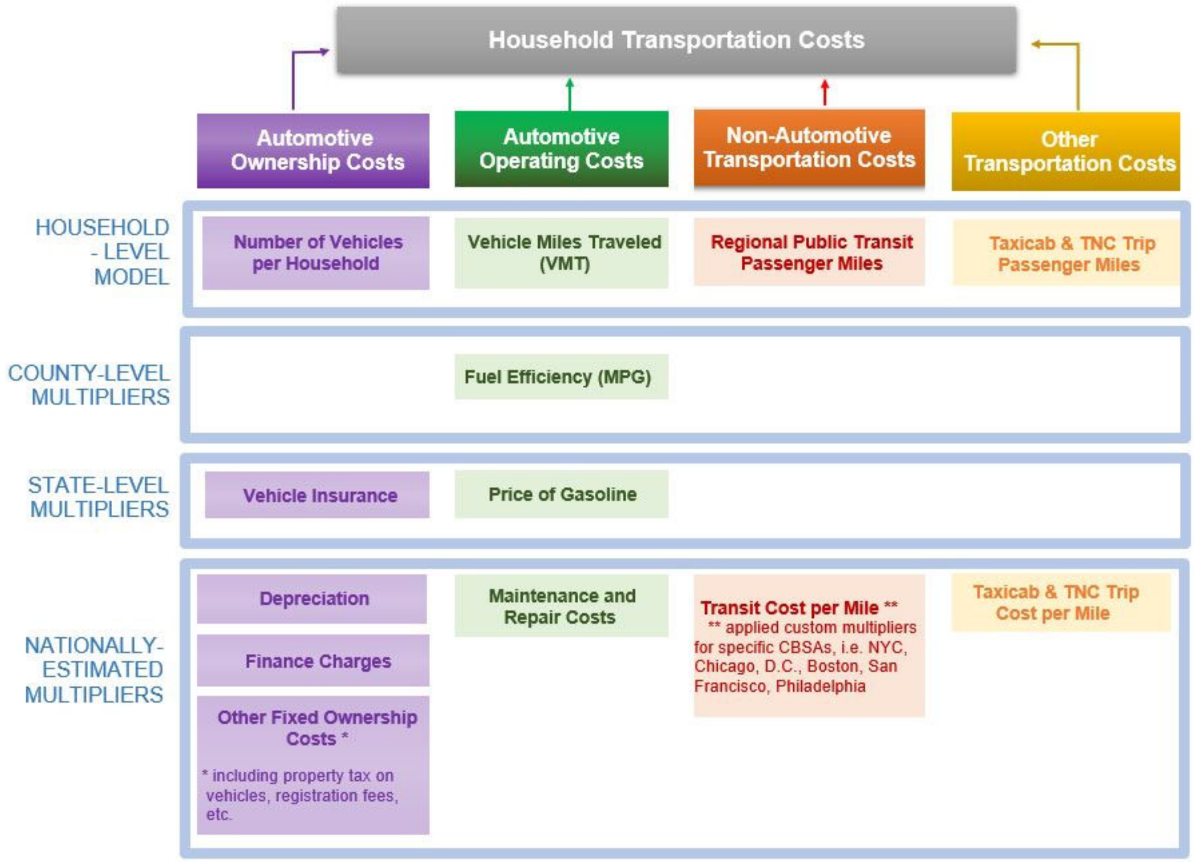
- Fuel spending
- Maintenance/repair costs

Regional Public Transit Cost and *Taxi & Ride-hailing Service Costs* are derived from:

- Number of miles traveled (by each service)
- Cost per mile (by each service)

Figure 1 provides an overview of the primary components of household transportation costs and the geographic coverage utilized in their estimation. Subsequent sections will elaborate on the methodologies employed in developing these models, offering insights into the intricate processes involved in estimating transportation cost burden at the local level.

Figure 1: Organization of Transportation Cost Factors by Modeling Approach



The Department collected data to estimate each of these cost components through HTS, the American Community Survey (ACS), the U.S. Environmental Protection Agency (EPA) Smart Location Database (SLD), and other auto ownership- and operation-related multipliers by available geography, and then developed models to estimate each component of household transportation cost across the country.

Household Transportation Costs

Population Synthesis The approach used to estimate household transportation costs consists of four overarching component elements:

1. Estimate a disaggregate household travel activity model (191,000 sample size; 17 regional/state/national household travel surveys)
2. Use a population simulation technique to generate representative data of household-level demographic, transportation, and housing information for the entire U.S. (ACS data).

3. Estimate the household-level travel for the population simulation data using the model from Step 1.
4. Estimate the transportation cost using external cost inputs (vehicle cost, gas, financing, transit costs, etc.) and household travel behavior from Step 3.

Each of these components and their sources are described in detail below.

Disaggregate Household Model Estimation

The Department assembled a combined HTS data set consisting of approximately 191,000 households, comprised of 25 separate existing HTS data collections that were obtained from 16 different agencies. This combined HTS data set formed the basis for estimating a disaggregate household travel model to predict the miles traveled by individual households for seven separate combinations of travel mode (e.g., personal auto, transit, Ride-hailing & ride-hailing) and trip purpose (e.g., work, non-work, long-distance) as a function of household, transportation, and land use/neighborhood attributes.

Table 1: Data Sources for Household-Level Travel Models

Cost Subcomponent	Main Cost Component	Data Source	Details
Number of vehicles per household	Auto Ownership	American Community Survey (ACS)	Controlled at the census block group level in population synthesis.
Vehicle miles traveled (VMT)	Auto Operation	Household travel surveys (HTS)	Divided by trip purpose (i.e., work trips, non-work trips and long-distance trips)
Regional public transit passenger miles	Non-Auto Transportation *	Household travel surveys (HTS)	Divided by trip purpose (i.e., work trips, non-work trips)
Taxi & ride-hailing trip passenger miles	Other Transportation *	Household travel surveys (HTS)	Divided by trip purpose (i.e., work trips, non-work trips)

* Estimated by applying models of cost per passenger mile, based on the National Transit Database (NTD) for public transit and HTS data for taxi & ride-hailing cost per mile.

The Department built seven different models, as the details in Table 1 show, to estimate the three cost subcomponents. Each model accounts for different trip characteristics by purpose albeit the team built a separate model for long-distance auto trips as they are not related to ordinary daily trips e.g., vacation and long-distance family visit trips. The segmentation of the models to treat work travel separately from non-work trips allows the model to more accurately reflect the influence of household characteristics such as the number of workers working from home and out of home, and the location of the residence block group relative to employment. All Household Travel Surveys include the seven models are listed below:

1. Household vehicle miles travelled (VMT) for work trips
2. Household VMT for non-work trips
3. Household VMT for long-distance trips
4. Miles traveled using public transportation for work trips
5. Miles traveled using public transportation for non-work trips
6. Miles traveled via taxi & ride-hailing (TNC) services for work trips
7. Miles traveled via taxi & ride-hailing (TNC) services for non-work trips

The Department then used the outputs of the first three models to estimate fuel spending, a component of the Auto Operation Cost as mentioned earlier. That is, the equation of fuel spending is:

- Fuel spending = VMT / MPG * gasoline price.

Where, miles per gallon (MPG) and gasoline price are county- and state-level multipliers while VMT is a household level estimate from HTS data (refer to Figure 1).

To address the complexities inherent in household travel behavior data, the models employ a two-part modeling approach. This approach is particularly effective at handling mixed discrete-continuous random variables, accommodating instances where households report zero VMT, no transit trip miles, or no taxi & ride-hailing trip miles. The first part of the model considers the probability for households with positive VMT, transit miles, or taxi & ride-hailing services trips. The second part of the model fits the distribution of VMT, miles traveled using transit, or miles traveled via taxi & ride-hailing services conditioned on the first part where households have non-zero values. The structure of the two-part model is:

- $E[Y|X] = \Pr(Y > 0|X) * E(Y|Y > 0, X]$

Where the first part of the model, $\Pr(Y > 0|X)$ is a binary logit model, while the second part, $E(Y|Y > 0, X]$ is a log-transformed linear least-squares regression model.

Although it is possible to estimate the two parts of each model (the binary logit model and the log-liner regression model) separately, the estimation method used in this application uses an iterative procedure across both parts to maximize the joint likelihood of the observed mileage data against the predicted outcome from both model components together.

The estimation results for the various models are presented and discussed in Appendix B.

Non-Automotive Transportation Costs Regional Adjustments

Non-Automotive Transportation Costs consider regional variations in fare costs and distance-based pricing models to accurately estimate expenses. By adjusting for specific metropolitan areas and incorporating real-world data on fare costs and ride-hailing pricing, these formulas provide a detailed approach to understanding transportation spending across different regions. Table 5 includes the summary of multipliers and equations applied for the non-automotive transportation cost estimation. Listed below are the steps used to calculate transit spending and taxi & ride-hailing spending using these established formulas.

Transit Spending Calculation

The model for the cost per mile for transit trips was estimated using the National Transit Database (NTD) from 2021, based on transit operators that reported both passenger revenue miles and passenger fare revenues for that year. For each transit operator reporting both types of data, the average fare paid per passenger mile is simply calculated as the reported fare revenues divided by the reported passenger revenue miles. Those operators were then grouped by region (CBSA) and the results were examined to identify any regions with significant passenger miles and an average fare per revenue mile that was much larger than the national average. Although many smaller operators did not provide both types of data, the data used represent a large majority of transit passenger miles traveled in the U.S. In fact, just the six (groups of) Core-Based Statistical Areas (CBSAs) listed below account for 61% of the transit passenger miles in the NTD (the New York City metro area at 47.1%, Chicago at 5.4% and Washington DC, Boston, San Francisco, and Philadelphia each at around 2%). After excluding those six regions, the average fare cost per mile for those operators reporting both mileage and revenues in the NTD is the “base” fare cost of 20 cents per mile. Note that other large regions such as Los Angeles, Seattle, and Atlanta were also tested, but did not show costs per passenger mile very different from the “base” level of 20 cents per mile.

The total transit spending is calculated using the following formula:

Transit Spending = Transit Miles × Fare Cost per Mile (farecostpm)

Where:

- **Base Fare Cost per Mile:** farecostpm = $0.20
- **Adjustments for Specific Metro Areas:**
    - for **NYC CBSA** (New York City CBSA): farecostpm=$0.20+$0.13
    - for **CHI CBSA** (Chicago CBSA): farecostpm=$0.20+$0.07
    - for **WAS CBSA** (Washington, D.C. CBSA): farecostpm=$0.20+$0.07
    - for **BOS CBSA** (Boston CBSA): farecostpm=$0.20+$0.15
    - for **SFO CBSA** (San Francisco and San Jose CBSAs): farecostpm=$0.20+$0.07
    - for **PHI CBSA** (Philadelphia CBSA): farecostpm=$0.20+$0.15

As described above, these six CBSAs represent major metropolitan areas in the U.S., where transit fare costs per mile in the 2021 NTD are higher than the national average—likely due to factors such as transit systems including underground rail and greater use of commuter rail.

<u>Taxi & Ride-hailing Spending Calculation</u>

The total spending on taxi & ride-hailing services was derived using HTS data that asked travelers to report the cost of taxi & ride-hailing trips. The resulting equation is as follows:

Taxi & Ride-hailing Spending = Taxi & Ride-hailing Miles × Taxi & Ride-hailing Cost per Mile (rhailcostpm)

Where:

- **For Distances 15 Miles or Greater:** rhailcostpm = $2.00
- **For Distances Less Than 15 Miles:** rhailcostpm = $10.834 - $4.709 * log(distance+1.0) + $0.321 * distance - $0.00233 * (distance squared)

Distance is the mileage in the reported household-day for Taxi & ride-hailing services within the Household Travel Survey. The cost per mile decreases as the distance increases, reflecting the fixed booking cost for a taxi or ride-hailing ride. However, beyond 15 miles, the cost stabilizes at approximately $2 per mile. The equation was estimated using linear regression across all HTS household-day observations that had non-zero taxi & ride-hailing mileage and reported (non-missing) data for taxi ride-hailing cost paid, using the cost divided by the mileage as the dependent variable and distance, distance squared, natural log of distance and an intercept term as independent variables.

Household Travel Surveys

One of the major innovations of this effort is the use of HTS data for recent years provided by several metropolitan planning organizations (MPOs), state departments of transportation (DOTs), and FHWA for the National Household Travel Survey (NHTS). Spanning from 2016 to 2023 and forming the foundation for the household-level models, the HTS data provide detailed information that is important for understanding and modeling travel behavior by car, taxi or ride-hailing services, and regional public transit.

HTS data captures daily trip information for each member of a household, along with comprehensive household and individual characteristics. This level of granularity enables modeling of travel behavior at the household level, offering a deeper understanding of mobility patterns by different household profiles. Despite being conducted by different entities across various regions, HTS data formats are largely standardized, facilitating their integration into cohesive national models. Every HTS involves collection of key household-level socio-demographic variables (e.g., household income, vehicle ownership), and person-level demographics for each household member including age, gender, employment status, race, and ethnicity.

Moreover, the widespread adoption of smartphone applications for data collection has enhanced the accuracy and richness of HTS data. These applications capture detailed trip attributes including the purpose of travel, mode of transportation, and the specific route taken, providing a comprehensive view of travel behavior. By incorporating actual trip paths, smartphone-based data collection methods enable precise measurement of household VMT, transit trip miles, and taxi and ride-hailing services trip miles, enhancing the reliability of transportation cost burden estimates. Also, the smartphone-based surveys capture travel for up to seven days from each household, including weekends, in order to provide evidence for all travel across the week. Table 2 shows the list of travel surveys permitted to use for this project.

Table 2: List of Travel Surveys Used for Developing the U.S. DOT Cost Burden Model

Survey	Agency	Areas covered	Year / Households / Household-days*
NHTS	FHWA	National samples	2016 / 33,212 / 33,212 2017 / 15,677 / 15,677 2022 / 7,422 / 7,422 2023 / 362 / 362
NHTS Add-on	NYSDOT, GDOT, WisDOT, NCDOT, SCDOT, NCTCOG, INRCOG	States of New York, Georgia, Wisconsin, North Carolina and, South Carolina. Dallas-Fort Worth metro region, Iowa Northland metro region	2016 / 53,896 / 53,896 2017 / 26,908 / 26,908
Regional HTS	Met Council	Twin Cities (Minneapolis–St. Paul) metro region, including three Wisconsin counties.	2018 / 2,292 / 8,512 2019 / 5,520 / 21,972 2021 / 7,014 / 17,711 2022 / 319 / 708
Regional HTS	Puget Sound Regional Council (PSRC)	Seattle (& central Puget Sound region)	2017 / 3,156 / 4,549 2019 / 2,894 / 5,280 2021 / 1,667 / 1,667 2023 / 4,343 / 10,228
Statewide HTS	Ohio Dept. of Transportation (ODOT)	Entire state of Ohio. Each year the study will focus on a different part of the state,	2016 / 954 / 4,363 2017 / 2,799 / 10,432 2018 / 2,309 / 8,718 2019 / 2,342 / 10,218 2020 / 1,498 / 6,786 2021 / 918 / 3,701 2022 / 3,569 / 15,571 2023 / 1,768 / 8,536
Statewide HTS	Utah Department of Transportation (UDOT)	Entire state of Utah	2023 / 8,416 / 17,042
Regional HTS	Spokane Regional Transportation Commission (SRTC)	Spokane County of Washington State	2022 / 1,818 / 5,255
Regional HTS	Community Planning Association of Southwest Idaho (COMPASS)	Boise, Southwest Idaho	2021 / 3,806 / 12,196

* Household and household-day sample sizes with complete data. The non-NHTS surveys covered multiple days for households using a smartphone app for data collection.

Household Characteristics and Neighborhood Characteristics

In the modeling process, U.S. DOT examined a range of household characteristics, survey period (year and month), and block group level neighborhood characteristics to understand their influence on travel behavior. The Department selected variables that demonstrated significance in explaining travel patterns for model development. Prior to model estimation, certain independent variables underwent mathematical transformations such as the natural log transformation and square root transformation to address statistical distribution issues and ensure adherence to regression assumptions. Additionally, all independent variables were normalized to maintain scale consistency across the models.

Each model incorporated a multitude of household and neighborhood characteristics, carefully selected based on their relevance and impact on travel behavior. Through rigorous evaluation of model fit and adherence to regression assumptions, the team identified the most suitable models for estimating transportation cost burden. The explanatory variables are nearly identical across the seven models. A simplified model form is as follows:

- $Y = X_{1(\,household\ characteristics)} + X_{2(survey\ data\ collection\ characteristics)} + X_{3(census\ block\ group-level\ neighborhood\ characteristics)}$

The first and second group of explanatory variables, X_1 and X_2 are from HTS, while the last group, X_3 come from the U.S. Environmental Protection Agency Smart Location Database (EPA SLD[2]). Note that the items in the Smart Location Database (SLD) prepared for the U.S. Environmental Protection Agency (EPA) are generally based off of Census data and OpenStreetMaps network data, except for the transit-related items, which are based on the General Transit Feed Specification (GTFS) data provided by the transit operators. See Appendix B for further detail on model estimation.

[2] https://www.epa.gov/sites/default/files/2021-06/documents/epa_sld_3.0_technicaldocumentationuserguide_may2021.pdf

Table 3 provides a concise summary of the household and neighborhood characteristics considered, along with their respective data sources. Note that the items in the Smart Location Database (SLD) prepared for the U.S. Environmental Protection Agency (EPA) are generally based off of Census data and OpenStreetMaps network data, except for the transit-related items, which are based on the General Transit Feed Specification (GTFS) data provided by the transit operators. See Appendix B for further detail on model estimation.

Table 3: List of All Independent Variables (for Household Level VMT/Miles Traveled Estimation Models), Data Sources, and Their Importance to the TIAT

Type of Independent Variable Characteristics	Independent Variables	Data Source	Importance to the TIAT
Household	Household income	HTS	Income influences the transportation cost burden in two ways; an increase in income may increase transportation expenses, thereby increasing the numerator. Conversely, a higher income in the denominator would result in lower transportation cost burden.
Household	Adults with vehicles	HTS	The higher the adults with vehicles, the higher the transportation cost burden (with more VMT). HTS values are used to estimate models from the HTS data. ACS PUMS data is used to create the synthetic population.
Household	Adults without vehicle	HTS	The higher the adults without vehicles, the lower VMT and the potentially higher transit and other transportation cost due to less trips taken (the lower transportation cost burden due to no vehicle ownership and operation costs).
Household	Number of workers per household	HTS	The higher the number of workers, the higher the transportation cost burden for non-remote workers.
Household	Number of (full/part time) commuters in the household	HTS	The higher the number of full-time commuters, the higher the transportation cost burden (the higher the number of part-time commuters, the lower the transportation cost burden).
Household	Number of (full/part time) workers working at home in the household	HTS	The higher the number of full-time (part-time) workers, working at home, the lower the transportation cost burden
Household	Number of children (less than 4 years old, and age 5 to 17) in the household	HTS	An increased number of children in the household can increase transportation cost burden because they may require to be dropped off and picked up from activities (requiring extra trip by adult on the way home after drop-off) and they cannot take transit independently.
Household	Number of people in age 18 to 34, age 35 to 54, and age 55 to 64	HTS	Each of the age group would affect the transportation cost burden differently by mode and purpose because trip making behavior are often different among them

Type of Independent Variable Characteristics	Independent Variables	Data Source	Importance to the TIAT
Household	Householders’ age 65+	HTS	An increased number of householders age 65+ may decrease the transportation cost burden because they often travel less than other adult age groups due to more likely being retired and less likely to have children in the household.
Household	Household Race for Adults (Hispanic, Black, Asian, and Other)	HTS	Adults in other races other than White may engage in different trip making patterns.
Data Collection	Day (Friday, Saturday, Sunday, and Monday)	HTS	Trip-making patterns on weekends would be different compared to the average weekdays
Data Collection	Month (11 months excluding October)	HTS	Trip-making patterns would be different in each month compared to October (seasonal impact).
Data Collection	Year (2016 to 2023, excluding 2021)	HTS	Trip-making patterns would be different by year compared to 2021 (COVID impact)
Data Collection	Diary	HTS	Data quality would be different from smartphone app survey
Neighborhood	Walkability	SLD	The higher the walkability score, there is less transportation cost burden.
Neighborhood	Transit Access	SLD	The more access to public transportation, there is less transportation cost burden.
Neighborhood	Transit Service Frequency	SLD	The more access to public transportation, there is less transportation cost burden.
Neighborhood	Automobile Access	SLD	Access to many destinations nearby is often associated with shorter trip lengths.
Neighborhood	Intersection density	SLD	Number of roadway intersections within an area is associated with higher connectivity in the street network and thus higher walkability and bikeability.
Neighborhood	Gross population density	SLD	The higher the household density and intensity, the more population in an area, and the more need for transportation in an area. Depending on the other neighborhood variables, this could have a positive or negative impact on transportation cost burden.
Neighborhood	Regional accessibility for jobs and working age population	SLD	The higher accessibility for jobs and working age group in an area, there is less transportation cost burden.

Type of Independent Variable Characteristics	Independent Variables	Data Source	Importance to the TIAT
Neighborhood	Core-based Statistical Area (CBSA) dummy	Joined Census Tiger BG and CBSA crosswalk geography	Cost burden by mode and purpose would differ by CBSA due to differences in the cost of living

Parking Costs & Tolls

This version of the transportation cost burden model excludes parking and tolls due to lack of adequate survey data. Only a few agencies collected toll usage but none collected toll expenditures. While some agencies collected parking costs, there were only a small number of trip records related to parking costs. The Department considered using regional data on toll and parking revenues in the areas of the country that have significant toll and parking costs, combined with other data such as the Longitudinal Employer-Household Dynamic Origin-Destination Employment Statistics (LEHD LODES) data on commute patterns. Because these data do not exist in a central national source, the Department determined that collecting this information consistently across the country and ensuring the data quality to be out of the scope for this initial effort. Such analyses could be part of future updates to the tool, and the Department encourages users to make use of local data to override the default assumption of zero in the tool regarding these costs with local data.

Population Synthesis

After the disaggregate household model estimation step described above, the Department used a modeling technique known as population synthesis to generate representative data of household-level demographic, transportation, and housing information for the entire U.S. based on data from the American Community Survey (ACS). To implement this population simulation technique, the Department utilized part of the ActivitySim[3] transportation modeling platform, which is an open-source software platform for activity-based travel modeling. Specifically, the Department used the PopulationSim population synthesizer tool, which is integrated into the ActivitySim platform. Additional information regarding ActivitySim

[3] https://activitysim.github.io/populationsim/

can be found in Appendix A.[4] The simulation uses 2017-2021 5-year ACS Public Use Microdata Sample (PUMS) data as a seed sample and sets the marginal distributions of targeted ACS variables of interest. Table 4 shows a list of ACS variables and categories used as control targets. These control targets ensure that the sum of households or persons in the synthetic population are nearly identical to the aggregate ACS data in the online Census data hub, at the Census block group (or tract; only for commute mode) level.[5] The list of variables is as follows:

Table 4: Control Variables Used in PopulationSim for Cost Burden Models.

Variable Type	Household	Person
Total	Household	person
gender	n/a	male, female
Age	n/a	age_0_4, age_5_17, age_18_34, age_35_49, age_50_64, age_65plus
Race	n/a	white, black, aapi, other,
Ethnicity	n/a	hispanic, non_hispanic,
Household Size	size_1, size_2, size_3, size_4, size_5_plus	n/a
Vehicle Availability	no_veh, veh_1, veh_2, veh_3, veh_4more	n/a
Commute Mode*	n/a	auto, transit, walk and bike, work from home, NA (non-worker)
Children	with children, without children	n/a

[4] https://activitysim.github.io/populationsim/
[5] https://data.census.gov/advanced

Variable Type	Household	Person
Student	n/a	university, non-university student
Household Income	income_$0 to_$25k, income_$25 to $50, income_$50 to $75, income_$75 to $100, income_$100 to $150, income $150plus	n/a
Workers		full-time, part-time, no worker
Monthly housing cost spending	own_$0 to_$799 own_$800 to $1499 own_$1500 to_$2499 own_$2500plus, rent $0 to $799 rent $800 to $1249 rent $1250 to $1999 rent $2000plus	

* Census tract level estimates were used for control target

The Department adjusted the seed sample based on the 5-year ACS PUMS data to 2021 1-year PUMS data of targeted ACS variables to account for the latest travel patterns. The model uses the 1-year adjusted seed sample and the adjusted marginal distributions of the targeted ACS variables of interest as inputs to the nationwide population synthesis. Appendix A provides a more detailed discussion of the single-year adjustment process, as well as a discussion of the rationale for not including group quarters residents in the synthetic population.

Application of the Disaggregate Household Model

The Department applied the model coefficients used for the seven household-level models described above to the synthetic population to generate estimates of VMT, miles traveled using public transportation, and miles traveled via taxi & ride-hailing services at the household level. The Department combined estimates for work trips, non-work trips and long-distance trips for each mode (private vehicle, public transport, and taxi & ride-hailing), and normalized them to cover the entire year of 2021 using the year-specific and month-specific effects. The Department generated separate estimates for the entire year of 2019, using a different synthetic population generated to reflect 2019 demographics and employment, and to better reflect the pre-COVID HTS data.

This document includes Appendix A, a guide to the survey data processing flow and the PopulationSim run and application process.

Estimate Transportation Costs Using External Cost Inputs

The household-level models capture how people travel (e.g., VMT, taxi & ride-hailing services mileage, and vehicles per household) but do not estimate costs directly. Rather, in an analysis separate from the household-level models described above, the Department estimated cost multipliers for each of the components of transportation costs as listed in Figure 1. The model applies the cost multipliers to corresponding travel activity to estimate transportation costs.

Table 5 summarizes the each of the cost multipliers from Figure 1.

Table 5: Data Sources for Multipliers

Cost Component	Geospatial Scale	Definition	Data Source
Vehicle depreciation	National	Sometimes called "service flow costs," this equals the average annual depreciation of the vehicle over the ownership period.	Bureau of Labor Statistics' Consumer Expenditure Survey (CEX), 2017-2021
Vehicle insurance	State	The average annual premium and insurance expenditure per vehicle.	National Association of Insurance Commissioners
Other fixed ownership costs	National	The costs associated with owning an automobile that do not vary directly with the number of miles driven (e.g., property tax on vehicles, registration fees, audio and video equipment, global positioning services).	Bureau of Labor Statistics' CEX
Finance charges	National	This includes finance charges on automotive loans, as well as lease charges.	Bureau of Labor Statistics' CEX

Cost Component	Geospatial Scale	Definition	Data Source
Gasoline Prices	State	The average price of gasoline per gallon,[6] which is calculated at the state level.	U.S. Department of Energy, State Energy Data Systems (SEDS)
Vehicle fuel efficiency	County	The number of miles the automobile can travel per gallon (miles per gallon).[7] Original data is tract level but estimated at the county level using harmonic means approach.	Argonne National Laboratory
Maintenance costs	National	The cost of keeping the vehicle in drivable shape, which includes maintenance and repairs costs, motor oil, and similar expenses. Inspection and licensing are also included in this category.	Bureau of Labor Statistics' CEX
Taxi & ride-hailing services cost per mile	National	A per-mile cost for using taxi & ride-hailing services. $2 per mile, If miles travelled distance (dist) is less than 15 miles, the cost per mile equals to $10.834 - $4.709 * log(dist+1.0) + $0.321 * dist - $0.00233 * (dist^2)	Household travel surveys (reported costs for taxi & ride-hailing services trips)

[6] U.S. Energy Information Administration (2023). *Motor Gasoline Price and Expenditure Estimates, 1970-2021.*
[7] Zhou, Y., Aeschliman, S., & Gohlke, D. (2020). *Affordability of household transportation fuel costs by region and socioeconomic factors* (No. ANL/ESD-20/11). Argonne National Lab (ANL), Argonne, IL (United States). Retrieved from https://www.energy.gov/eere/vehicles/articles/fotw-1175-march-1-2021-vehicles-registered-district-columbia-averaged-22.

Cost Component	Geospatial Scale	Definition	Data Source
Transit fare per mile	National (with six CBSAs where the fare per mile is substantially different from the national multiplier broken out)	A per-mile cost describing transit riders' spending for transit, $0.2 per mile. • for NYC, $0.33 • for Chicago, $0.27 • for Washington DC $0.27 • for Boston $0.35 • for San Francisco, $0.27 • for Philadelphia, $0.35	National Transit Database (for operators that reported both passenger miles and passenger fares)

The Department derived the national-level cost multipliers, including vehicle depreciation, finance charges, and maintenance costs through a thorough analysis of the BLS Consumer Expenditure Survey Public-Use Microdata files.

Appendix H describes the background, approach, interesting findings, and final results that were used to arrive at cost multipliers to be used in the U.S. DOT transportation cost burden models.

Table 6 lists the national cost per vehicle multipliers that the Department used in the transportation cost burden models.

Table 6: Per Vehicle Spending by Income Level Among Households with at Least 1 Vehicle, Real 2021 dollars

Income Group	Depreciation Costs	Finance Charges	Maintenance and Repair Costs
Overall average	3,766	149	418
Less than $24,999	4,082	86	374
$25,000 to $49,999	3,810	118	392
$50,000 to $99,999	3,695	162	411
$100,000 to $149,999	3,652	192	445
$150,000 or more	3,696	194	496

The Department estimated the cost multiplier for fixed ownership costs, which are comprised mostly of vehicle insurance premiums and expenditures, at the state-level. Because these costs can vary quite significantly state-to-state, the Department sought an additional data source to supplement the national estimates with state-level insurance figures. Table 7 shows the final fixed ownership cost multipliers, which include insurance costs as well as taxes and all other fixed ownership costs.

Table 7: Per Vehicle Fixed Ownership Costs by Income-level by State, Real 2021 dollars

State	Fixed Ownership Costs by Income level				
	Less than $24,999	$25,000 to $49,999	$50,000 to $99,999	$100,000 to $149,999	$150,000 or more
Alabama	940.76	987.65	1,008.74	1,027.61	1,140.23
Alaska	982.19	1,031.15	1,053.16	1,072.86	1,190.45
Arizona	1,036.59	1,088.26	1,111.49	1,132.28	1,256.38
Arkansas	924.06	970.12	990.83	1,009.36	1,119.99
California	1,024.24	1,075.28	1,098.24	1,118.78	1,241.40
Colorado	1,105.59	1,160.69	1,185.47	1,207.64	1,340.00
Connecticut	1,161.21	1,219.08	1,245.11	1,268.40	1,407.42
Delaware	1,187.01	1,246.17	1,272.77	1,296.58	1,438.69
District of Columbia	1,290.10	1,354.40	1,383.32	1,409.19	1,563.64
Florida	1,282.93	1,346.87	1,375.62	1,401.35	1,554.94
Georgia	1,163.26	1,221.24	1,247.31	1,270.64	1,409.91
Hawaii	880.08	923.94	943.66	961.31	1,066.68
Idaho	811.84	852.31	870.50	886.78	983.98
Illinois	948.65	995.93	1,017.19	1,036.22	1,149.79
Indiana	843.20	885.23	904.12	921.04	1,021.99
Iowa	799.35	839.19	857.11	873.14	968.84
Kansas	866.61	909.80	929.22	946.60	1,050.35
Kentucky	952.54	1,000.01	1,021.36	1,040.47	1,154.50
Louisiana	1,359.94	1,427.72	1,458.20	1,485.47	1,648.28
Maine	792.65	832.15	849.92	865.82	960.71
Maryland	1,143.15	1,200.12	1,225.74	1,248.67	1,385.53
Massachusetts	1,119.27	1,175.06	1,200.14	1,222.59	1,356.59
Michigan	1,288.10	1,352.30	1,381.16	1,407.00	1,561.21
Minnesota	919.04	964.85	985.45	1,003.88	1,113.91
Mississippi	986.80	1,035.98	1,058.10	1,077.89	1,196.03
Missouri	944.36	991.43	1,012.59	1,031.53	1,144.59
Montana	884.11	928.18	947.99	965.72	1,071.57
Nebraska	864.28	907.36	926.73	944.06	1,047.53
Nevada	1,172.98	1,231.45	1,257.73	1,281.26	1,421.69
New Hampshire	899.40	944.22	964.38	982.42	1,090.09
New Jersey	1,257.76	1,320.44	1,348.63	1,373.86	1,524.44
New Mexico	938.01	984.76	1,005.78	1,024.60	1,136.90
New York	1,306.56	1,371.68	1,400.96	1,427.17	1,583.59
North Carolina	827.45	868.69	887.23	903.83	1,002.89

State	Fixed Ownership Costs by Income level				
	Less than $24,999	$25,000 to $49,999	$50,000 to $99,999	$100,000 to $149,999	$150,000 or more
North Dakota	788.11	827.39	845.06	860.86	955.22
Ohio	859.37	902.20	921.46	938.70	1,041.58
Oklahoma	938.00	984.75	1,005.77	1,024.59	1,136.88
Oregon	983.23	1,032.24	1,054.27	1,074.00	1,191.71
Pennsylvania	990.69	1,040.07	1,062.27	1,082.14	1,200.75
Rhode Island	1,260.96	1,323.81	1,352.07	1,377.36	1,528.33
South Carolina	1,072.97	1,126.44	1,150.49	1,172.01	1,300.47
South Dakota	820.61	861.51	879.90	896.36	994.60
Tennessee	903.99	949.04	969.30	987.44	1,095.66
Texas	1,088.78	1,143.05	1,167.45	1,189.29	1,319.64
Utah	960.10	1,007.96	1,029.47	1,048.73	1,163.67
Vermont	850.97	893.38	912.45	929.52	1,031.40
Virginia	900.36	945.24	965.42	983.48	1,091.27
Washington	1,027.92	1,079.16	1,102.19	1,122.81	1,245.88
West Virginia	953.18	1,000.69	1,022.05	1,041.17	1,155.29
Wisconsin	833.17	874.70	893.37	910.08	1,009.83
Wyoming	845.12	887.24	906.18	923.13	1,024.31

The Department applied the cost multipliers to corresponding transportation components at the household-level and then summed to the Census tract level for the 25 household types described in Table 8.

Household Income

Income measured by the ACS is “money income,” which is the total pre-tax income earned by individuals, excluding certain lump sum payments and capital gains. It captures regularly received income that households/individuals can spend. It includes income received as wages and salaries, self-employment income, property income (i.e., dividends, interest, rent), government transfer payments (i.e., social security, unemployment and worker’s compensation, public assistance), retirement income (private and government), interpersonal transfers (e.g., alimony, child support), and other recurrent income. The Department used the ACS as the data source for household income to maintain consistency in the sources used for synthetic population controls. No other publicly available database provides this capability for such fine-grained geographic areas.[8, 9, 10]

Housing Costs

Housing Costs represent the average annual housing cost per household, which in the TIAT is calculated for all households as well as for various subsets of households (referred to in the TIAT as Household Profiles) for each Census tract. These housing costs are calculated across a synthetic population of all households in the U.S. using data from the 2017–2021 American Community Survey (ACS) Public Use Microdata Sample (PUMS) and ACS 5-year tables, processed through the PopulationSim tool using a technique known as population synthesis. This population synthesis technique is described earlier in the “Population Synthesis” section of this TIAT Technical Documentation, with additional details also provided in Appendix A.

The average housing costs presented in the TIAT are not those directly reported in the ACS or PUMS data but instead are derived through the population synthesis process which uses the ACS and PUMS as inputs. This process ensures that the synthesized population matches the real-world distributions of variables such as income, household size, and housing tenure (owner/renter). Specifically:

- Housing costs are categorized based on whether households are renters or owners.
- These costs are drawn from ACS PUMS data and controlled at multiple geographic levels:
    - Census block group distributions from the ACS 5-year tables (Tables B25087 and B25063) provide granular targets for owner and renter housing costs, respectively.

---

[8] Brady, P. J., & Bass, S. (2021). Comparing the Current Population Survey to Income Tax Data. Investment Company Institute. https://www.ici.org/system/files/2022-03/21_ppr_cps_tax_compare.pdf#:~:text=Comparing%20the%20Current%20Population%20Survey%20to%20Income%20Tax,a%20considerable%20amount%20of%20retirement%20and%20investment%20income.

[9] Lin, D. (2022). Methods and Assumptions of the CPS ASEC Tax Model (SEHSD Working Paper FY-2022-18). U.S. Census Bureau. https://www.census.gov/library/working-papers/2022/demo/SEHSD-wp2022-18.html

[10] Rothbaum, J. L. (2015). Comparing Income Aggregates: How do the CPS and ACS Match the National Income and Product Accounts, 2007-2012 (SEHSD Working Paper 2015–01). U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2015/demo/SEHSD-WP2015-01.pdf

- At the PUMA (Public Use Microdata Area) level, these distributions are further aligned with the ACS 1-year data for 2019 and 2021.

For owners, selected monthly owner costs (SMOCP) are adjusted to 2021 dollars using the ACS adjustment factor (ADJHSG) and include costs such as mortgage payments, property taxes, insurance, and utilities. For renters, gross rent (GRNTP) is similarly adjusted to reflect 2021 dollars. Housing costs are then grouped into 8 categories—4 for owners and 4 for renters—based on national distributions and assigned to each household record in the synthetic population.

The tool outputs the mean housing cost (not the median) across all households within each geography and household profile combination. While median housing costs could be calculated in a similar manner, they are not currently included in the tool. This option could be considered for future updates.

For additional details, including the categorization of housing costs and the variables used in the calculation, refer to the Appendix I (Use of ACS Housing Cost Data for Developing Housing Cost Estimates).

Development of Estimates for Household Types

The Department collected household income data from the ACS and estimated household transportation costs for 25 household types related to income and transportation options shown in Table 8. These range from estimates for a “typical” household (labeled “All Households” to households with more limited income or limited access to vehicles, transit service, or destinations within walking distance. Use of household types allows users of the TIAT to understand how different households in the same area may experience a different transportation cost burden.

Table 8: Household Characteristics

General Topic	Household Type	Data Source(s) for Classifying Household Characteristic	Number of Household Types	Options
All Households	All Households	Not applicable	1	Average Household
Income	Relative Income Quintile	ACS	5	1st quintile, 2nd quintile, 3rd quintile, 4th quintile, 5th quintile
Income	Fixed Income Ranges	ACS	5	Less than $24,999; $25,000 to $49,999; $50,000 to $99,999; $100,000 to $149,999; $150,000 or more
Income	Official Poverty Measure	ACS	4	Below 100 percent of the poverty level, 100 to 149 percent of the poverty level, 150 to 199 percent of the poverty level, 200 percent of the poverty level or greater
Transportation	Vehicle Availability	ACS	3	No cars in household, Cars in household but fewer cars than the number of adults, Cars in household equal to or greater than the number of adults
Transportation	Transit Availability	EPA Smart Location Database data.	3	No transit service available, Transit service likely available, but limited data, Transit service available
Transportation	Walkability	EPA Smart Location Database.	4	Least walkable (natwalkind* <= 5.75), Below average walkable (5.75 < natwalkind <= 10.5), Above average walkable (10.5 < natwalkind <= 15.25), Most walkable (15.25 < natwalkind <= 20)
Total			**25**	

* Walkability index from SLD data, https://www.epa.gov/system/files/documents/2023-10/epa_sld_3.0_technicaldocumentationuserguide_may2021_0.pdf

Other filters in the TIAT

In addition to the new cost burden model, the TIAT also updated a series of variables to better understand the intersection of transportation costs and other factors, including income, housing costs, density factors, safety and environmental factors, and access variables including:

- Income
- Housing Costs
- Density Factors
    - Rural vs urban
    - Households per square mile
    - Jobs per square mile
- Safety and Environment
    - Traffic fatalities (2018-2022)
    - Automotive CO_2 Emissions per household
- Pedestrian Access
- Cyclist Access
- Motorist Access

The Department outlines the access methodology in the technical documentation for the TC Explorer tool, in Section 4 in the Appendix titled “Calculating Access to Destinations.” The access filters includes filters for overall access (defined by 30 minutes by walking (one mile), biking (five miles) or driving (30 minutes outbound at 8am on a Wednesday) to key destinations including: (1) educational facilities, (2) grocery stores, (3) public libraries, (4) medical facilities, including pharmacies, (5) parks, (6) post offices, (7) public transit service, (8) population, and (9) jobs.

Table 9 summarizes these filters and the sources.

Table 9: Filter Variables and Sources

Variable	Unit of Measure	Description	Source
Housing Cost (ACS)	Dollars	Average annual housing cost per household (for the subset of households in the selected Household Profile). Average across synthetic population households.	PopulationSim (2017 – 2021 PUMS & ACS)
Rural or Urban	Category	Urban or rural designation for Census Tract. For tracts with at least 100 residents, the tract was coded as "URBAN" if the majority of its residents lived in urban areas of at least 50,000 residents and "RURAL" otherwise. For tracts with less than 100 residents, the tract was coded as "URBAN" if the majority of its land area was in urban areas of at least 50,000 residents and "RURAL" otherwise.	U.S. Census Bureau 2020 Decennial Census
Household Density	Households per square mile	A categorical variable defining whether household density is low, medium, or high. Based on the household density neighborhood characteristic. Low: < 250 households per square mile; Medium: 250 - 2500 households per square mile; High: > 2500 household per square mile. Classification based on Census tract-level data.	HOUSEHOLD COUNT: U.S. Census Bureau American Community Survey (ACS) 5-Year (2017-2021) - Table B11016
Employment Density	Jobs per square mile	A categorical variable defining whether employment density is low, medium, or high. Based on the employment density neighborhood characteristic. Low: < 85 jobs per square mile; Medium: 85 - 1400 jobs per square mile; High: > 1400 jobs per square mile.	EMPLOYMENT: U.S. Census Bureau LEHD Origin-Destination Employment Statistics (LODES)

Variable	Unit of Measure	Description	Source
Traffic Fatalities (2018 – 2022)	Persons	Count of traffic fatalities suffered in motor vehicle traffic crashes from 2018–2022 within a 250-ft buffer around each Census tract. Note that due to the buffer, these values are not additive across multiple tracts.	U.S. DOT NHTSA, Fatality Analysis Reporting System (FARS)
Pedestrian Access Score	Category	Pedestrian accessibility index, on a scale from 0 to 1 with 1 meaning “most access.”	Calculated from various sources. See individual access fields on specified sources.
Cyclist Access Score	Category	Cyclist accessibility index, on a scale from 0 to 1 with 1 meaning “most access.”	Calculated from various sources. See individual access fields on specified sources.
Motorist Access Score	Category	Motorist accessibility index, on a scale from 0 to 1 with 1 meaning “most access.”	Calculated from various sources. See individual access fields on specified sources.
Automotive CO2 Emissions per Household	Kilograms	Estimated annual kilograms of automotive carbon dioxide emissions per household. Calculated as 8.887 kg of CO2 per gallon of fuel times VMT over average fuel economy.	U.S. Environmental Protection Agency (EPA), Greenhouse Gas Equivalencies. On average, 8.887 kilograms of CO2 are emitted from burning one gallon of gasoline.
Walkable Educational Facilities	Facilities	Primary, secondary, and post-secondary education facilities within a 30-minute walk (1 mile) of the tract population centroid.	U.S. Department of Education National Center for Education Statistics (NCES) data
Walkable Grocery Stores	Facilities	Grocery stores within a 30-minute walk (1 mile) of the tract population centroid. Number of grocery stores ("Supermarket", "Super Store", or "Large Grocery Store")	U.S. Department of Agriculture Supplemental Nutrition Assistance Program (USDA SNAP) retailers as of January 2024

Variable	Unit of Measure	Description	Source
Walkable Medical Facilities	Facilities	Hospitals, outpatient care facilities, and pharmacies within a 30-minute walk (1 mile) of the tract population centroid.	SafeGraph POI dataset received from U.S. DHS HIFLD-Secure database
Walkable Parks	Facilities	Parks within a 30-minute walk (1 mile) of the tract population centroid.	HERE POI dataset received from U.S. DHS HIFLD-Secure database
Walkable Transit Trips	Trips	Transit trips in a regular service week within a 30-minute walk (1 mile) of the tract population centroid.	U.S. Federal Transit Administration GTFS feeds for National Transit Database reporters, April 2024
Walkable Jobs	Jobs	Jobs (2020 LEHD jobs data) within a 30-minute walk (1 mile) of the tract population centroid.	U.S. Census Bureau 2020 LEHD Workplace Area Characteristics
Bikeable Educational Facilities	Facilities	Primary, secondary, and post-secondary education facilities within a 30-minute bike ride (5 miles) of the tract population centroid.	U.S. Department of Education National Center for Education Statistics (NCES) data
Bikeable Grocery Stores	Facilities	Grocery stores within a 30-minute bike ride (5 miles) of the tract population centroid. ("Supermarket", "Super Store", or "Large Grocery Store").	USDA SNAP retailers as of January 2024
Bikeable Medical Facilities	Facilities	Hospitals, outpatient care facilities, and pharmacies within a 30-minute bike ride (5 miles) of the tract population centroid.	SafeGraph POI dataset received from U.S. DHS HIFLD-Secure database
Bikeable Parks	Facilities	Parks within a 30-minute bike ride (5 miles) of the tract population centroid.	HERE POI dataset received from U.S. DHS HIFLD-Secure database

Variable	Unit of Measure	Description	Source
Bikeable Transit Trips	Trips	Transit trips in a regular service week within a 30-minute bike ride (5 miles) of the tract population centroid.	U.S. Federal Transit Administration GTFS feeds for National Transit Database reporters, April 2024
Bikeable Jobs	Jobs	Jobs (2020 LEHD jobs data) within a 30-minute bike ride (5 miles) of the tract population centroid.	U.S. Census Bureau 2020 LEHD Workplace Area Characteristics
Drivable Educational Facilities	Facilities	Primary, secondary, and post-secondary education facilities within a 30-minute drive of the tract population centroid.	U.S. Department of Education National Center for Education Statistics (NCES) data
Drivable Grocery Stores	Facilities	Grocery stores within a 30-minute drive of the tract population centroid ("Supermarket", "Super Store", or "Large Grocery Store").	USDA SNAP retailers as of January 2024
Drivable Medical Facilities	Facilities	Hospitals, outpatient care facilities, and pharmacies within a 30-minute drive of the tract population centroid.	SafeGraph POI dataset received from U.S. DHS HIFLD-Secure database
Drivable Parks	Facilities	Parks within a 30-minute drive of the tract population centroid.	HERE POI dataset received from U.S. DHS HIFLD-Secure database
Drivable Transit Trips	Trips	Transit trips in a regular service week within a 30-minute drive of the tract population centroid.	U.S. Federal Transit Administration GTFS feeds for National Transit Database reporters, April 2024
Drivable Jobs	Jobs	Jobs (2020 LEHD jobs data) within a 30-minute drive of the tract population centroid.	U.S. Census Bureau 2020 LEHD Workplace Area Characteristics

Frequently Asked Questions (FAQ)

1. What is the TIAT, and how was it developed?

The TIAT includes a cost burden model that estimates transportation cost burden at the local level using the best available data at the time. The estimates of the transportation cost burden that it produces can be accessed through the U.S. DOT Transportation Insecurity Analysis Tool (TIAT), which allows for both interacting with and downloading the cost burden estimates. As compared to the related Transportation Community (TC) Explorer tool that the U.S. DOT has developed and published, the TIAT is primarily focused on providing transportation insecurity data including cost and cost burden information in terms of their absolute dollar values or percent values, whereas the TC Explorer provides the same type of information but expressed in terms of percentiles and relative index values that better indicate how an area of interest compares to other areas of the U.S.

2. What is transportation cost burden?

Transportation cost burden is the percent of a household's income that is spent on transportation.

3. What transportation modes are included in the TIAT?

The transportation cost burden estimate accounts for costs associated with owning and driving an automobile, using transit, and using taxi or ride-hailing services. The TIAT also allows users to examine households' transportation cost burden based on different levels of household vehicle ownership, the availability of regional public transportation service, and the ability to reach destinations on foot (also called "walkability"), by bicycle, or by vehicle. Thus, tool users can see how transportation cost burden varies based on households' ability to use automobiles, availability of public transportation, walkability, bikeability, and drivability.

4. What will the data be used for?

U.S. DOT developed the TIAT to assist applicants of its discretionary grant programs and to make data available to address the high transportation cost burden. Other potential users include State DOTs, MPOs, and other transportation planning organizations to inform transportation planning. The data is downloadable and is available for use by the general public, governmental organizations, businesses, non-profit organizations, researchers, and any other interested individual or organization.

5. How are these estimates different from earlier efforts to estimate transportation cost burden?

   Existing transportation cost burden estimates produced by the U.S. DOT as part of the TC Explorer tool (and a previous version of the TIAT) were based primarily on aggregate Census tract-level data and other data sources at various levels of geographic granularity. For the new TIAT presented in this User Guide, large disaggregate household-level data sets and related modeling and simulation methods have been utilized, and additional improvements in the geographic granularity of certain other data (such as vehicle fuel economy, auto insurance costs, etc.) have also been incorporated in order to produce more granular cost burden estimates and cross-tabulations. There have also been several prior efforts to estimate transportation cost burden at the local level in the United States, including the Location Affordability Index (LAI) by the U.S. Department of Housing and Urban Development and the U.S. Department of Transportation, as well as the Center for Neighborhood Technology's Housing and Transportation (H+T) Affordability Index. The TIAT differs from these models in several ways. While they both also include costs associated with housing, the TIAT focuses on transportation costs. Additionally, the TIAT uses several previously unavailable data sources, including the use of regional, statewide, and national HTS to better understand how and how much Americans travel.

6. At what geospatial scale are the results?

   The estimates of household transportation cost burden are at the level of Census tracts, as well as counties, states, and the country as a whole. Within the TIAT, users can select multiple Census tracts to produce a weighted average transportation cost burden for their selection.

7. When will the model be updated in the future?

   U.S. DOT has built the model to allow it to be updated and to allow new transportation costs to be included as adequate data becomes available. An exact update schedule does not yet exist.

8. Does the model account for costs associated with walking, biking, car subscription services, parking, or tolls?

   Adequate data does not yet exist to account for households' costs for walking, biking, car subscription services, parking, and tolls at the local level and for different household types.

9. What household types are results available for?

Transportation cost burden estimates are available for a “typical” household in an area, as well as different household types defined by household income (income quintile, select income ranges, and poverty level) and by the household’s transportation availability (vehicle ownership, transit availability, and walkability).

10. Why might I see minor discrepancies when making transportation cost adjustments?

There are small amounts of expected variation in data due to rounding, approximations, or sampling. In cases where values are rounded to a certain number of decimal places, minor discrepancies can occur, but they typically fall within the range of $200. Data is stored with reduced decimal precision (e.g., two decimal places) to optimize storage and processing.

11. Are there any plans for future enhancements or expansions of the model and its capabilities?

U.S. DOT has built the model to allow new transportation costs to be included as adequate data becomes available. Plans for these updates are under development.

Glossary

Developed Area: Land characterized by urban or built-up features.

Filters: Criteria applied to refine analysis of Census tracts within the tool, such as transit availability, walkability, household density, employment density, and urban/rural location. Selecting a filter removes Census tracts from the view.

Finance Charges: Costs associated with financing the purchase or lease of a vehicle, including interest payments and fees.

Fixed Ownership Costs: Expenses related to vehicle ownership that do not vary with the number of miles driven, such as insurance premiums and taxes.

Household Characteristics: Attributes and demographics of households, including income, household size, number of children, number of commuters, and vehicle ownership.

Household Profiles: Categories used to classify households based on income levels and transportation options, facilitating analysis of transportation cost burden.

Household Travel Surveys: Surveys collected by various transportation agencies such as MPOs and state DOTs on travel behavior and patterns of households, including modes of transportation, trip purpose, and frequency.

Income: Household income encompasses the total annual earnings, before deductions for taxes and other items, of all individuals residing within a specific household. This includes wages, salaries, government assistance, and other sources of revenue contributing to the household's financial resources.[11]

Income Quintiles: Divisions of the population into five equal groups based on income levels.

Maintenance Costs: Expenses associated with keeping a vehicle in drivable condition, including repairs, servicing, and replacement parts.

Neighborhood Characteristics: Features of a neighborhood or area, such as walkability, bike access, transit availability, housing density, and employment opportunities.

Official Poverty Measure: The federal poverty guidelines (FPG) as defined by the U.S. Department of Health and Human Services (HHS) are used as the official poverty measure.[12]

---

[11] For additional details regarding the specific elements included in the measure of household income, see the U.S. Census Bureau definition of income at https://www.census.gov/glossary/?term=Income, and the income-related questions (43 and 44) used on the 2021 American Community Survey (ACS) questionnaire at https://www2.census.gov/programs-surveys/acs/methodology/questionnaires/2021/quest21.pdf#page=18.

[12] *U.S. Federal Poverty Guidelines*. U.S. Department of Health and Human Services (HHS), Office of the Assistant Secretary for Planning and Evaluation (ASPE). Retrieved from https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-guidelines, and from

The FPG vary by the number of people in the household, and also differ for the states of Alaska and Hawaii as compared to the rest of the U.S.[13]

Population Synthesis: The process of creating a synthetic population representative of real-world demographics and characteristics.

Taxi and Ride-hailing: Services providing transportation for hire, typically on-demand and utilizing vehicles such as taxis or privately owned cars.

Transit Availability: The ease with which individuals can access and utilize public transit services within their community or region.

Transportation & Housing Cost Burden: The ratio of a household's combined transportation-related expenses and housing-related expenses to its gross household income. It indicates the proportion of income spent on both transportation and housing combined. A household is considered transportation and housing cost-burdened if it spends 45 percent or more of its gross income on both transportation and housing combined.

Transportation Cost Burden: The ratio of a household's transportation-related expenses to its household income. It indicates the proportion of income spent on transportation. A household is considered transportation cost-burdened if it spends 15 percent or more of its gross income on transportation.

Transportation Cost: Expenses associated with various modes of transportation, including automobile ownership and operation, taxi, ride-hailing, and regional public transit.

Vehicle Depreciation: The decrease in the value of a vehicle over time, typically due to age, wear and tear, and market factors.

Vehicle Fuel Efficiency: The distance in miles that a vehicle can travel per gallon of fuel consumed, measured as miles per gallon.

VMT (Vehicle Miles Traveled): The total distance traveled in owned or leased automobiles by a household within a specified time period.

---

https://aspe.hhs.gov/sites/default/files/documents/e03cb39a940516a81d5537829bad9430/guidelines-1983-2024.xlsx

[13] There are two versions of the "federal poverty measure." The first version is the "poverty thresholds" defined by the U.S. Census Bureau, which are used primarily for statistical purposes. The second version is the "federal poverty guidelines" defined by the U.S. Department of Health and Human Services (HHS). The federal poverty guidelines (FPG) are a simplification of the poverty thresholds and are used primarily for administrative purposes such as determining financial eligibility for various federal programs.

Appendix A: Population Synthesis Using PopulationSim

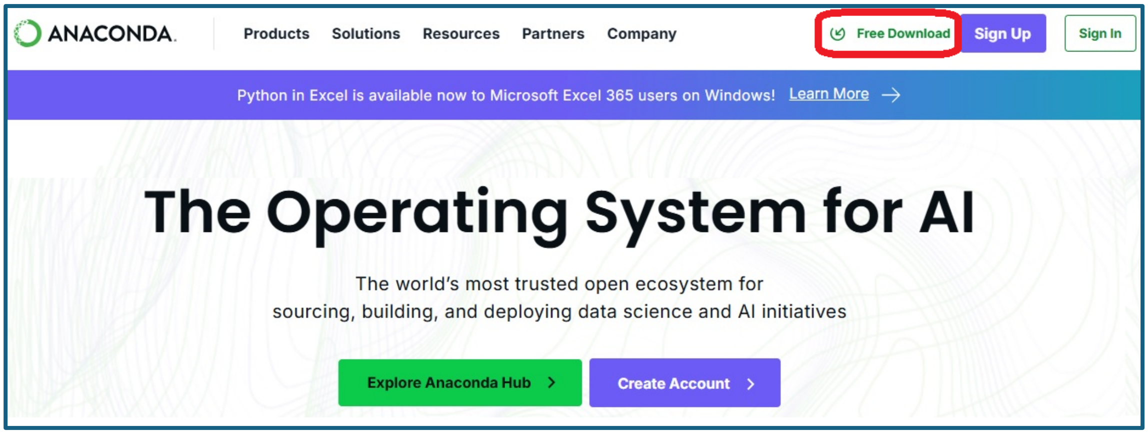
Instructions

Requirement: Python

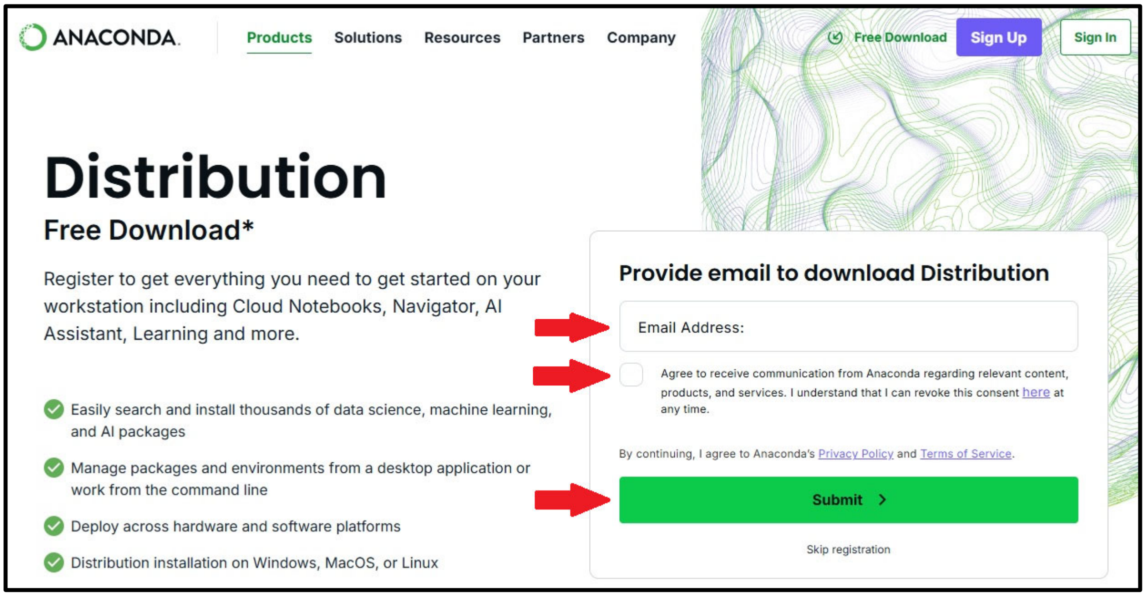
Recommended program: either use anaconda, miniconda or mamba to run populationsim.

Option 1: Download and install anaconda:

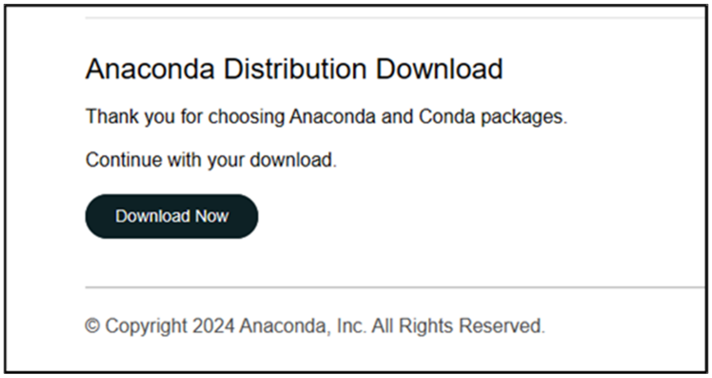
- Go to https://www.anaconda.com, and click "Free Download" button in the top right corner of the website.



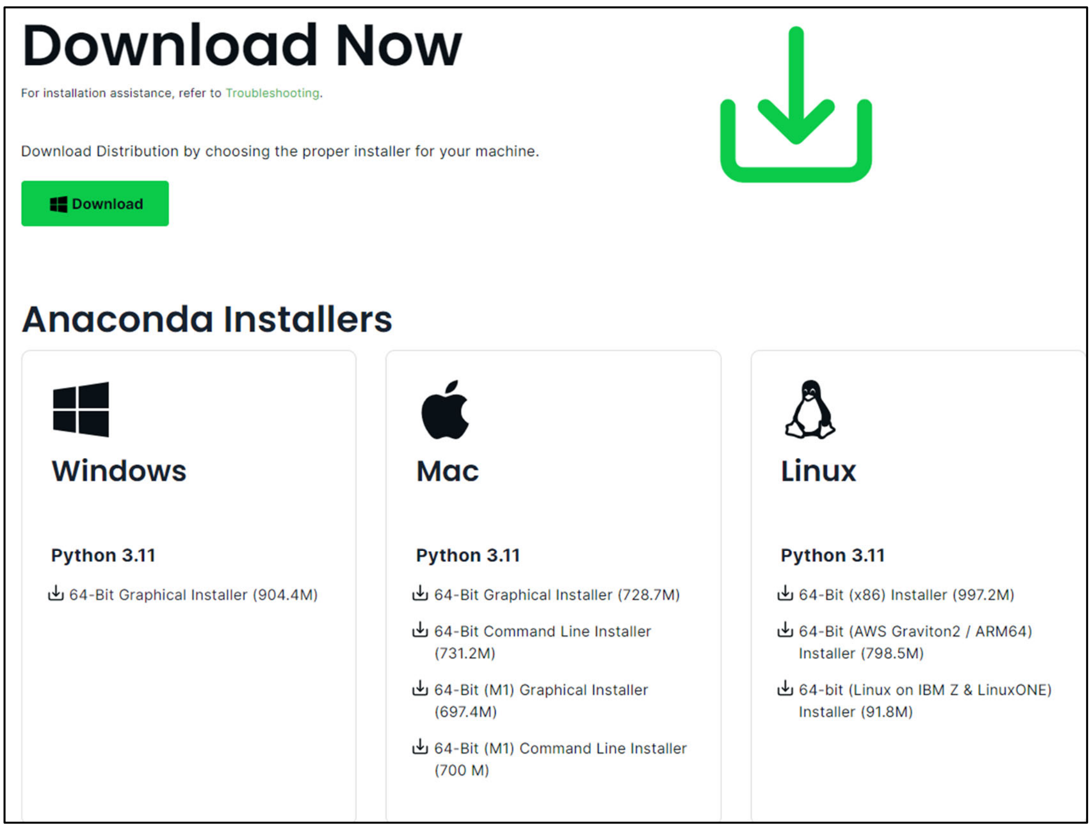
- It will take you to the "Distribution" page. Enter your email address, check the box (optional) to receive communication from anaconda, and click the submit button.



- You will receive the downlink link in your email shortly after.



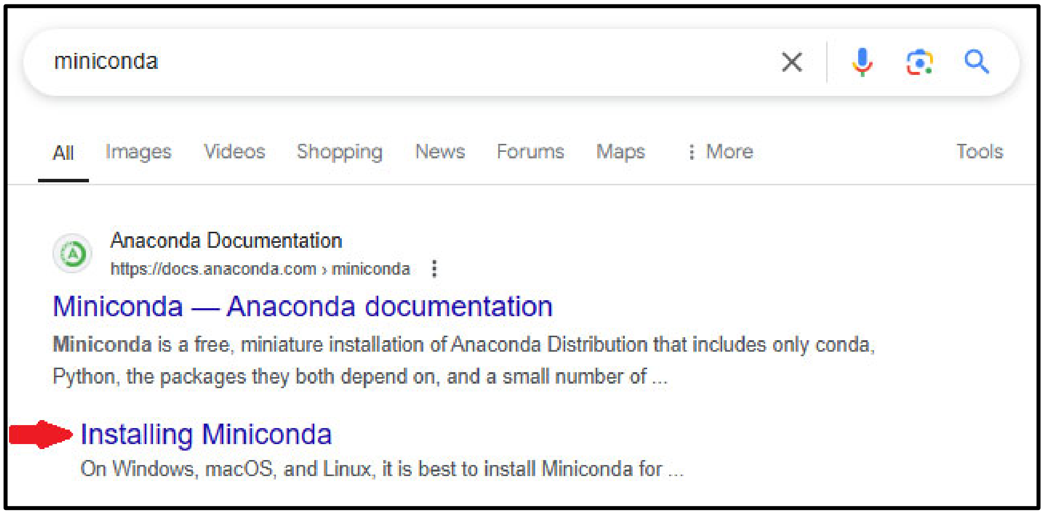
- Click the 'Download' button (in your email) that will direct to a webpage. It shows the option to download for different operating systems.



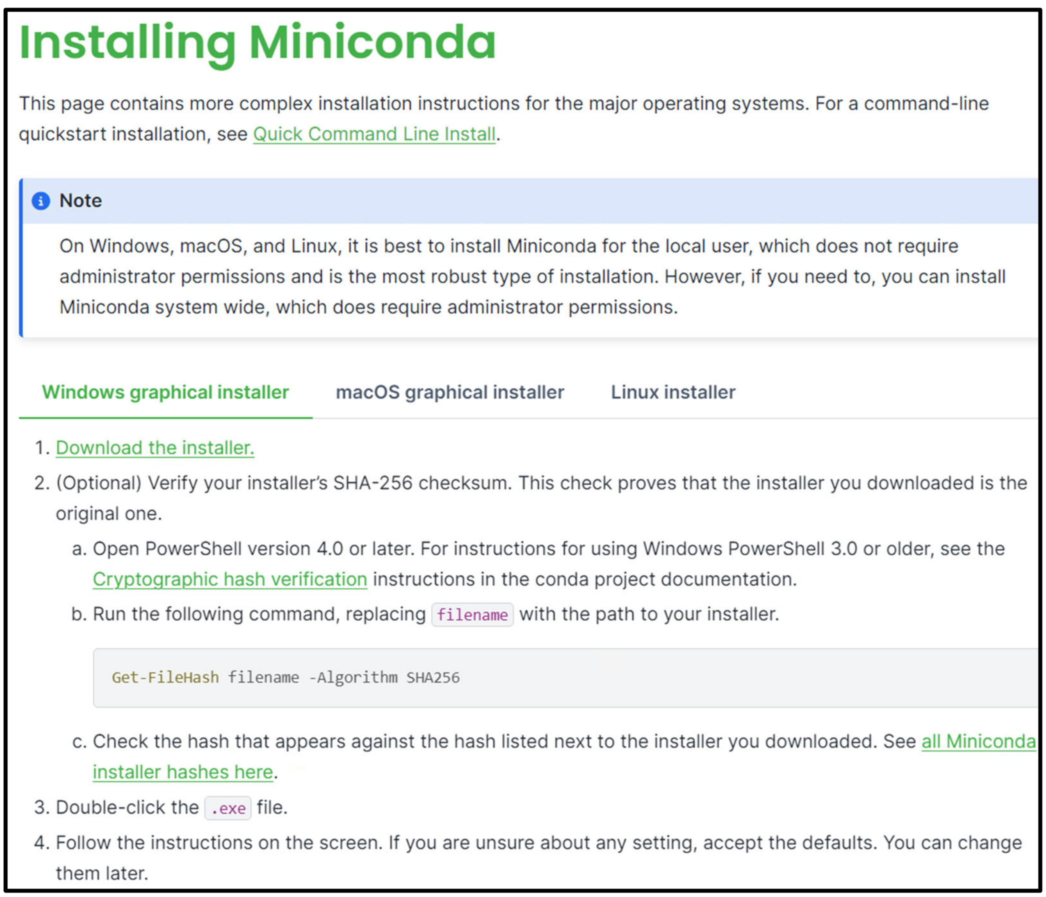
- Click the “Download” button for windows or other operating systems you have.
- Install the file following the installation wizard prompts. A shortcut of anaconda prompt will be placed in the start menu.

Option 2: Download and install miniconda:

- Use Google to search for “miniconda”



- Click on 'Installing Miniconda' to access the webpage where you can download the Miniconda installer. You will find the download link for the installation file for different operating systems along with installation guidelines.

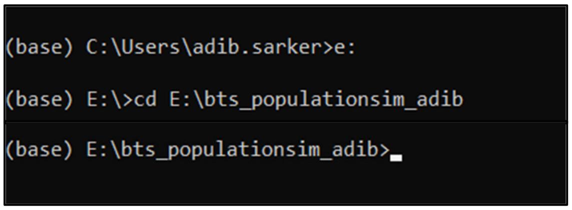


Run populationsim

Step 1: *Clone the bts_populationsim repository and navigate to that:*

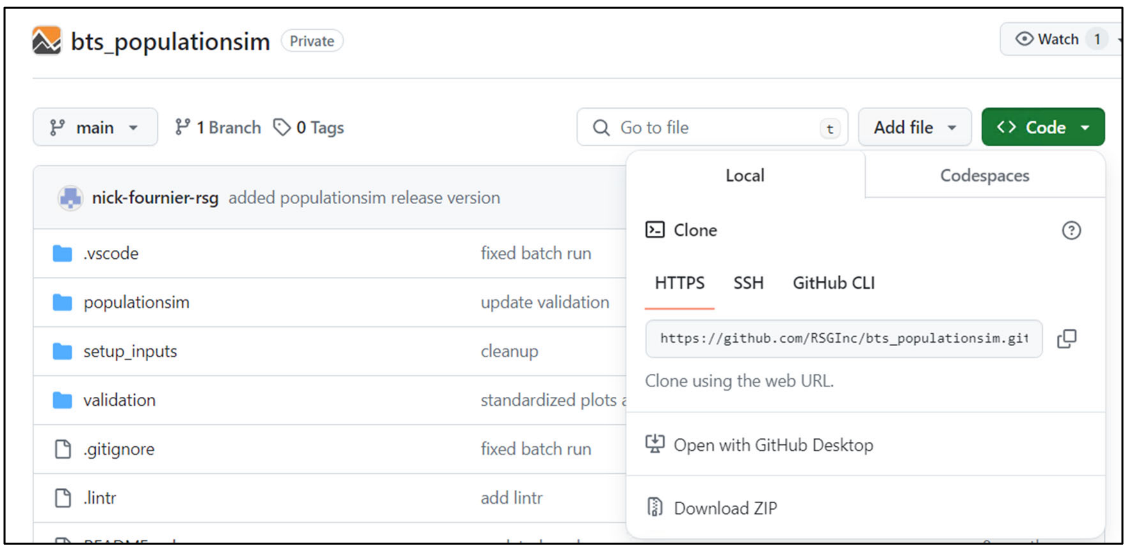
Open your Anaconda Prompt (or Miniconda)

Change the directory where you want to clone the repository



Copy the URL of the bts_populationsim GitHub repository:

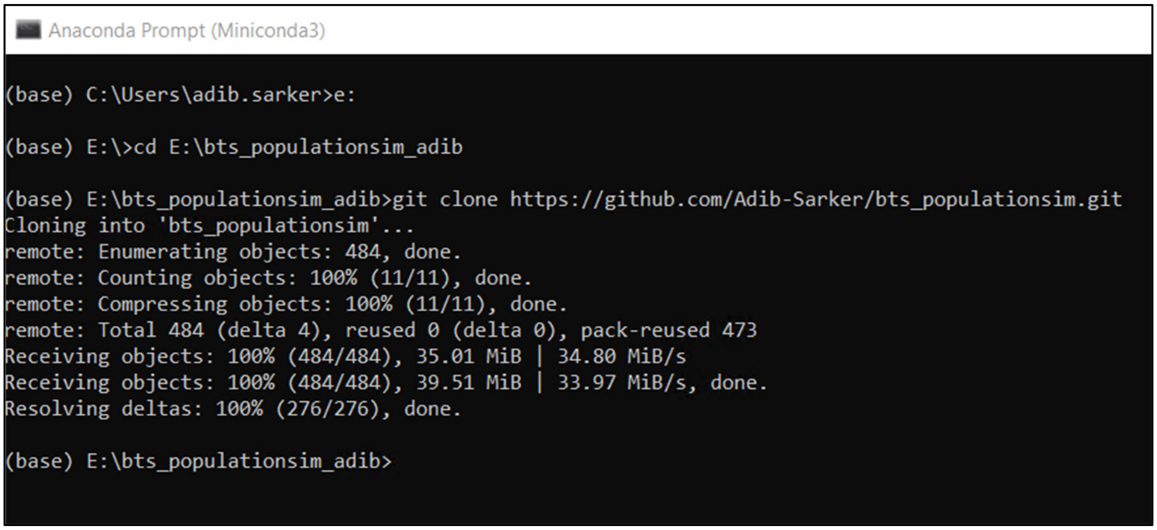
https://github.com/RSGInc/bts_populationsim.git

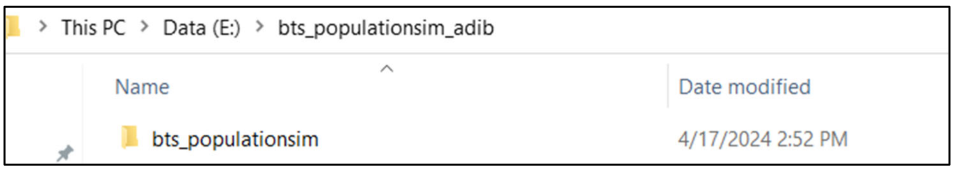


And clone the repository:

git clone https://github.com/RSGInc/bts_populationsim.git

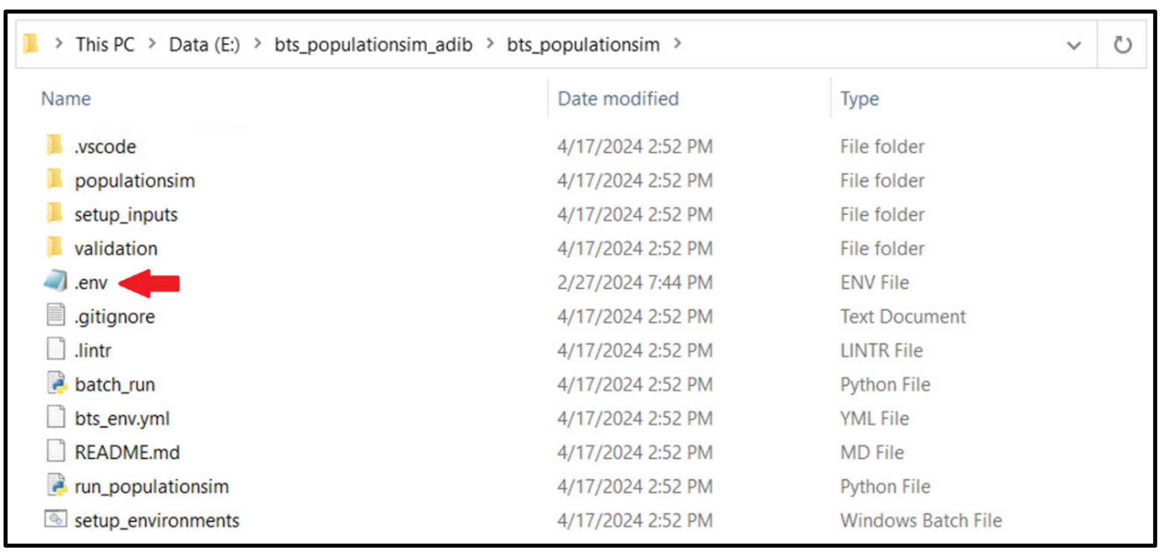
You will see the message “done” upon completion of the clone operation, and you will find the 'bts_populationsim' folder saved in the location you assigned.





Step 2: *Navigate to the bts_populationsim folder. Create a local env key file ".env" in your local directory with your Census API key like this:*

CENSUS_API_KEY=##########################



If you don't have a key, you can get one for free from the Census here:

https://api.census.gov/data/key_signup.html

Step 3: *Create environment*

You can install and run PopulationSim in default "base" python environment or you can set up new environment

Use the following command to set up the environment:

conda create --name my_env python=3.11.5

Here, name "my_env" to "bts_popS"; i.e., the screenshot below shows

conda create --name bts_popS python=3.11.5

You will see the message *"Proceed ([y]/n)?", type y and press "Enter"*

Anaconda Prompt (Miniconda3)

```
(base) E:\bts_populationsim_adib>conda create --name bts_popS python=3.11.5
Retrieving notices: ...working... done
Channels:
 - defaults
Platform: win-64
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##
```

Activate the environment using following command:

conda activate bts_popS

The "base" environment will change to "bts_popS"

Anaconda Prompt (Miniconda3)

```
Downloading and Extracting Packages:

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#     $ conda activate bts_popS
#
# To deactivate an active environment, use
#
#     $ conda deactivate


(base) E:\bts_populationsim_adib>conda activate bts_popS

(bts_popS) E:\bts_populationsim_adib>
```

Step 4: *Install (latest version) PopulationSim fork directly from GitHub*

Use pip that will install all dependencies and the forked version of PopulationSim to the current python environment.

(on anaconda prompt (miniconda) with python version 3.11.5) install to existing python environment using following command:

pip install git+https://github.com/nick-fournier-rsg/populationsim.git@v0.6.1#egg=populationsim

Anaconda Prompt (Miniconda3)

```
(base) E:\bts_populationsim_adib>conda activate bts_popS

(bts_popS) E:\bts_populationsim_adib>pip install git+https://github.com/nick-fournier-rsg/populationsim.git@v0.6.1#egg=p
opulationsim
```

Step 5: *Install required packages.*

Before installation, change a path to the bts_populationsim, as all the files required for PopulationSim run are in this directory.

```
cd bts_populationsim
```

Anaconda Prompt (Miniconda3)

```
(bts_popS) E:\bts_populationsim_adib>cd E:\bts_populationsim_adib\bts_populationsim

(bts_popS) E:\bts_populationsim_adib\bts_populationsim>
```

You can install the packages by running a code below:

```
pip install -r requirements.txt
```

Step 6: *Edit the controls specification file in "populationsim/configs/controls.csv"*

Which should look something like this…

target	geography	seed_table	importance	control_field	control_group	expression
h_total	BG	households	100000	H_TOTAL	h_total	(households.WGTP > 0) & (households.WGTP < np.inf)
p_total	BG	persons	100000	P_TOTAL	p_total	(persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_male	BG	persons	1000	P_MALE	p_sex	(persons.SEX == 1) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_female	BG	persons	1000	P_FEMALE	p_sex	(persons.SEX == 2) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_age_0_4	BG	persons	1000	P_AGE_0_4	p_age	(persons.AGEP <= 4) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_age_5_17	BG	persons	1000	P_AGE_5_17	p_age	(persons.AGEP >4) & (persons.AGEP <= 17) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_age_18_34	BG	persons	1000	P_AGE_18_34	p_age	(persons.AGEP > 17) & (persons.AGEP <= 34) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_age_35_49	BG	persons	1000	P_AGE_35_49	p_age	(persons.AGEP > 34) & (persons.AGEP <= 49) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_age_50_64	BG	persons	1000	P_AGE_50_64	p_age	(persons.AGEP > 49) & (persons.AGEP <= 64) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_age_65plus	BG	persons	1000	P_AGE_65PLUS	p_age	(persons.AGEP > 64) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_race_white	BG	persons	1000	P_RACE_WHITE	p_race	(persons.RAC1P == 1) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_race_black	BG	persons	1000	P_RACE_BLACK	p_race	(persons.RAC1P == 2) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_race_aapi	BG	persons	1000	P_RACE_AAPI	p_race	persons.RAC1P.isin([6,7]) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_race_other	BG	persons	1000	P_RACE_OTHER	p_race	~persons.RAC1P.isin([1,2,6,7]) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_hispanic	BG	persons	1000	P_HISPANIC	p_hispanic	(persons.HISP != 1) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
p_non_hispanic	BG	persons	1000	P_NON_HISPANIC	p_hispanic	(persons.HISP == 1) & (persons.PWGTP > 0) & (persons.PWGTP < np.inf)
h_size_1	TRACT	households	1000	H_SIZE_1	h_size	(households.NP == 1) & (households.WGTP > 0) & (households.WGTP < np.inf)

More in depth details for creating a control file are in PopulationSim's documentation:

https://activitysim.github.io/populationsim/application_configuration.html#controls

Step 7: *Edit the aggregate control targets in "populationsim/configs/controls_aggregator.csv"*

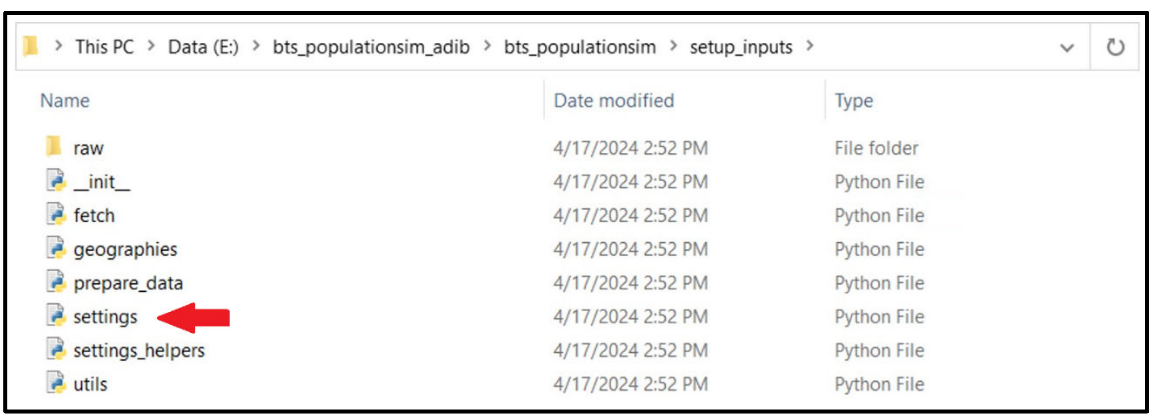
Which should look something like this…

field	original_label	concept	type	group	geography	control_field
B01001_001E	Estimate!!Total:	SEX BY AGE	int	B01001	BG	P_TOTAL
B01001_002E	Estimate!!Total:!!Male:	SEX BY AGE	int	B01001	BG	P_MALE
B01001_003E	Estimate!!Total:!!Male:!!Under 5 years	SEX BY AGE	int	B01001	BG	P_AGE_0_4
B01001_004E	Estimate!!Total:!!Male:!!5 to 9 years	SEX BY AGE	int	B01001	BG	P_AGE_5_17
B01001_005E	Estimate!!Total:!!Male:!!10 to 14 years	SEX BY AGE	int	B01001	BG	P_AGE_5_17
B01001_006E	Estimate!!Total:!!Male:!!15 to 17 years	SEX BY AGE	int	B01001	BG	P_AGE_5_17
B01001_007E	Estimate!!Total:!!Male:!!18 and 19 years	SEX BY AGE	int	B01001	BG	P_AGE_18_34
B01001_008E	Estimate!!Total:!!Male:!!20 years	SEX BY AGE	int	B01001	BG	P_AGE_18_34
B01001_009E	Estimate!!Total:!!Male:!!21 years	SEX BY AGE	int	B01001	BG	P_AGE_18_34
B01001_010E	Estimate!!Total:!!Male:!!22 to 24 years	SEX BY AGE	int	B01001	BG	P_AGE_18_34
B01001_011E	Estimate!!Total:!!Male:!!25 to 29 years	SEX BY AGE	int	B01001	BG	P_AGE_18_34
B01001_012E	Estimate!!Total:!!Male:!!30 to 34 years	SEX BY AGE	int	B01001	BG	P_AGE_18_34
B01001_013E	Estimate!!Total:!!Male:!!35 to 39 years	SEX BY AGE	int	B01001	BG	P_AGE_35_49
B01001_014E	Estimate!!Total:!!Male:!!40 to 44 years	SEX BY AGE	int	B01001	BG	P_AGE_35_49
B01001_015E	Estimate!!Total:!!Male:!!45 to 49 years	SEX BY AGE	int	B01001	BG	P_AGE_35_49
B01001_016E	Estimate!!Total:!!Male:!!50 to 54 years	SEX BY AGE	int	B01001	BG	P_AGE_50_64
B01001_017E	Estimate!!Total:!!Male:!!55 to 59 years	SEX BY AGE	int	B01001	BG	P_AGE_50_64
B01001_018E	Estimate!!Total:!!Male:!!60 and 61 years	SEX BY AGE	int	B01001	BG	P_AGE_50_64
B01001_019E	Estimate!!Total:!!Male:!!62 to 64 years	SEX BY AGE	int	B01001	BG	P_AGE_50_64
B01001_020E	Estimate!!Total:!!Male:!!65 and 66 years	SEX BY AGE	int	B01001	BG	P_AGE_65PLUS
B01001_021E	Estimate!!Total:!!Male:!!67 to 69 years	SEX BY AGE	int	B01001	BG	P_AGE_65PLUS
B01001_022E	Estimate!!Total:!!Male:!!70 to 74 years	SEX BY AGE	int	B01001	BG	P_AGE_65PLUS
B01001_023E	Estimate!!Total:!!Male:!!75 to 79 years	SEX BY AGE	int	B01001	BG	P_AGE_65PLUS
B01001_024E	Estimate!!Total:!!Male:!!80 to 84 years	SEX BY AGE	int	B01001	BG	P_AGE_65PLUS
B01001_025E	Estimate!!Total:!!Male:!!85 years and over	SEX BY AGE	int	B01001	BG	P_AGE_65PLUS
B01001_026E	Estimate!!Total:!!Female:	SEX BY AGE	int	B01001	BG	P_FEMALE
B01001_027E	Estimate!!Total:!!Female:!!Under 5 years	SEX BY AGE	int	B01001	BG	P_AGE_0_4
B01001_028E	Estimate!!Total:!!Female:!!5 to 9 years	SEX BY AGE	int	B01001	BG	P_AGE_5_17
B01001_029E	Estimate!!Total:!!Female:!!10 to 14 years	SEX BY AGE	int	B01001	BG	P_AGE_5_17
B01001_030E	Estimate!!Total:!!Female:!!15 to 17 years	SEX BY AGE	int	B01001	BG	P_AGE_5_17
B01001_031E	Estimate!!Total:!!Female:!!18 and 19 years	SEX BY AGE	int	B01001	BG	P_AGE_18_34
B01001_032E	Estimate!!Total:!!Female:!!20 years	SEX BY AGE	int	B01001	BG	P_AGE_18_34

You must specify the 'field' (from Census table), 'geography', and 'control field', *which should correspond to the 'control_field' in the controls.csv specification*. The setup script reads the 'control_field' e.g., P_AGE_5_17 to aggregate relevant census table fields.

Step 8: *Configure setup_inputs settings in "setup_inputs/settings.py" file.*

Go to the ".../bts_populationsim/setup_inputs" path



Open the "settings.py" python file using any IDE or Notepad++:

Parameters that might be useful to change:

- YEAR: the PUMS/ACS year to synthesize,
- STATES: a list of States by their abbreviation to be synthesized,

- BATCH_SIZE: the batch size in case you want to run states in chunks,
- PUMS_FIELDS: the fields to be synthesized,
- ACS_REMAINDERS: this supports the “controls_aggregator.csv” file if you want to create a field that is the difference between the total and another field, e.g., total number of non-commuters P_MODE_NA.

File Edit Selection View Go Run Terminal Help

settings.py 5

C: > RSG_PopS > BTS > bts_populationsim > setup_inputs > settings.py > ...

```python
1   import os
2   from pathlib import Path
3   from dotenv import load_dotenv
4   import pandas as pd
5   from us import states
6   import yaml
7
8   from setup_inputs.settings_helpers import aggregate_acs_fields
9
10
11  # Load .env file
12  load_dotenv()
13
14  # User-defined constants
15  YEAR = 2021
16  # POPSIM_DIR = 'C:/gitclones/bts_populationsim/populationsim'
17  POPSIM_DIR = os.path.join(Path(__file__).parent.parent.absolute(), 'populationsim')
18  SETUP_DIR = os.path.dirname(__file__)
19  RAW_DATA_DIR = os.path.join(SETUP_DIR, 'raw')
20
21  # ACS data | This just will list all states
22  STATES = [x.abbr for x in states.STATES]
23  ACS_TYPE = 'acs5'
24  BATCH_SIZE = 50
25
26  """
```

If your computer has limited memory, running PopulationSim for all 50 states at once would not be feasible. In such cases, it is recommended to run the simulation for each state, by changing the `BATCH_SIZE` to 1 to avoid memory errors.

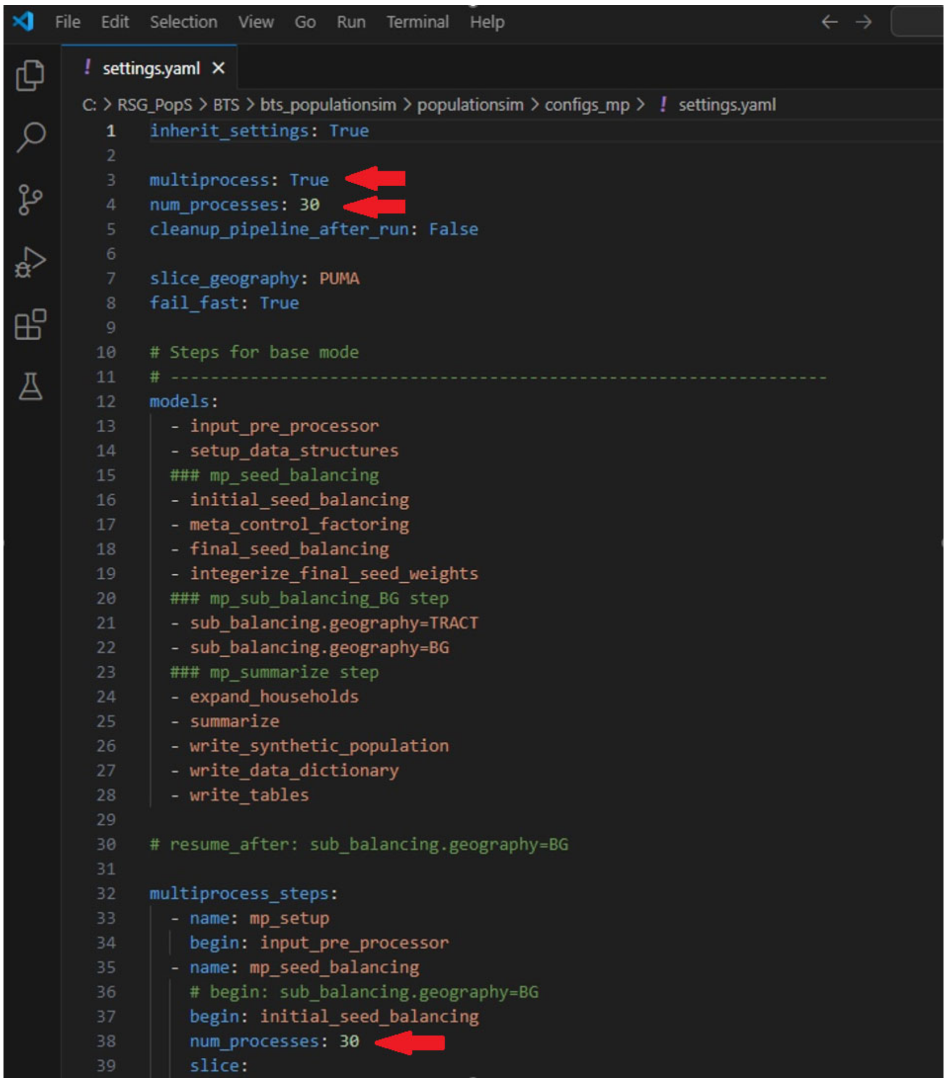
To run PopulationSim for specific state(s), specify the state names in line 22/23 in the `setting.py` file.

```python
21  # ACS data | This just will list all states
22  STATES = ['MT'] # Explicitly list specific states
23  # STATES = [x.abbr for x in states.STATES]
24  ACS_TYPE = 'acs5'
25  BATCH_SIZE = 1
```

***Step 9:** Configure multiprocessing in “populationsim/configs_mp/settings.yaml”*

Important parameters in the multi-process settings file are:

```
inherit_settings: True # This line will inherit settings from the base
configs/settings.yaml
multiprocess: True     # Change this to False if multiprocessing is not desired
num_processes: 30      # The number of cores to run parallel, also change the
“num_process” line below too if changing!
slice_geography: PUMA  # The geometry being "sliced" for parallel processing. I
do not recommend changing this or it might affect results.
```



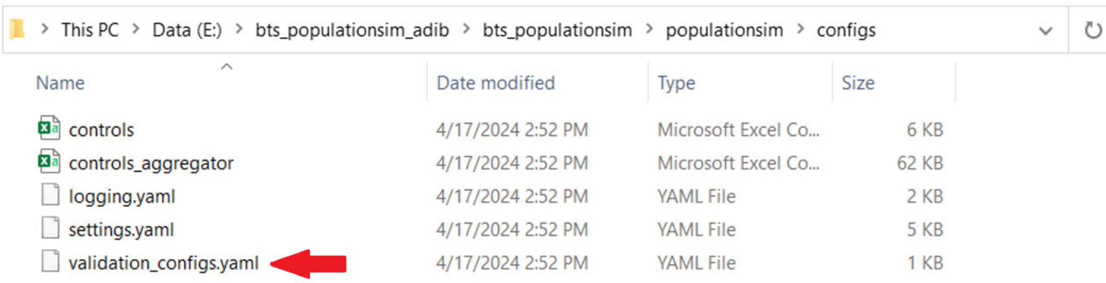
Multiprocessing reduces runtime but requires a large amount of RAM memory. For instance, 128 GB RAM memory is required to successfully run multiprocessing for the state of California.

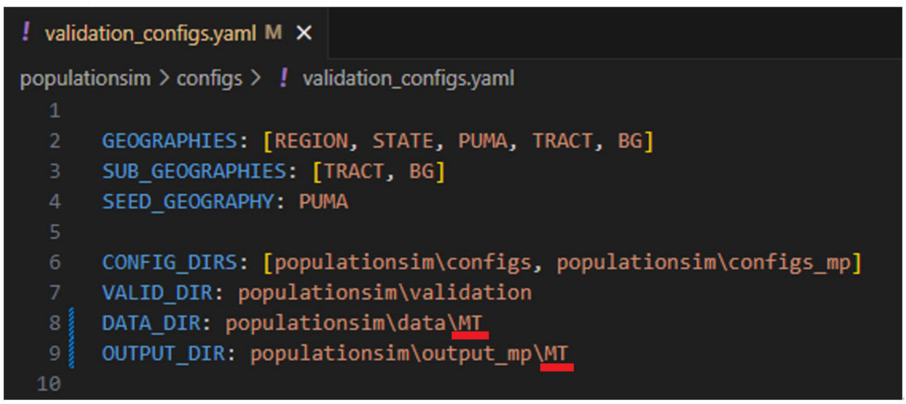
Small states such as AK, DC, DE, MT, ND, RI, SD, VT, and WY cannot be run on multiprocessors but can only be run on a single processor.

To run PopulationSim on a single processor (e.g., for small states), change the num_processes: 1 and set multiprocess: False.

***Step 10:** Edit validation configuration file:*

Modify a validation directory in the `validation_configs.yaml` file for your state of interest.





***Step 11:** Run the batch script:*

This batch_run.py file is in the bts_populationsim folder that was cloned already.

Change the directory to the folder where the file is located (if you haven't done in the previous steps.

Run the batch script by typing:

```
python -m batch_run
```

Adjust the 5-year ACS PUMS data to 1-year PUMS data

The aforementioned steps save the 2017-2021 5-year ACS PUMS data. The 1-year PUMS data adjustment process uses the 5-year PUMS data as the baseline input and the distribution of the 1-year PUMS data variables of interest as the distribution target. The following steps reconcile the 2021 5-year ACS PUMS data to the 2021 1-year PUMS data or the 2019 1-year PUMS data.

Step 1: *Copy the 2021 5-year data:*

- Run the first part (up to line 22) of the copy_folder.py

```
1   import os
2   import shutil
3   import us
4   from us import states
5
6   system_path = os.getcwd()
7   system_path_populationsim = os.path.join(system_path, 'populationsim')
8   main_folder_path = os.path.join(system_path_populationsim, 'data')
9
10  states_folders = [state.abbr for state in us.states.STATES] + ["DC"]
11
12  ### copy 2021 5 year data for future use
13  data_2021_5yr = os.path.join(main_folder_path, 'data_2021_5yr')
14  if not os.path.exists(data_2021_5yr):
15      os.makedirs(data_2021_5yr)
16
17  for state in states_folders:
18      source_path = os.path.join(main_folder_path, state)
19      destination_path = os.path.join(data_2021_5yr, state)
20
21      if os.path.exists(source_path) and os.path.isdir(source_path):
22          shutil.copytree(source_path, destination_path)
23
```

Step 2: *Run the hh_pop_adjust.py script:*

- Open the hh_pop_adjust.py file in any IDE/editor (VS Code, Notepad++, etc.)
- Change the year to 2021 (2019 for 2019 1-year data adjustment) in line 21

```
20  state_fips = '46' ## change the fips here
21  YEAR = '2021' ## change the year here
22
```

- Change the fips code of the state in line 20
- Run the whole script either in VS code or in anaconda.
- Run the script in anaconda (similar to populationsim) by typing:

```
python -m hh_pop_adjust
```

- Run the populationsim again with the adjusted dataset

Step 3: *Combine all states data into one:*

- Run the combine.py script

- Make sure the file name when saving.

```
expanded_household_ids.to_csv(os.path.join(system_path_populationsim, 'final_expanded_household_ids_2019_1yearadj.csv'), index=False)
```

```
seed_hh.to_csv(os.path.join(system_path_populationsim, 'seed_households_2019_1yearadj.csv'), index=False)
```

```
seed_per.to_csv(os.path.join(system_path_populationsim, 'seed_persons_2019_1yearadj.csv'), index=False)
```

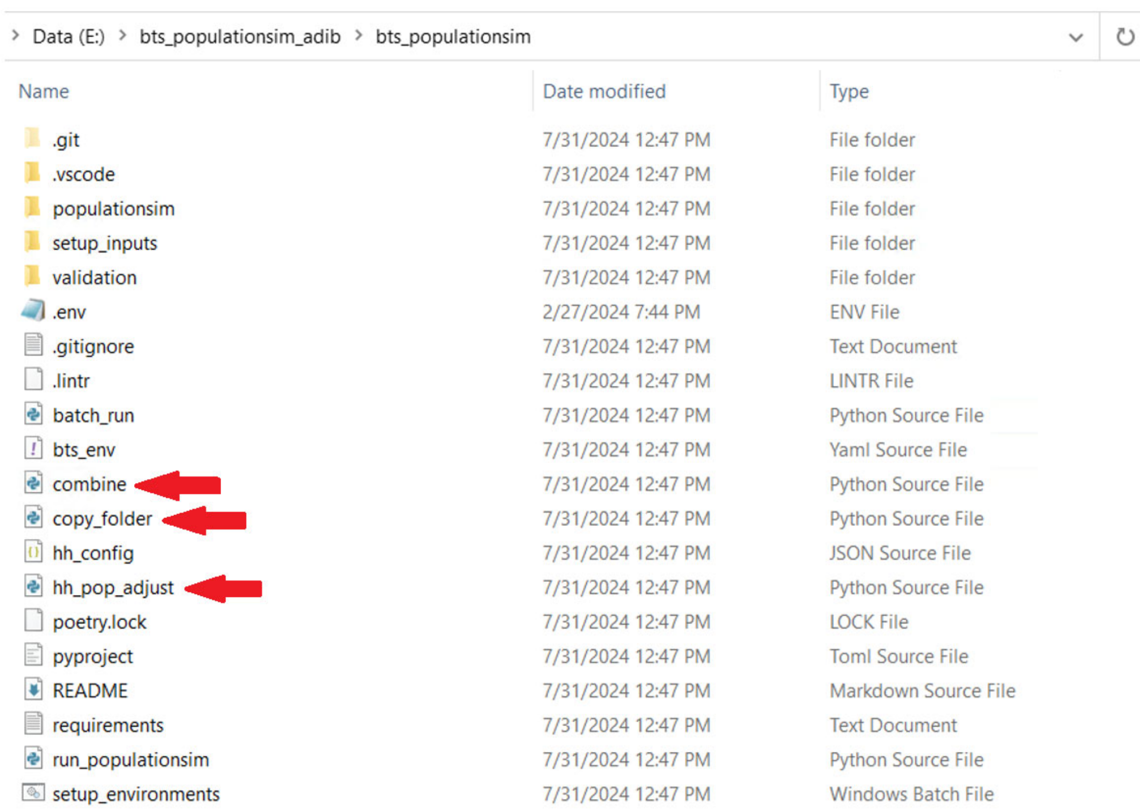
Step 4: *Copy all the 2021 adjusted dataset:*

- Run the second part (from line 27 to end) of the copy_folder.py

```
25  # ------------------_DO NOT RUN BELOW THIS LINE BEFORE YOU COMPLETE 2021 1-YEAR ADJUSTMENT_------------------ #
26  ### move 2021 1 year data for future use
27  data_2021_1yr = os.path.join(main_folder_path, 'data_2021_1yr')
28  if not os.path.exists(data_2021_1yr):
29      os.makedirs(data_2021_1yr)
30
31  for state in states_folders:
32      source_path_2021_1yr = os.path.join(main_folder_path, state)
33      destination_path_2021_1yr = os.path.join(data_2021_1yr, state)
34
35      if os.path.exists(source_path_2021_1yr) and os.path.isdir(source_path_2021_1yr):
36          shutil.move(source_path_2021_1yr, destination_path_2021_1yr)
37
38  ## again copy 2021 5-year data for 2019 1-year adjustment
39  for state in states_folders:
40      source_path_2019 = os.path.join(data_2021_5yr, state)
41      destination_path_2019 = os.path.join(main_folder_path, state)
42      if os.path.exists(source_path_2019) and os.path.isdir(source_path_2019):
43          shutil.copytree(source_path_2019, destination_path_2019)
```

Step 5: *Repeat step 2 and step 3 for 2019 adjustment*

Note: all the scripts are in "./bts_populationsim"



Data (E:) > bts_populationsim_adib > bts_populationsim

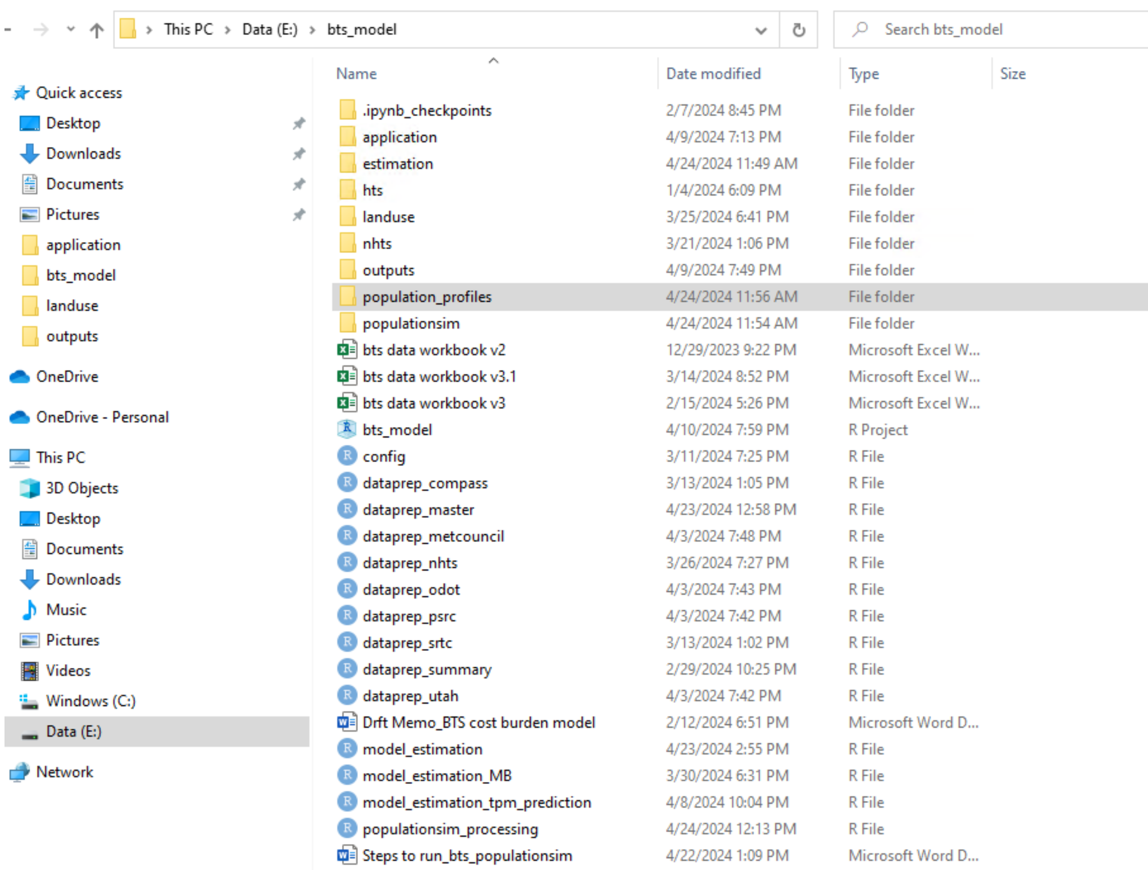
Name	Date modified	Type
.git	7/31/2024 12:47 PM	File folder
.vscode	7/31/2024 12:47 PM	File folder
populationsim	7/31/2024 12:47 PM	File folder
setup_inputs	7/31/2024 12:47 PM	File folder
validation	7/31/2024 12:47 PM	File folder
.env	2/27/2024 7:44 PM	ENV File
.gitignore	7/31/2024 12:47 PM	Text Document
.lintr	7/31/2024 12:47 PM	LINTR File
batch_run	7/31/2024 12:47 PM	Python Source File
bts_env	7/31/2024 12:47 PM	Yaml Source File
combine	7/31/2024 12:47 PM	Python Source File
copy_folder	7/31/2024 12:47 PM	Python Source File
hh_config	7/31/2024 12:47 PM	JSON Source File
hh_pop_adjust	7/31/2024 12:47 PM	Python Source File
poetry.lock	7/31/2024 12:47 PM	LOCK File
pyproject	7/31/2024 12:47 PM	Toml Source File
README	7/31/2024 12:47 PM	Markdown Source File
requirements	7/31/2024 12:47 PM	Text Document
run_populationsim	7/31/2024 12:47 PM	Python Source File
setup_environments	7/31/2024 12:47 PM	Windows Batch File

Data Processing and Modeling

This README provides an overview of the data processing and modeling workflow implemented in this project. Please review the following sections to understand the code structure and execution steps.

File structure:

- The project is organized into several R script files, each responsible for processing data and modeling. The data processing workflow is organized into multiple R scripts, each responsible for processing data from specific survey and an additional central file. The model estimation file provides scripts to estimate the model coefficients for this project using HTS and census block group level land use data. The populationsim processing file contains scripts to apply the model to household level synthetic population data to create household profile tables.



dataprep_master.R file:

- The **dataprep_master.R** file serves as the central script coordinating the travel survey data processing pipeline. You will see the following sections:

```
Library initialization
0. directory setup
4. add landuse variable
prepare land use variables to b...
1 Run NHTS Dataprep Script
2 Run HTS Dataprep Script
  2.4 PSRC Data
  2.5 ODOT Data
  2.6 COMPASS Data
  2.7 SRTC Data
  2.8 UTAH Data
3 Merge Data
```

- **Library initialization:** Initialize required libraries.

```r
# Library initialization ---------
if(!require('tidyverse')) install.packages("tidyverse")
if(!require('scales')) install.packages("scales")
if(!require('sf')) install.packages("sf")
if(!require('leaflet')) install.packages("leaflet")
```

- **Directory setup:** Set up the directory structure for data storage.

```r
# 0. directory setup ----------
SYSTEM_PATH = getwd()
SYSTEM_GIS_DATA_PATH  = file.path(SYSTEM_PATH, "gis")
SYSTEM_RHTS_DATA_PATH = file.path(SYSTEM_PATH, "hts")
SYSTEM_NHTS_DATA_PATH = file.path(SYSTEM_PATH, "nhts")
SYSTEM_LANDUSE_DATA_PATH = file.path(SYSTEM_PATH, "landuse")
SYSTEM_OUTPUTS_PATH = file.path(SYSTEM_PATH, "outputs")
SYSTEM_PROFILES_PATH = file.path(SYSTEM_PATH, "population_profiles")
SYSTEM_APP_PATH = file.path(SYSTEM_PATH, "application")

```

- **Add land use variables:** Add land use variables to the dataset.

```r
# 4. add landuse variable============
sld_21    <- fread(file.path(SYSTEM_LANDUSE_DATA_PATH, "sld2021extended.csv"), colClasses = "numeric")
names(sld_21) = toupper(names(sld_21))
sld_19    <- fread(file.path(SYSTEM_LANDUSE_DATA_PATH, "sld2019extended.csv"), colClasses = "numeric")
```

- **Prepare land use variables:** Prepare land use variables to add to the dataset.

```r
# prepare land use variables to be added to pop data (merge to populationsim data. Expanded_HHID data---
bg2021_keyvars <- sld_21[, .(GEOID20, D4C, D3B,D1B, TRANSIT_LEVEL2, D5AR,D5AE, D5BR, D5BE, NATWALKIND, CB
names(bg2021_keyvars)
bg2021_keyvars[D5BR == -99999, D5BR := NA]; bg2021_keyvars[D5BE == -99999, D5BE := NA]
```

- **Run survey data preparation scripts:** Execute each survey data preparation script (e.g., dataprep_nhts) using *'source()'* function calls.

```r
  # 1 Run NHTS Dataprep Script--------------------
  source("dataprep_nhts.R")
  all_hh_day_nhts[,.N, SURVEY]
  all_daytrip_nhts[,.N, SURVEY]

  # 2 Run HTS Dataprep Script --------------------

  ## 2.3 Metcouncil
  source("dataprep_metcouncil.R")
  all_hh_day_metcouncil[, SURVEY:=3]
  all_daytrip_metcouncil[, SURVEY:=3]

  ## 2.4 PSRC Data -----------------
```

- **Merge data:** Merge all the processed survey data and land use data.

```r
  # 3 Merge Data ------------
  # hh level

  all_hh_day <- bind_rows(all_hh_day_nhts, all_hh_day_metcouncil) %>% bind_rows(all_hh_day_psrc) %>% bind
    bind_rows(all_hh_day_srtc) %>% bind_rows(all_hh_day_utah)

  # trip level data
  all_daytrip_hh <- bind_rows(all_daytrip_nhts, all_daytrip_metcouncil) %>% bind_rows(all_daytrip_psrc) %
    bind_rows(all_daytrip_srtc) %>% bind_rows(all_daytrip_utah)

```

```r
# merge split data into one
all_hh_day_lu <- bind_rows(all_hh_day_bg10_lu, all_hh_day_bg20_lu)
all_hh_day_lu <- data.table(all_hh_day_lu)
setcolorder(all_hh_day_lu, colnames_new)
```

Survey data preparation files:

- Individual survey data preparation scripts (dataprep_nhts.R, dataprep_metcouncil.R, dataprep_psrc.R, dataprep_odot.R, dataoreo_srtc.R, dataprep_compass.R, and dataprep_utah.R) contain code specific to cleaning and processing each survey data.
- These scripts are sourced within the dataprep_master.R file to integrate survey-specific processing steps into the workflow.

Model estimation file:

- The **model_estimation.R** file provides codes to build the model for this project. It shows the model function and variables of the processed survey dataset.
- The **dataprep_master.R** file is sourced within this script to retrieve the processed survey data for model development.

Populationsim Processing file:

- The script provides the steps to apply estimated BTS cost burden model to household level synthetic population data and calculate cost burdens by defined set of household profiles and geographies, e.g., census tract, county, state, and national.
- The output of this script is the final cost burden model output table, which will be an input for the online visualization tool.

Dataprep Summary file:

- The **dataprep_summary.R** file provides data descriptive statistics of the processed survey dataset.
- The **dataprep_master.R** file is sourced within this script to retrieve the processed data.

Appendix B: Model Estimation Results

As described in the main body of this report, seven different models of miles traveled per household day were estimated based on the combined Household Travel Survey data. The models are:

1. Household vehicle miles travelled (VMT) for work trips
2. Household vehicle miles travelled (VMT) for non-work trips
3. Household vehicle miles travelled (VMT) for long-distance trips over 100 miles
4. Miles traveled using public transportation for work trips
5. Miles traveled using public transportation for non-work trips
6. Miles traveled via taxi & ride-hailing (TNC) services for work trips
7. Miles traveled via taxi & ride-hailing (TNC) services for non-work trips

The three main modes (household vehicles, public transportation, and taxi & ride-hailing) are modeled separately because they tend to have different relationships to explanatory variables such as income, vehicle availability, and land use. Travel for trips to or from work is modeled separately from non-work travel because the two purposes tend to have different relationships with other explanatory variables such as employment status and telecommute status. Long distance auto travel (trips over 100 miles) tends to be non-typical trips that are more difficult to explain and predict, so those are modeled separately. Public transit and taxi & ride-hailing trips over 100 miles are exceedingly rare and were not included in the modeling.

As introduced in the main body of the report, a two-part model was estimated for each of the seven model types above. The two-part model form is particularly effective at handling mixed discrete-continuous random variables, accommodating instances where households report zero VMT, no transit trip miles, or no taxi & ride-hailing trip miles. The first part of the model predicts the probability of a household traveling any miles at all for the relevant mode/purpose combination, while the second part of the model predicts the number of miles traveled for household with a non-zero number of miles. The structure of the two-part model is,

$$E[Y \mid X] = Pr(Y>0 \mid X) * E(Y \mid Y>0, X]$$

The first part of the model, $Pr(Y>0 \mid X)$ uses a binary logit model specification, while the second part, $E(Y \mid Y>0, X]$ uses a log-transformed linear least-squares regression specification. The log transformation is used because the distribution of miles traveled per household-day (excluding the 0 cases) tends to have a log-normal shape, skewed to the left with a tail extending to the right.

Given an array X of explanatory variables and arrays of coefficients b(1) and b(2) for the two parts of the model, the predicted number of miles traveled is equal to the binary logit probability of traveling any miles at all multiplied by the predicted number of miles traveled conditional on it being greater than zero:

Miles/Household-day(predicted) = [EXP (b(1).X) / (1 + EXP (b(1).X)] * EXP (b(2).X)

Although it is possible to estimate the two parts of each model (the binary logit model and the log-liner regression model) separately, the estimation method used in this project uses an iterative procedure across both parts to maximize the joint likelihood of the observed mileage data against the predicted outcome from both model components together.

The estimation results for the various models are shown in the tables at the end of this appendix.

Auto vehicle miles traveled for work trips

Table B-1 shows the first model for household-vehicle miles traveled to and from work. The top rows show the number of survey choice observations used in estimation as well as the mean values of the dependent variables and the fit of the estimated model. All seven models used nearly 311,000 household-days of data, aggregated across the various household travel surveys. Among the household-days of data used, roughly 130,000 (41.5%) had miles traveled greater than 0, which became the basis for the second part of the model. Reasons that over half the household-days had no auto trips for work may include: (a) there are no workers in the household, (b) the observed day was not a workday, (c) any workers in the household worked from home that day, and/or (d) any trips to or from work were made using other modes of travel. The rho-squared (fraction of likelihood explained) model fit measure for the binary logit part of the model is 0.38, which is quite high for a disaggregate discrete choice model. The r-squared measure for the log-linear regression part of the model is 0.117. The exact mileage is more difficult to predict than a binary 0 vs. non-0 choice, particularly without using any data regarding where the household members work or other land use measures beyond those related to the residence block group.

The first two columns in the table give the variable name as used in estimation, as well as a more detailed description of each variable. The last four columns give the estimated coefficients and related t-statistics for the Part 1 and Part 2 models respectively. The t-statistic is the estimated coefficient divided by the standard error of the estimate. The higher the absolute value of the t-statistic, the greater the probability that the estimated coefficient is significantly different from 0. A t-statistic with absolute value of 1.96 or higher has more than a 95% probability of being significantly different from 0.

Because different explanatory variables use different units (some are 0/1 dummy variables, some are count variables, and some are continuous variables of various types), the magnitude of the coefficients do not directly indicate how important each variable is in predicting the mileage outcome. The t-statistics are the strongest indicators of the importance of the variables. For example, the NO_WORKERS variable has a t-statistic of over 100, and, not surprisingly, is the strongest predictor of whether or not a household generates any auto travel to or from work. Also not surprisingly, the NO_VEHICLES variable is the second strongest predictor, with a t-statistic of almost 50. T-statistics generally increase in proportion to the square root of the number of observations used in the model, so having over 300,000 household-day observations from the various surveys allows the estimation of several very significant explanatory effects. The coefficients of NO_WORKERS and NO_VEHICLES are not very significant in the Part 2 model, since those variables have a value of 0 for most of the observations with work VMT greater than 0.

In other words, some variables are more useful in explaining whether or not a household uses a given mode for a given type of trip at all, while others are more useful in explaining how many miles a household travels during a day. For example, income strongly explains the use of a given mode for a given trip in both parts of the model but has a stronger effect in the Part 2 model, with higher income households tending to travel longer distances. A log transform of income gave a better model fit than using a linear income variable. This is often the case in travel models, since budget constraints on travel have more effect in lower income ranges and income itself tends to have a log-normal distribution across the population. In addition to the main LOG_INCOME variable, the model includes an additive variable for households with income below the Federal poverty level. This variable has a significant negative coefficient for the Part 1 model, indicating that those with very low incomes often avoid driving to work, perhaps sharing a ride with others or using other modes such as bicycle or walking.

In most surveys, about 10% of households decline to report their income. Rather than excluding such cases from the model estimation, the Department included a MISSIMG_INCOME 0/1 dummy variable. This type of so-called “nuisance variable” allows the model to be estimated without the missing income data biasing the estimates of the other income-related coefficients.

The variables following income are related to the composition of the household in terms of auto sufficiency, work hour and telecommute status, and age groups. Auto sufficiency is statistically important in both parts of the model. The adults without a vehicle (ADULTS_WOUTVEH) variable indicates that households with one or more vehicles but fewer vehicles than adults generate less work VMT, all else equal. Full-time and part-time workers who usually work out of home have strong positive coefficients in both model parts, with full-time workers generating more distance than part-time workers, as they are more likely to be going to work on a given day. Workers who usually work from home generate less work VMT greater (Part 1 model), but if they do travel for work, they add positively to the distance traveled (Part 2 model).

Having children in the household—particularly school age children—corresponds to a lower probability of having positive work VMT. The age-related variables for adults are relative to a base age group of 35-54. The more adults that are younger than 35 or older than 54, the less chance of generating work VMT in both parts of the model, all else equal. The effect is particularly strong for adults age 65+, who evidently travel to work fewer days and shorter distance than workers of similar status in other age groups.

The diary data (DIARYDATA) variables in Appendix B show the expected direction of bias correction, as diary recall surveys are less successful than smartphone-based surveys in capturing all trips people actually make. RSG’s rMove travel survey app was used for all of the non-NHTS surveys used in this study.

The next set of variables in the table attempt to capture land-use and network effects at the block group level. No block group (NOBG), like missing income (MISSING-INCOME), is a "nuisance variable" that allows observations with missing data on the block group ID (some of the NHTS data) to be included in estimation without biasing the estimates of the land use coefficients. The land use variables come mainly from the EPA Smart Location Database (SLD). Many are significant with the expected sign. Higher walkability, higher street intersection density (for walking and biking), higher population density and higher transit frequency are all related to a lower probability of having any auto mileage for work and usually for shorter distance in Part 2. The best measures of surrounding land use accessibility by auto and transit are the LND5ABOTH and LND5BBOTH variables which measure jobs plus population over 18 within a 45-minute drive and within 45-minutes by transit respectively. These do not show strong effects on the zero vs non-zero choice but do show strong distance effects in the Part 2 model in the expected directions. The transit level (TLEVEL) variables for lack of transit service or data show a counter-intuitive effect in the Part 1 model; namely, that the lack of transit service leads to a lower probability of positive work VMT. This may be due to correlation with the transit-related SLD variables. The stronger effects in the Part 2 model are in the expected direction, however—the less likely that transit service is available, the higher the auto distance traveled.

Land use variables tend to show a high degree of correlation (collinearity) in modeling, so being able to estimate several significant coefficients in the expected direction is quite satisfactory compared to many modeling efforts. The final region-specific effects show that some areas such as New York City, Washington DC, Boston and San Francisco have a somewhat lower share of households with positive work auto mileage than can be predicted using the set of SLD variables. It is possible that the accessibility via the transit system in those regions is greater than what can be represented using the nationwide SLD transit variables.

The last sets of variables capture day-of-week, month-of-year, and year-specific differences in behavior. The models include day-of-week and month-of-year coefficients to ensure that the models represent all days of the year and give a representative total annual result. The year-specific effects allow the models to be applied for years other than the "base year" of

2021. As one might expect, all of the pre-COVID years show strongly positive differences in VMT in both parts of the model, while 2020 has the most negative trend effect.

Auto vehicle miles traveled for non-work trips

Table B-2 has the same format as Table B-1 because the models use the same variables. Rather than giving such a detailed description of results as was given for the first model (above), this section points out some key differences between the work and non-work VMT models.

Over 73% of the household-days used in estimation had non-zero VMT for non-work trips, allowing over 225,000 household-days to be used in estimating the Part 2 model. The average VMT for household-days with non-zero values was just under 30 miles, very similar to the average value of 32.5 miles Part 2 of the work VMT model.

Not surprisingly, the largest difference is for the no workers (NO_WORKERS) variable, which had a huge negative effect in the work model, but has a strong positive effect in both parts of the non-work model. The no vehicles (NO_VEHS) variable, on the other hand, has an even stronger negative effect in the non-work model, with a t-statistic of almost -85.

The log of income (LOG_INCOME) has highly significant positive coefficients in both parts of the model. The additive variable for incomes below poverty level is still somewhat significant in generating fewer days with non-zero VMT, but much less significant than in the model for work trips. The income below poverty level (INCBPOV) variable was tested for the transit and taxi & ride-hailing models described below, but was not found to be significant, so was not included in any of the other models.

More full-time workers working out of home corresponds to lower non-work VMT, while more part-time workers working either out of home or at home corresponds to a higher probability of generating non-work VMT. Full-time workers usually working from home have a positive coefficient in the Part 1 model, in contrast to the negative coefficient for full-time workers usually working out of home.

The presence of children in both age ranges increases the probability of positive non-work VMT, while the number of children of school age has a very positive distance effect in the Part 2 model. Having adults in the older age categories (relative to age 35-54) is related to a positive probability of generating non-work VMT in the Part 1 model, but small distance effects in the Part 2 model.

The diary data (DIARYDATA) bias term is even stronger in both parts of the model for non-work trips than for work trips. This likely arises from trip non-reporting bias being more severe for discretionary travel purposes.

The rest of the variables in the model show similar results for the non-work VMT model as for the work VMT model, although the region-specific variables are somewhat less significant, particularly for Boston and San Francisco.

Auto vehicle miles traveled for long-distance trips

Whether or not a household generates long-distance auto trips on a given day is much less systematic and difficult to predict than more regular short-distance travel. Table B-3 shows that only 2.1% of households generated any auto trips of more than 100 miles, giving only 6,682 observations for the Part 2 model (which had an average observed distance of just under 200 miles per household-day). The model fit for both parts of the model is lower than for the other two VMT models.

In model testing, most of the land use variables related to the residence block group were not significant, so the Department dropped them from the final model. The strongest positive coefficients for the Part 1 model of generating any long-distance auto trips are the income variable and the variable for the number of adults with a vehicle, as well as weekend days (FRI, SAT, SUN).

There are very few significant coefficients in the Part 2 model to predict the exact distance of long-distance trips.

Because of the low frequency of these types of trips compared to shorter-distance, long-distance trips contribute much less to total VMT and cost burden than the work and non-work trips under 100 miles used to estimate the two previous models. Thus, the lower fit and explanatory power of this model is not of great concern.

Miles traveled for work trips by public transit

Table B-4 shows that 2.5% of the household-days have non-zero distance by public transit for work trips, yielding 7,776 observations for the Part 2 model with an average of just over 18 miles traveled. The model fits are relatively good, particularly for the Part 2 regression model. Note that in this model and the remaining models described below, the percent of household-days with non-zero mileage is only a small fraction of the sample, so the Part 1 model explaining zero vs. non-zero mileage is more important for the predicted total mileage than the Part 2 model explaining the variation across the relatively few non-zero cases.

As expected, the no workers in the household (NO_WORKERS) variable has a large negative coefficient, and the coefficients related to vehicle ownership are very significant with a sign opposite from the sign in the models for auto VMT.

The income variable is positive in both parts of the model, suggesting that lower income households may try to find an even less expensive alternative for commuting, such as ride-hailing or cycling.

The more workers, either full-time or part-time, who usually work out of home, the higher the chance of generating transit mileage for work trips. The more children and older adults over 54 in the household, the less chance of traveling any miles by transit for work.

The variables related to land use generally show significant effects with a sign opposite from the sign in the models for auto mileage. The variables related to higher walkability are positive for transit use, as most transit trips involve walking for some portion of the trip. The transit level (TLEVEL) variables for low transit availability have the expected negative sign. All of the CBSA-specific variables show very positive effects in Part 1, and some also in Part 2-the NYC variables in particular.

Interestingly, this is the only model where the diary data (DIARYDATA) variable shows a significant positive bias towards work trips by transit for diary-based surveys compared to smartphone-based surveys. The non-work model for transit shows only an insignificant positive effect.

All of the years other than 2021 have a positive coefficient relative to 2021 in Part 1 of the model, suggesting that 2021 was the low point in terms of the effect of COVID-19 on transit use. There was very little survey data for 2020, so the positive effect for that year is questionable.

Miles traveled for non-work trips by public transit

Table B-5 shows that the percentage of household-days with positive transit milage for non-work trips is higher than for work trips (3.9% vs. 2.5%), although the average distance for the non-zero cases is somewhat less for non-work trips than for work trips (15.14 vs. 18.07). The model fit for the non-work trip transit mileage is somewhat less than for the work trip transit mileage (Table B-4), although still somewhat better than for the auto non-work mileage model (Table B-2).

The variables related to vehicle ownership are very strong and in the expected direction—even more significant than in the transit to work model. In contrast to the transit work model, however, higher income has a negative effect on using transit for non-work trips.

Having more workers in the household corresponds with a higher chance of generating transit non-work travel, with the exception of full-time workers who usually work from home. Having more children under age 5 or more adults over age 54 decreases the chance of making non-work trips by transit but having more school-age children somewhat increases the chance.

The land use variables show a similar pattern for transit non-work trips as for transit work trips (Table B-4), with the CBSA-specific coefficients somewhat less significant but still important, with NYC again the highest.

Non-work transit use is less common on weekends (FRI,SAT,SUN), while the model shows the opposite for non-work auto use (Table B-2). This may be in part due to lower transit service levels on weekends, as well as the mix of trip purposes and destinations that transit typically serves.

The year-specific variables indicate relatively low transit use for non-work trips in 2020 and 2022 similar to 2021, but with signs of a rebound in 2023. Note that surveys in different years are from different regions and the NHTS 2022 sample size is quite small, so it is not possible to generalize these trends to all regions with the data available.

Miles traveled for work trips by taxi & ride-hailing

Table B-6 shows that only 0.5% of household-days had any taxi & ride-hailing miles for work trips, giving only 1,482 observations for the Part 2 model, with an average distance traveled of 9.28 miles. The model fit for taxi & ride-hailing is somewhat lower than for the corresponding auto and transit models, but is still in an acceptable range for disaggregate choice models, with several significant coefficient estimates.

As with the transit models, the no worker (NO_WORKER) coefficient remains negative, and the coefficients related to vehicle ownership are significant with the expected signs. The income effect is positive, as one would expect for taxi & ride-hailing being higher-cost alternatives to transit.

Households with more workers in all categories show a higher chance of using taxi & ride-hailing for work, particularly among workers who usually work out of home. The number of children or adults under age 35 do not significant affect use, but the number of adults age 55-64 and 65+ decreases the chance of using taxi & ride-hailing for work trips.

The various land use coefficients are mainly of little significance in this model compared to the models previously described. The specific CBSA regions do generate somewhat more taxi & ride-hailing travel for work, with the strongest effect in the Washington DC area.

The trend effects across years are not particularly strong, although the peak appears in 2019.

Miles traveled for non-work trips by taxi & ride-hailing (TNC)

The final Table B-7 shows that about 1% of household-days had positive taxi & ride-hailing mileage, with 3,019 observations to estimate the Part 2 model, with an average distance of 12.52 miles. The model fits are similar to those of the taxi & ride-hailing model for work trips (Table B-6).

The coefficients also appear similar those for the taxi & ride-hailing model for work trips for the most part (except for the no workers (NO_WORKERS) variable, of course). The income effect is again strongly positive, and the vehicle-related coefficients are significant and in the expected directions.

One difference is that workers who usually work from home appear to generate more non-work taxi & ride-hailing trips, while workers who usually work out of home do not. The presence of children also has a much more negative effect on the probability of non-work taxi & ride-hailing travel. As found for work travel, adults over age 54 appear less likely to use taxi & ride-hailing, but with the coefficients larger for non-work travel. The number of young adults age 18-34 has a strong positive relationship with using taxi & ride-hailing for non-work trips, possibly for meal and/or entertainment trips. There are positive coefficients for FRI and SAT relative to weekdays.

The land use effects are somewhat more significant than for taxi & ride-hailing for work trips, with high walkability and auto accessibility showing a positive relationship. These may be areas with better taxi & ride-hailing availability. There are also positive effects for NYC, Washington DC and San Francisco, three cities known for high usage of taxi & ride-hailing.

Like the model for taxi & ride-hailing for work trips (Table B-6), the trend effects show a positive peak in 2019.

Table B-1	Model of Work Trip Auto Vehicle Miles per Household-day	Part 1: Binary logit (Miles > 0)		Part 2: Regression LN(Miles)	
Observations	Number of survey household-days used in estimation	310922		129088	
Mean value	Fraction of non-0 values for Part 1, average of non-0 values for Part 2	0.415		32.49	
Model fit	McFadden rho-squared for Part 1, regression R-squared for Part 2	0.380		0.117	
Variable	**Description**	**Estimate**	**t value**	**Estimate**	**t value**
Constant	Intercept term	0.45337298	4.825	1.55005463	27.658
LOG_INCOME	LN(Max(Household income,0) + 1) (0 if data missing)	-0.04061821	-4.735	0.1025675	20.073
INCBPOV	1 if household income is below Federal poverty level, otherwise 0	-0.49903131	-16.652	0.01293946	0.669
MISSING_INCOME	1 if income data is missing, otherwise 0	-0.59033843	-5.946	1.10714205	18.611
NO_VEHS	1 if there are no vehicles in household, otherwise 0	-2.69097359	-46.029	0.1126949	2.018
ADULTS_WITHVEH	Lower of # adults and # vehicles in household if # vehicles>0, otherwise 0	0.12610943	9.794	0.1271384	17.046
ADULTS_WOUTVEH	Max(# adults minus # vehicles,0) if # vehicles >0, otherwise 0	-0.47148105	-31.898	-0.12337259	-13.567
NO_WORKERS	1 if there are no working adults in household, otherwise 0	-3.26544159	-114.594	0.0024527	0.097
N_FTWORKAWH	# of full-time workers in household who usually work out of home	0.75895062	61.457	0.23855645	33.55
N_PTWORKAWH	# of part-time workers in household who usually work out of home	0.66052549	36.831	0.11306209	11.639
N_FTWORKATH	# of full-time workers in household who usually work at home	-0.33555082	-20.109	0.11422021	10.633
N_PTWORKATH	# of part-time workers in household who usually work at home	-0.24336794	-9.522	0.11698673	6.856
N_AGE_04	# of children age 0 to 4 in household	-0.05528981	-4.77	0.01180797	1.812
N_AGE_517	# of children age 5 to 17 in household	-0.11560338	-16.639	0.00112413	0.284
N_AGE_1834	# of adults age 18-34 in household (relative to age 35-54)	-0.05218667	-6.165	-0.01152032	-2.45
N_AGE_5564	# of adults age 55-64 in household (relative to age 35-54)	-0.03774823	-4.054	-0.02862929	-5.671
N_AGE_65UP	# of adults age 65 and older in household (relative to age 35-54)	-0.32166993	-28.515	-0.09715642	-14.002
DIARYDATA	Data collected using diary recall (compared to smartphone)	-0.17364358	-11.831	-0.05112431	-5.903
NOBG	1 if block group ID not available in the data, otherwise 0	-0.34245567	-10.56	0.24861758	12.921
NATWALKIND	National walk index for block group (from SLD)	-0.01140529	-4.168	-0.0122879	-7.611
D3B	Street intersection density, (SLD)	-0.00046993	-5.171	-0.00074807	-12.839

D1B	Population density on unprotected land (SLD)	-0.00552372	-10.981	-0.00133801	-4.028
D4C	Aggregate frequency of transit within 0.25 mile in evening peak (SLD)	-0.0021463	-13.43	-0.00012691	-1.139
LND5ABOTH	Jobs plus pop. age 18+ within 45 min by car, decay-weighted (SLD)	-0.00516942	-1.609	0.02808396	14.855
LND5BBOTH	Jobs plus pop. age 18+ within 45 min by transit, decay-weighted (SLD)	-0.00843018	-3.694	-0.00577654	-4.441
TLEVEL_2	Block group has transit stops, but no transit data in SLD	-0.10851723	-3.797	-0.04057296	-2.46
TLEVEL_3	Block group may have transit, but no transit data in SLD	-0.04996231	-2.335	0.10733291	9.011
TLEVEL_4	Block group likely has no transit and no transit data in SLD	-0.1237529	-5.503	0.21316366	16.872
NYCCBSA	CBSA is in New York City area	-0.1800451	-4.361	0.17187265	7.145
CHICBSA	CBSA is in Chicago area	0.08912269	0.806	0.05040267	0.836
WASCBSA	CBSA is in Washington DC area	-0.40519413	-3.977	0.06146107	0.941
BOSCBSA	CBSA is in Boston area	-0.39383085	-2.812	0.13762156	1.619
SFOCBSA	CBSA is in San Francisco area	-0.45739282	-2.974	0.24556845	2.488
FRI	Friday (compared to Tue-Thu)	-0.27329024	-18.77	-0.03721416	-4.587
SAT	Saturday (compared to Tue-Thu)	-2.12756571	-123.64	-0.25454696	-19.359
SUN	Sunday (compared to Tue-Thu)	-2.53217195	-136.304	-0.27487264	-18.451
MON	Monday (compared to Tue-Thu)	-0.20962641	-14.694	-0.02636782	-3.337
JAN	January (compared to October)	0.22656795	8.014	-0.00582198	-0.367
FEB	February (compared to October)	0.2149093	8.647	-0.00411314	-0.291
MAR	March (compared to October)	0.17655151	7.392	0.00971868	0.713
APR	April (compared to October)	0.14565459	5.639	-0.0043946	-0.294
MAY	May (compared to October)	0.02475768	0.967	0.01338584	0.896
JUN	June (compared to October)	0.03477912	1.241	0.03172775	1.916
JUL	July (compared to October)	-0.04791802	-1.83	0.01550385	1.003
AUG	August (compared to October)	0.16721823	6.564	0.03566435	2.482
SEP	September (compared to October)	0.13178481	4.99	0.01323778	0.895

NOV	November (compared to October)	-0.01298481	-0.61	-0.01549753	-1.271
DEC	December (compared to October)	-0.13510301	-5.454	-0.0212291	-1.471
YEAR_16	Data collected in 2016 (compared to 2021)	0.76840508	34.997	0.11995052	8.754
YEAR_17	Data collected in 2017 (compared to 2021)	0.63454918	23.901	0.11913676	7.421
YEAR_18	Data collected in 2018 (compared to 2021)	0.586113	21.563	0.09565085	6.036
YEAR_19	Data collected in 2019 (compared to 2021)	0.23691235	10.171	0.05076571	3.463
YEAR_20	Data collected in 2020 (compared to 2021)	-0.51220774	-13.633	-0.02381115	-1.027
YEAR_22	Data collected in 2022 (compared to 2021)	0.31810788	12.23	0.0117138	0.714
YEAR_23	Data collected in 2023 (compared to 2021)	-0.11989868	-4.646	0.05965312	3.608

Table B-2	Model of Non-Work Trip Auto Vehicle Miles per Household-day	Part 1: Binary logit (Miles > 0)		Part 2: Regression LN(Miles)	
Observations	Number of survey household-days used in estimation	310922		227742	
Mean value	Fraction of non-0 values for Part 1, average of non-0 values for Part 2	0.732		29.22	
Model fit	McFadden rho-squared for Part 1, regression R-squared for Part 2	0.148		0.084	
Variable	**Description**	**Estimate**	**t value**	**Estimate**	**t value**
Constant	Intercept term	-0.7211878	-8.41	1.80930691	37.66
LOG_INCOME	LN(Max(Household income,0) + 1) (0 if data missing)	0.14384194	18.529	0.08015608	18.365
INCBPOV	1 if household income is below Federal poverty level, otherwise 0	-0.05717238	-2.347	0.04227399	2.806
MISSING_INCOME	1 if income data is missing, otherwise 0	1.53680727	17.347	0.88009226	17.557
NO_VEHS	1 if there are no vehicles in household, otherwise 0	-3.28297976	-84.718	0.18069151	4.322
ADULTS_WITHVEH	Lower of # adults and # vehicles in household if # vehicles>0, otherwise 0	0.4434732	34.733	0.31808975	47.414
ADULTS_WOUTVEH	Max(# adults minus # vehicles,0) if # vehicles >0, otherwise 0	0.06281312	4.67	0.08628191	10.989
NO_WORKERS	1 if there are no working adults in household, otherwise 0	0.24106984	12.902	0.05927122	5.815
N_FTWORKAWH	# of full-time workers in household who usually work out of home	-0.11425661	-9.281	-0.11090306	-17.032
N_PTWORKAWH	# of part-time workers in household who usually work out of home	0.18275074	9.583	-0.04497678	-4.723
N_FTWORKATH	# of full-time workers in household who usually work at home	0.02352853	1.355	-0.05396824	-5.786
N_PTWORKATH	# of part-time workers in household who usually work at home	0.19601298	6.879	0.01572669	1.073
N_AGE_04	# of children age 0 to 4 in household	0.11677385	9.157	-0.00281171	-0.438
N_AGE_517	# of children age 5 to 17 in household	0.05153946	6.931	0.09049212	23.125
N_AGE_1834	# of adults age 18-34 in household (relative to age 35-54)	-0.03074979	-3.628	0.01798547	3.799
N_AGE_5564	# of adults age 55-64 in household (relative to age 35-54)	0.03297744	3.535	0.0215783	4.341
N_AGE_65UP	# of adults age 65 and older in household (relative to age 35-54)	0.09731359	9.213	-0.00899588	-1.611
DIARYDATA	Data collected using diary recall (compared to smartphone)	-0.59398421	-44.041	-0.14704085	-18.771
NOBG	1 if block group ID not available in the data, otherwise 0	-0.52819948	-17.08	0.04189683	2.556
NATWALKIND	National walk index for block group (from SLD)	-0.00526493	-2.068	-0.01480729	-10.315
D3B	Street intersection density, (SLD)	-0.00038865	-4.793	-0.00061828	-11.922

D1B	Population density on unprotected land (SLD)	-0.00438482	-11.275	-0.00066919	-3.134
D4C	Aggregate frequency of transit within .25 mile in evening peak (SLD)	-0.00215718	-16.682	0.00011181	1.147
LND5ABOTH	Jobs plus pop. age 18+ within 45 min by car, decay-weighted (SLD)	-0.00420665	-1.347	0.00972508	5.949
LND5BBOTH	Jobs plus pop. age 18+ within 45 min by transit, decay-weighted (SLD)	-0.01521918	-6.821	0.00155392	1.351
TLEVEL_2	Block group has transit stops, but no transit data in SLD	-0.13908771	-5.081	0.01710819	1.189
TLEVEL_3	Block group may have transit, but no transit data in SLD	-0.12015636	-5.599	0.12900284	12.197
TLEVEL_4	Block group likely has no transit and no transit data in SLD	-0.34506199	-15.833	0.27503402	24.629
NYCCBSA	CBSA is in New York City area	-0.15799	-4.397	0.00636907	0.312
CHICBSA	CBSA is in Chicago area	0.17667744	1.771	-0.15162818	-2.861
WASCBSA	CBSA is in Washington DC area	-0.29081109	-3.303	-0.06210832	-1.098
BOSCBSA	CBSA is in Boston area	-0.07472692	-0.594	-0.02117307	-0.294
SFOCBSA	CBSA is in San Francisco area	-0.08231685	-0.588	0.05722854	0.709
FRI	Friday (compared to Tue-Thu)	0.13913738	9.79	0.15557837	20.516
SAT	Saturday (compared to Tue-Thu)	0.21482424	13.626	0.3424681	41.82
SUN	Sunday (compared to Tue-Thu)	0.11732352	7.649	0.20478925	24.929
MON	Monday (compared to Tue-Thu)	-0.08452432	-6.363	-0.02068	-2.73
JAN	January (compared to October)	0.04513899	1.703	-0.09837619	-6.863
FEB	February (compared to October)	0.06351585	2.669	-0.06976878	-5.491
MAR	March (compared to October)	0.06143946	2.698	-0.03216564	-2.625
APR	April (compared to October)	0.13648573	5.591	0.01182514	0.884
MAY	May (compared to October)	0.13996919	5.843	0.04810088	3.605
JUN	June (compared to October)	0.04698441	1.838	0.04606207	3.163
JUL	July (compared to October)	0.11963359	5.049	0.08712099	6.665
AUG	August (compared to October)	0.11788241	5.041	0.03900706	3.063
SEP	September (compared to October)	0.06655632	2.753	0.01869835	1.421
NOV	November (compared to October)	0.07619608	3.717	0.00585535	0.541
DEC	December (compared to October)	0.02683036	1.159	0.00089306	0.071

YEAR_16	Data collected in 2016 (compared to 2021)	0.42727061	21.389	0.16282046	13.95
YEAR_17	Data collected in 2017 (compared to 2021)	0.34255886	13.645	0.17236364	12.189
YEAR_18	Data collected in 2018 (compared to 2021)	0.60585716	21.412	0.2088859	15.034
YEAR_19	Data collected in 2019 (compared to 2021)	0.26095159	11.579	0.1838424	14.572
YEAR_20	Data collected in 2020 (compared to 2021)	0.37616552	9.241	0.19957055	9.533
YEAR_22	Data collected in 2022 (compared to 2021)	0.00187035	0.079	0.00512549	0.374
YEAR_23	Data collected in 2023\ (compared to 2021)	-0.20649213	-8.691	0.07452101	5.309

Table B-3	Model of Long-Distance Trip Auto Vehicle Miles per Household-day	Part 1: Binary logit (Miles > 0)		Part 2: Regression LN(Miles)	
Observations	Number of survey household-days used in estimation	310922		6682	
Mean value	Fraction of non-0 values for Part 1, average of non-0 values for Part 2	0.021		198.74	
Model fit	McFadden rho-squared for Part 1, regression R-squared for Part 2	0.051		0.040	
Variable	**Description**	**Estimate**	**t value**	**Estimate**	**t value**
Constant	Intercept term	-8.7374584	-38.372	5.4131501	50.054
LOG_INCOME	LN(Max(Household income,0) + 1) (0 if data missing)	0.3731218	18.471	0.0034442	0.362
MISSING_INCOME	1 if income data is missing, otherwise 0	3.9682054	16.711	0.0126221	0.112
NO_VEHS	1 if there are no vehicles in household, otherwise 0	-2.0277042	-7.157	0.161848	1.179
ADULTS_WITHVEH	Lower of # adults and # vehicles in household if # vehicles>0, otherwise 0	0.324267	10.217	0.027759	1.809
ADULTS_WOUTVEH	Max(# adults minus # vehicles,0) if # vehicles >0, otherwise 0	-0.1355538	-2.985	0.0215756	0.994
NO_WORKERS	1 if there are no working adults in household, otherwise 0	-0.2119057	-3.975	-0.0032763	-0.128
N_FTWORKAWH	# of full-time workers in household who usually work out of home	-0.1208959	-3.891	-0.0156163	-1.064
N_PTWORKAWH	# of part-time workers in household who usually work out of home	-0.0905082	-2.034	-0.0053695	-0.258
N_FTWORKATH	# of full-time workers in household who usually work at home	-0.0807885	-1.777	-0.0073254	-0.344
N_PTWORKATH	# of part-time workers in household who usually work at home	-0.0641581	-0.927	-0.0670877	-2.071
N_AGE_04	# of children age 0 to 4 in household	-0.2126055	-5.803	-0.0025478	-0.147
N_AGE_517	# of children age 5 to 17 in household	-0.0407196	-2.003	0.0176158	1.824
N_AGE_1834	# of adults age 18-34 in household (relative to age 35-54)	0.0966186	3.977	-0.0015877	-0.14
N_AGE_5564	# of adults age 55-64 in household (relative to age 35-54)	0.061724	2.603	0.0252609	2.289
N_AGE_65UP	# of adults age 65 and older in household (relative to age 35-54)	-0.038864	-1.367	0.0088979	0.689
DIARYDATA	Data collected using diary recall (compared to smartphone)	-0.0908978	-2.033	0.1683211	7.74
NOBG	1 if block group ID not available in the data, otherwise 0	-0.1223467	-1.446	-0.1402611	-3.435
NATWALKIND	National walk index for block group (from SLD)	0.0241988	3.557	-0.0026838	-0.844
D4C	Aggregate frequency of transit within .25 mile in evening peak (SLD)	-0.0023386	-2.882	-0.0009241	-2.677
LND5ABOTH	Jobs plus pop. age 18+ within 45 min by car, decay-weighted (SLD)	-0.029427	-3.486	-0.0097163	-2.401

LND5BBOTH	Jobs plus pop. age 18+ within 45 min by transit, decay-weighted (SLD)	-0.0323734	-6.665	0.0027695	1.192
FRI	Friday (compared to Tue-Thu)	0.632732	17.443	-0.0398132	-2.33
SAT	Saturday (compared to Tue-Thu)	0.4938838	12.014	-0.0635341	-3.287
SUN	Sunday (compared to Tue-Thu)	0.6925928	18.077	-0.0785265	-4.34
MON	Monday (compared to Tue-Thu)	0.1212404	2.861	0.0005452	0.027
JAN	January (compared to October)	-0.2754511	-3.466	0.0001591	0.004
FEB	February (compared to October)	-0.1879946	-2.702	0.0467448	1.4
MAR	March (compared to October)	-0.0877391	-1.327	0.0371148	1.166
APR	April (compared to October)	-0.0584141	-0.814	0.0013545	0.04
MAY	May (compared to October)	-0.1061895	-1.521	0.0482464	1.479
JUN	June (compared to October)	0.0291164	0.401	-0.0348466	-1.031
JUL	July (compared to October)	0.1851842	3.046	0.0056392	0.201
AUG	August (compared to October)	0.1093373	1.795	-0.0225297	-0.795
SEP	September (compared to October)	-0.0413644	-0.652	-0.017468	-0.592
NOV	November (compared to October)	0.0025025	0.045	0.0251267	0.974
DEC	December (compared to October)	-0.2688391	-4.054	-0.0314726	-1.015
YEAR_16	Data collected in 2016 (compared to 2021)	0.6760107	10.025	-0.1242218	-3.952
YEAR_17	Data collected in 2017 (compared to 2021)	0.4969621	6.142	-0.1558662	-4
YEAR_18	Data collected in 2018 (compared to 2021)	0.2276382	2.756	-0.0341901	-0.891
YEAR_19	Data collected in 2019 (compared to 2021)	0.2532086	3.463	-0.0506713	-1.507
YEAR_20	Data collected in 2020 (compared to 2021)	0.3604049	3.079	-0.1313613	-2.386
YEAR_22	Data collected in 2022 (compared to 2021)	0.1044292	1.298	-0.0104977	-0.277
YEAR_23	Data collected in 2023 (compared to 2021)	-0.0544071	-0.645	0.0116817	0.293

Table B-4	Model of Work Trip Public Transit Miles per Household-day	Part 1: Binary logit (Miles > 0)		Part 2: Regression LN(Miles)	
Observations	Number of survey household-days used in estimation	310922		7776	
Mean value	Fraction of non-0 values for Part 1, average of non-0 values for Part 2	0.025		18.07	
Model fit	McFadden rho-squared for Part 1, regression R-squared for Part 2	0.322		0.257	
Variable	**Description**	**Estimate**	**t value**	**Estimate**	**t value**
Constant	Intercept term	-9.1700611	-33.206	1.8617525	7.619
LOG_INCOME	LN(Max(Household income,0) + 1) (0 if data missing)	0.2145503	11.718	0.0736727	4.223
MISSING_INCOME	1 if income data is missing, otherwise 0	2.2983778	10.704	0.8348025	4.041
NO_VEHS	1 if there are no vehicles in household, otherwise 0	0.8254627	16.187	0.0214014	0.466
ADULTS_WITHVEH	Lower of # adults and # vehicles in household if # vehicles>0, otherwise 0	-0.5228644	-13.438	0.18709	5.055
ADULTS_WOUTVEH	Max(# adults minus # vehicles,0) if # vehicles >0, otherwise 0	0.3617311	11.046	0.0065842	0.202
NO_WORKERS	1 if there are no working adults in household, otherwise 0	-2.7560779	-22.942	-0.0730753	-0.574
N_FTWORKAWH	# of full-time workers in household who usually work out of home	0.6404563	19.178	0.0801498	2.443
N_PTWORKAWH	# of part-time workers in household who usually work out of home	0.5829116	12.76	-0.0143858	-0.317
N_FTWORKATH	# of full-time workers in household who usually work at home	0.0388864	0.69	-0.0329956	-0.605
N_PTWORKATH	# of part-time workers in household who usually work at home	0.0399207	0.477	0.135841	1.661
N_AGE_04	# of children age 0 to 4 in household	-0.0697828	-2.096	-0.0327292	-0.991
N_AGE_517	# of children age 5 to 17 in household	-0.0971252	-4.481	0.0573171	2.585
N_AGE_1834	# of adults age 18-34 in household (relative to age 35-54)	0.0035068	0.176	-0.0512521	-2.71
N_AGE_5564	# of adults age 55-64 in household (relative to age 35-54)	-0.0529785	-2.057	0.0176266	0.685
N_AGE_65UP	# of adults age 65 and older in household (relative to age 35-54)	-0.3146155	-8.037	-0.0172308	-0.443
DIARYDATA	Data collected using diary recall (compared to smartphone)	0.2482605	6.813	0.2060468	5.932
NOBG	1 if block group ID not available in the data, otherwise 0	0.8314821	4.151	-0.3543146	-2.015
NATWALKIND	National walk index for block group (from SLD)	0.0250393	3.603	-0.0072791	-1.099
D3B	Street intersection density, (SLD)	0.0010417	6.214	-0.0007118	-4.731
D1B	Population density on unprotected land (SLD)	0.0004357	1.759	-0.0004908	-2.802

LND5ABOTH	Jobs plus pop. age 18+ within 45 min by car, decay-weighted (SLD)	0.0094585	0.553	-0.0066928	-0.441
LND5BBOTH	Jobs plus pop. age 18+ within 45 min by transit, decay-weighted (SLD)	0.0597827	8.925	-0.0505562	-7.726
TLEVEL_2	Block group has transit stops, but no transit data in SLD	0.0350601	0.332	-0.5034763	-4.59
TLEVEL_3	Block group may have transit, but no transit data in SLD	-0.2746094	-3.65	-0.1551796	-2.037
TLEVEL_4	Block group likely has no transit and no transit data in SLD	-0.770398	-7.813	-0.0339696	-0.337
NYCCBSA	CBSA is in New York City area	1.7231925	26.814	0.661078	11.344
CHICBSA	CBSA is in Chicago area	1.4453438	9.445	0.4612768	3.358
WASCBSA	CBSA is in Washington DC area	1.5788252	12.642	0.2053117	1.904
BOSCBSA	CBSA is in Boston area	1.6989867	9.283	0.0744084	0.474
SFOCBSA	CBSA is in San Francisco area	1.6031592	8.025	0.4535771	2.554
FRI	Friday (compared to Tue-Thu)	-0.3282662	-7.843	-0.1623805	-3.824
SAT	Saturday (compared to Tue-Thu)	-2.0383237	-23.303	-0.058597	-0.642
SUN	Sunday (compared to Tue-Thu)	-2.3471676	-23.559	-0.2277279	-2.191
MON	Monday (compared to Tue-Thu)	-0.1312283	-3.69	-0.0092686	-0.267
JAN	January (compared to October)	-0.1868257	-2.114	-0.0937721	-0.987
FEB	February (compared to October)	-0.098319	-1.225	0.0100394	0.114
MAR	March (compared to October)	-0.0508167	-0.659	-0.0095874	-0.113
APR	April (compared to October)	0.5910659	8.062	0.1705555	2.149
MAY	May (compared to October)	0.8381859	12.179	0.2494221	3.331
JUN	June (compared to October)	0.7592028	10.088	0.3315621	4.197
JUL	July (compared to October)	0.2711179	3.091	0.0440594	0.498
AUG	August (compared to October)	0.5265472	7.076	0.0475268	0.615
SEP	September (compared to October)	0.3329814	4.276	0.0231947	0.287
NOV	November (compared to October)	0.0885235	1.21	0.1122327	1.499
DEC	December (compared to October)	0.1820694	2.306	0.0198984	0.246
YEAR_16	Data collected in 2016 (compared to 2021)	1.8596469	19.38	0.4000786	3.79

YEAR_17	Data collected in 2017 (compared to 2021)	2.4867983	25.168	0.4122366	3.927
YEAR_18	Data collected in 2018 (compared to 2021)	2.6877439	26.083	0.0595515	0.507
YEAR_19	Data collected in 2019 (compared to 2021)	2.6470559	27.997	0.2222079	2.12
YEAR_20	Data collected in 2020 (compared to 2021)	1.2953273	6.716	0.7098595	3.484
YEAR_22	Data collected in 2022 (compared to 2021)	1.3890208	11.976	-0.038128	-0.307
YEAR_23	Data collected in 2023 (compared to 2021)	0.2083473	1.92	-0.117843	-1.026

Table B-5	Model of Non-Work Trip Public Transit Miles per Household-day	Part 1: Binary logit (Miles > 0)		Part 2: Regression LN(Miles)	
Observations	Number of survey household-days used in estimation	310922		12277	
Mean value	Fraction of non-0 values for Part 1, average of non-0 values for Part 2	0.039		15.14	
Model fit	McFadden rho-squared for Part 1, regression R-squared for Part 2	0.254		0.145	
Variable	**Description**	**Estimate**	**t value**	**Estimate**	**t value**
Constant	Intercept term	-5.2993076	-26.66	2.7579211	12.191
LOG_INCOME	LN(Max(Household income,0) + 1) (0 if data missing)	-0.0910896	-7.04	0.0273585	1.766
MISSING_INCOME	1 if income data is missing, otherwise 0	-1.1005358	-7.521	0.2584315	1.486
NO_VEHS	1 if there are no vehicles in household, otherwise 0	1.708791	43.847	0.0798089	1.671
ADULTS_WITHVEH	Lower of # adults and # vehicles in household if # vehicles>0, otherwise 0	0.0037048	0.123	0.3005835	8.098
ADULTS_WOUTVEH	Max(# adults minus # vehicles,0) if # vehicles >0, otherwise 0	0.5510013	22.286	0.1289966	4.413
NO_WORKERS	1 if there are no working adults in household, otherwise 0	0.1738024	4.273	0.0473772	0.963
N_FTWORKAWH	# of full-time workers in household who usually work out of home	0.1033163	3.988	-0.0916241	-2.908
N_PTWORKAWH	# of part-time workers in household who usually work out of home	0.214691	5.6	-0.0704603	-1.52
N_FTWORKATH	# of full-time workers in household who usually work at home	-0.0978514	-2.394	-0.072532	-1.443
N_PTWORKATH	# of part-time workers in household who usually work at home	0.2161424	3.391	-0.0724004	-0.904
N_AGE_04	# of children age 0 to 4 in household	-0.1590821	-4.879	0.0191922	0.441
N_AGE_517	# of children age 5 to 17 in household	0.0739409	4.365	0.0539944	2.36
N_AGE_1834	# of adults age 18-34 in household (relative to age 35-54)	-0.0778513	-4.479	-0.0274924	-1.314
N_AGE_5564	# of adults age 55-64 in household (relative to age 35-54)	-0.0777247	-3.45	0.0314376	1.099
N_AGE_65UP	# of adults age 65 and older in household (relative to age 35-54)	-0.3993947	-15.147	-0.0059789	-0.183
DIARYDATA	Data collected using diary recall (compared to smartphone)	0.0181099	0.663	0.1527331	4.649
NOBG	1 if block group ID not available in the data, otherwise 0	2.3772459	15.637	-0.9627461	-6.01
NATWALKIND	National walk index for block group (from SLD)	0.0109895	2.053	0.0013548	0.206
D3B	Street intersection density, (SLD)	0.0007114	5.641	-0.0006903	-4.58
D1B	Population density on unprotected land (SLD)	-0.0004569	-2.69	0.0001389	0.591
LND5ABOTH	Jobs plus pop. age 18+ within 45 min by car, decay-weighted (SLD)	0.1534096	11.637	-0.0394487	-2.787

LND5BBOTH	Jobs plus pop. age 18+ within 45 min by transit, decay-weighted (SLD)	0.074704	13.005	-0.0501946	-7.162
TLEVEL_2	Block group has transit stops, but no transit data in SLD	0.2682239	3.389	-0.6083977	-6.06
TLEVEL_3	Block group may have transit, but no transit data in SLD	-0.0715682	-1.088	-0.4143543	-4.989
TLEVEL_4	Block group likely has no transit and no transit data in SLD	-0.3796723	-4.81	-0.0148153	-0.147
NYCCBSA	CBSA is in New York City area	0.8050882	12.767	0.4710901	6.299
CHICBSA	CBSA is in Chicago area	0.6585343	4.028	0.3025395	1.54
WASCBSA	CBSA is in Washington DC area	0.5367692	3.678	0.0541151	0.314
BOSCBSA	CBSA is in Boston area	0.4857479	2.216	-0.3819034	-1.471
SFOCBSA	CBSA is in San Francisco area	0.55486	2.14	0.664807	2.079
FRI	Friday (compared to Tue-Thu)	-0.0972396	-3.005	-0.0315347	-0.782
SAT	Saturday (compared to Tue-Thu)	-0.5260883	-13.301	0.1573753	3.149
SUN	Sunday (compared to Tue-Thu)	-0.9055259	-20.079	0.0563813	0.971
MON	Monday (compared to Tue-Thu)	-0.1897588	-6.131	-0.036754	-0.964
JAN	January (compared to October)	-0.3978355	-5.503	0.0382922	0.396
FEB	February (compared to October)	-0.2058176	-3.343	0.156043	1.876
MAR	March (compared to October)	-0.1700895	-2.95	0.1975607	2.529
APR	April (compared to October)	0.2227623	3.927	0.0098979	0.13
MAY	May (compared to October)	0.2037846	3.722	0.0882321	1.208
JUN	June (compared to October)	0.0783555	1.332	0.0657243	0.855
JUL	July (compared to October)	-0.1392349	-2.238	0.1974496	2.49
AUG	August (compared to October)	0.05467	0.94	0.1223463	1.626
SEP	September (compared to October)	0.0604066	0.985	-0.045866	-0.579
NOV	November (compared to October)	0.0591897	1.126	0.1123601	1.667
DEC	December (compared to October)	0.152752	2.744	0.0563771	0.796
YEAR_16	Data collected in 2016 (compared to 2021)	-0.2222937	-4.766	0.2099614	3.529
YEAR_17	Data collected in 2017 (compared to 2021)	0.1284303	2.368	0.1317273	1.854
YEAR_18	Data collected in 2018 (compared to 2021)	0.4587772	8.349	0.0520373	0.754

YEAR_19	Data collected in 2019 (compared to 2021)	0.5704668	12.485	0.0927509	1.517
YEAR_20	Data collected in 2020 (compared to 2021)	-0.4096541	-2.867	0.2015722	1.06
YEAR_22	Data collected in 2022 (compared to 2021)	-0.1068814	-1.791	-0.0274072	-0.342
YEAR_23	Data collected in 2023 (compared to 2021)	0.5717528	11.342	0.5502919	8.283

Table B-6	Model of Work Trip Taxi & Ride-hailing Miles per Household-day	Part 1: Binary logit (Miles > 0)		Part 2: Regression LN(Miles)	
Observations	Number of survey household-days used in estimation	310922		1482	
Mean value	Fraction of non-0 values for Part 1, average of non-0 values for Part 2	0.005		9.28	
Model fit	McFadden rho-squared for Part 1, regression R-squared for Part 2	0.173		0.130	
Variable	**Description**	**Estimate**	**t value**	**Estimate**	**t value**
Constant	Intercept term	-11.066706	-23.8	1.1399545	2.268
LOG_INCOME	LN(Max(Household income,0) + 1) (0 if data missing)	0.4403092	11.58	0.0781856	1.873
MISSING_INCOME	1 if income data is missing, otherwise 0	4.5505334	10.16	0.9789911	1.948
NO_VEHS	1 if there are no vehicles in household, otherwise 0	1.5357386	14.5	-0.0076312	-0.067
ADULTS_WITHVEH	Lower of # adults and # vehicles in household if # vehicles>0, otherwise 0	-0.4281404	-5.18	0.3107808	3.279
ADULTS_WOUTVEH	Max(# adults minus # vehicles,0) if # vehicles >0, otherwise 0	0.2309312	3.51	0.1458798	1.875
NO_WORKERS	1 if there are no working adults in household, otherwise 0	-2.8537418	-9.74	0.097825	0.282
N_FTWORKAWH	# of full-time workers in household who usually work out of home	0.425837	6.24	-0.2021611	-2.598
N_PTWORKAWH	# of part-time workers in household who usually work out of home	0.4556898	4.86	-0.261645	-2.406
N_FTWORKATH	# of full-time workers in household who usually work at home	0.2316013	2.36	-0.006587	-0.061
N_PTWORKATH	# of part-time workers in household who usually work at home	0.3004897	1.89	-0.1752749	-0.959
N_AGE_04	# of children age 0 to 4 in household	0.0831873	1.27	0.0070809	0.096
N_AGE_517	# of children age 5 to 17 in household	-0.0672927	-1.5	-0.0452049	-0.896
N_AGE_1834	# of adults age 18-34 in household (relative to age 35-54)	0.0146601	0.36	-0.0155496	-0.34
N_AGE_5564	# of adults age 55-64 in household (relative to age 35-54)	-0.1929111	-3.28	0.0328357	0.483
N_AGE_65UP	# of adults age 65 and older in household (relative to age 35-54)	-0.2361775	-2.85	-0.1019324	-1.067
DIARYDATA	Data collected using diary recall (compared to smartphone)	0.1521518	2.07	-0.0935687	-1.049
NOBG	1 if block group ID not available in the data, otherwise 0	0.6802433	2.6	0.3804856	1.43
NATWALKIND	National walk index for block group (from SLD)	0.0260525	1.79	0.0091108	0.537
D3B	Street intersection density, (SLD)	-0.0002858	-0.78	-0.0005524	-1.219
D1B	Population density on unprotected land (SLD)	8.715E-05	0.18	-0.0031707	-3.014

LND5ABOTH	Jobs plus pop. age 18+ within 45 min by car, decay-weighted (SLD)	0.0471315	2.01	0.0535037	2.232
LND5BBOTH	Jobs plus pop. age 18+ within 45 min by transit, decay-weighted (SLD)	0.0053117	0.41	-0.0265452	-1.713
TLEVEL_2	Block group has transit stops, but no transit data in SLD	-0.1273864	-0.65	-0.5813731	-2.585
TLEVEL_3	Block group may have transit, but no transit data in SLD	-0.2926815	-2.12	-0.1867738	-1.146
TLEVEL_4	Block group likely has no transit and no transit data in SLD	-0.9268675	-4.95	0.0850992	0.39
NYCCBSA	CBSA is in New York City area	0.2651638	1.58	0.2295691	1.14
CHICBSA	CBSA is in Chicago area	1.1468476	3.53	-0.1992539	-0.553
WASCBSA	CBSA is in Washington DC area	1.1525722	4.76	-0.3336344	-1.286
BOSCBSA	CBSA is in Boston area	1.092136	2.86	-0.3840727	-0.926
SFOCBSA	CBSA is in San Francisco area	1.0490314	2.41	-0.1653617	-0.344
FRI	Friday (compared to Tue-Thu)	-0.1294784	-1.5	-0.0067347	-0.068
SAT	Saturday (compared to Tue-Thu)	-0.7793194	-6.49	0.1710472	1.231
SUN	Sunday (compared to Tue-Thu)	-0.8648878	-6.95	-0.0858374	-0.6
MON	Monday (compared to Tue-Thu)	-0.1103971	-1.42	-0.0607594	-0.689
JAN	January (compared to October)	0.1542915	0.91	-0.0083085	-0.04
FEB	February (compared to October)	0.0009438	0	0.0033002	0.017
MAR	March (compared to October)	-0.200321	-1.31	-0.1023772	-0.531
APR	April (compared to October)	0.1777484	1.22	-0.1398666	-0.759
MAY	May (compared to October)	0.2392056	1.73	-0.2233144	-1.288
JUN	June (compared to October)	-0.0825759	-0.51	0.1022131	0.547
JUL	July (compared to October)	-0.1335712	-0.8	-0.001107	-0.006
AUG	August (compared to October)	-0.1515928	-0.96	-0.096648	-0.523
SEP	September (compared to October)	0.1462378	0.95	0.0655071	0.365
NOV	November (compared to October)	0.1820706	1.38	-0.1094163	-0.733
DEC	December (compared to October)	0.2243743	1.57	-0.1883505	-1.171
YEAR_16	Data collected in 2016 (compared to 2021)	-0.0366532	-0.29	-0.1069099	-0.719
YEAR_17	Data collected in 2017 (compared to 2021)	0.1049432	0.76	-0.1730934	-0.982

YEAR_18	Data collected in 2018 (compared to 2021)	0.3554053	2.48	-0.2819121	-1.728
YEAR_19	Data collected in 2019 (compared to 2021)	0.512564	4.31	-0.3566225	-2.441
YEAR_20	Data collected in 2020 (compared to 2021)	-0.3489879	-1.14	-0.0292481	-0.083
YEAR_22	Data collected in 2022 (compared to 2021)	0.0717982	0.46	0.1131555	0.604
YEAR_23	Data collected in 2023\ (compared to 2021)	-0.3600138	-2.54	-0.0602678	-0.346

Table B-7	Model of Non-Work Trip Taxi & Ride-hailing Miles per Household-day	Part 1: Binary logit (Miles > 0)		Part 2: Regression LN(Miles)	
Observations	Number of survey household-days used in estimation	310922		3019	
Mean value	Fraction of non-0 values for Part 1, average of non-0 values for Part 2	0.010		12.52	
Model fit	McFadden rho-squared for Part 1, regression R-squared for Part 2	0.126		0.139	
Variable	**Description**	**Estimate**	**t value**	**Estimate**	**t value**
Constant	Intercept term	-9.0283134	-28.593	0.8859113	2.425
LOG_INCOME	LN(Max(Household income,0) + 1) (0 if data missing)	0.2936536	11.842	0.0776068	2.649
INCBPOV	1 if household income is below Federal poverty level, otherwise 0				
MISSING_INCOME	1 if income data is missing, otherwise 0	3.0452837	10.752	0.98659	2.902
NO_VEHS	1 if there are no vehicles in household, otherwise 0	1.2254104	16.489	-0.140012	-1.524
ADULTS_WITHVEH	Lower of # adults and # vehicles in household if # vehicles>0, otherwise 0	-0.2276483	-4.081	0.0962334	1.303
ADULTS_WOUTVEH	Max(# adults minus # vehicles,0) if # vehicles >0, otherwise 0	0.2657376	5.808	0.0414476	0.68
NO_WORKERS	1 if there are no working adults in household, otherwise 0	0.3068843	3.938	0.0928201	0.931
N_FTWORKAWH	# of full-time workers in household who usually work out of home	-0.0625681	-1.298	-0.0217174	-0.343
N_PTWORKAWH	# of part-time workers in household who usually work out of home	-0.0129904	-0.181	-0.1906894	-1.962
N_FTWORKATH	# of full-time workers in household who usually work at home	0.2053883	3.123	0.0992372	1.159
N_PTWORKATH	# of part-time workers in household who usually work at home	0.3000533	2.716	0.0843031	0.614
N_AGE_04	# of children age 0 to 4 in household	-0.3542214	-5.543	0.190021	2.323
N_AGE_517	# of children age 5 to 17 in household	-0.1217706	-3.406	0.1960386	4.43
N_AGE_1834	# of adults age 18-34 in household (relative to age 35-54)	0.1880127	6.028	0.0322189	0.808
N_AGE_5564	# of adults age 55-64 in household (relative to age 35-54)	-0.2353814	-5.492	0.05806	1.057
N_AGE_65UP	# of adults age 65 and older in household (relative to age 35-54)	-0.7095025	-13.213	0.1367751	2.145
DIARYDATA	Data collected using diary recall (compared to smartphone)	0.1045766	1.895	0.0955522	1.28
NOBG	1 if block group ID not available in the data, otherwise 0	1.353057	7.159	0.6222697	2.998
NATWALKIND	National walk index for block group (from SLD)	0.0354367	3.527	0.0136174	1.076
D3B	Street intersection density, (SLD)	0.0002538	1.037	-0.0012609	-3.827
D1B	Population density on unprotected land (SLD)	-0.0001536	-0.567	-0.000321	-0.71

D4C	Aggregate frequency of transit within .25 mile in evening peak (SLD)				
LND5ABOTH	Jobs plus pop. age 18+ within 45 min by car, decay-weighted (SLD)	0.0585865	3.393	0.0805212	4.179
LND5BBOTH	Jobs plus pop. age 18+ within 45 min by transit, decay-weighted (SLD)	0.0173889	1.823	-0.0488744	-4.107
TLEVEL_2	Block group has transit stops, but no transit data in SLD	-0.0184126	-0.135	-0.5197057	-2.989
TLEVEL_3	Block group may have transit, but no transit data in SLD	-0.2798125	-2.641	-0.160221	-1.172
TLEVEL_4	Block group likely has no transit and no transit data in SLD	-0.5630123	-4.441	0.1427308	0.879
NYCCBSA	CBSA is in New York City area	0.5895286	5.235	-0.0993417	-0.696
CHICBSA	CBSA is in Chicago area	0.7134056	2.61	-0.1090205	-0.32
WASCBSA	CBSA is in Washington DC area	1.1604618	6.171	-0.0843428	-0.373
BOSCBSA	CBSA is in Boston area	0.310875	0.79	-1.0833978	-2.189
SFOCBSA	CBSA is in San Francisco area	1.3445173	4.747	-0.0987226	-0.29
FRI	Friday (compared to Tue-Thu)	0.512211	9.424	-0.1081101	-1.569
SAT	Saturday (compared to Tue-Thu)	0.7025852	12.994	0.0666138	0.968
SUN	Sunday (compared to Tue-Thu)	0.0407581	0.6	0.0023889	0.028
MON	Monday (compared to Tue-Thu)	-0.1520201	-2.379	-0.0130672	-0.161
JAN	January (compared to October)	-0.1197527	-0.99	0.0416223	0.258
FEB	February (compared to October)	-0.1939003	-1.79	-0.1040639	-0.715
MAR	March (compared to October)	-0.1409301	-1.375	0.2109222	1.529
APR	April (compared to October)	0.0345446	0.338	-0.0384231	-0.277
MAY	May (compared to October)	-0.1222475	-1.227	0.0723755	0.547
JUN	June (compared to October)	-0.0158596	-0.151	0.0265647	0.194
JUL	July (compared to October)	-0.0957285	-0.923	0.1729451	1.297
AUG	August (compared to October)	-0.2337823	-2.237	0.1480764	1.098
SEP	September (compared to October)	0.0013718	0.013	0.1076563	0.791
NOV	November (compared to October)	0.031906	0.35	0.2345787	2.02
DEC	December (compared to October)	0.0509436	0.521	0.1695388	1.357
YEAR_16	Data collected in 2016 (compared to 2021)	-0.2346738	-2.801	-0.1213528	-1.047

YEAR_17	Data collected in 2017 (compared to 2021)	-0.0884064	-0.875	-0.1392013	-0.989
YEAR_18	Data collected in 2018 (compared to 2021)	0.0161746	0.15	-0.1750603	-1.269
YEAR_19	Data collected in 2019 (compared to 2021)	0.5541738	6.826	-0.1571299	-1.45
YEAR_20	Data collected in 2020 (compared to 2021)	0.8058814	5.014	-0.1172682	-0.554
YEAR_22	Data collected in 2022 (compared to 2021)	-0.1410053	-1.404	-0.2379226	-1.759
YEAR_23	Data collected in 2023 (compared to 2021)	-0.2764456	-2.852	0.0230557	0.174

Appendix C: Use of Unweighted Survey Data for Estimating Disaggregate Models

The general model estimation and application strategy used for the TIAT transportation cost burden values is essentially the same as that used for advanced regional and statewide travel demand model systems, including activity-based models. These types of models consist of two main steps.

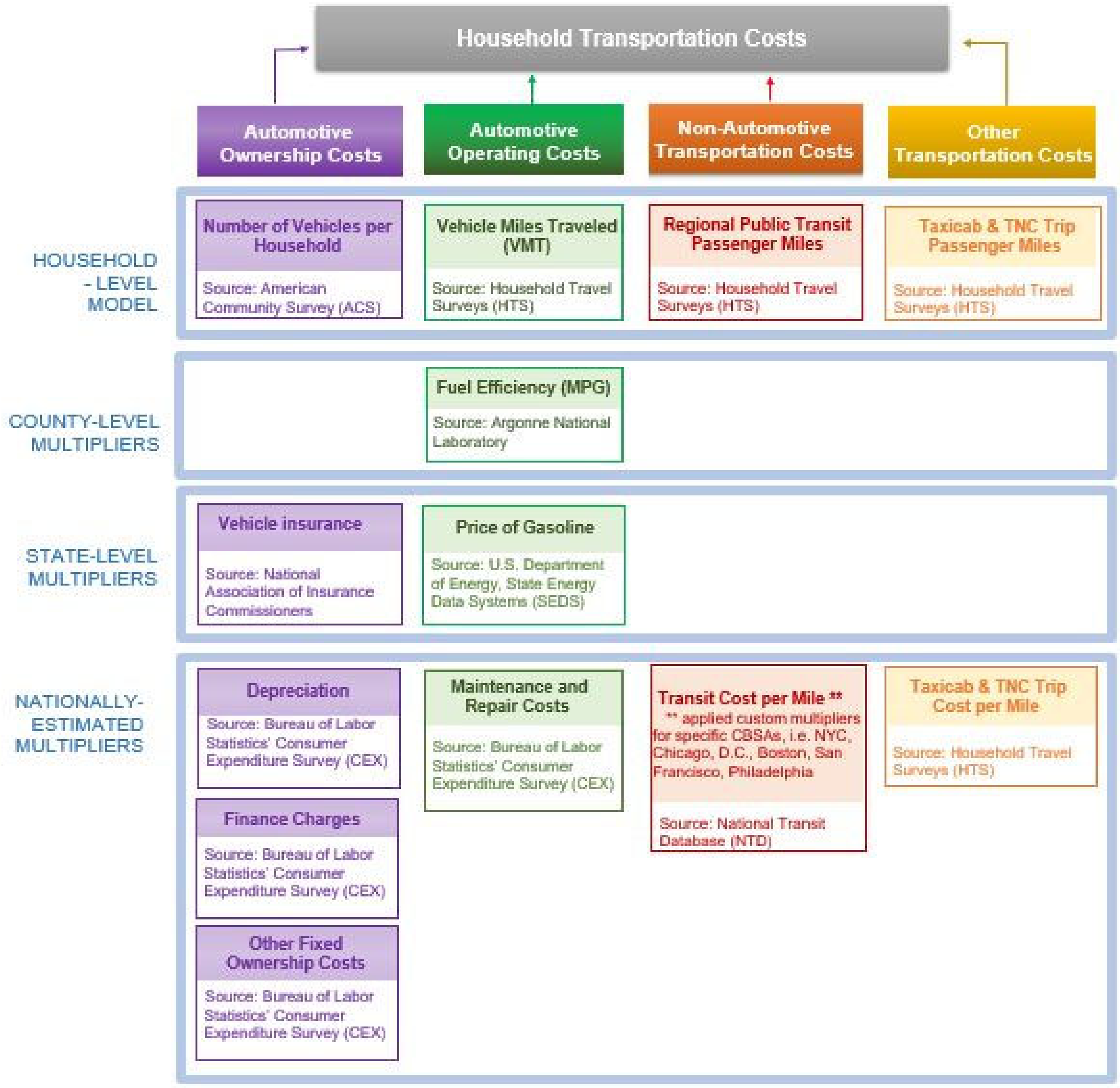
The first of these two main steps is the estimation of disaggregate choice models using household travel survey data, with models estimated using each household-day as an observed choice. For the TIAT, we do not need to estimate models of mode choice, destination choice, time of day choice, etc., that are needed for network-based policy scenario analysis in regional and statewide travel demand models. Rather, to obtain models that can accurately reflect travel mileage and cost for specific types of households in specific Census tracts, it is sufficient to estimate models of the number of miles traveled during each household-day for specific motorized travel modes (auto, transit, and taxi & ride-hailing) for different travel purposes (work and non-work). These models use the same types of explanatory variables that are used in advanced travel demand models, namely household characteristics (including income, vehicle ownership, household size, the age mix and employment status of household members, and race/ethnicity) and neighborhood land use characteristics at the Census block group level (including densities of employment, population, and the street network, various measures of accessibility by transit and auto, and a walkability index).

The second of the two main steps is to apply the disaggregate models to a full synthetic population of all households in the United States—over 120 million separate households. The household population is synthesized using state-of-the-art software to match ACS distributions of the socio-demographic variables used in the models at the Census block group level, sampling households from the ACS PUMS microdata to match those distributions as closely as possible. Group quarters residents are not included. See Appendix E for a discussion of the rationale for not including group quarters residents.

When applying the models, the key "weighting" is done via the synthetic population, which is a representative sample of households in each Census block group and tract. However, since the household travel survey data itself is weighted to adjust for any stratified sampling and non-response bias, there is the question of whether those weights should be used in model estimation. The standard practice in travel demand modeling is to not use survey weights when estimating disaggregate discrete choice models. The main reasons for this include:

- Using weights in estimation reduces the statistical efficiency of the maximum likelihood estimation, giving higher standard errors on the estimates.
- As long as all of the variables that were used to determine the survey weights are also used as explanatory variables in the choice models—which is true for the models estimated in this project—then weighted estimation is not necessary, as it will not provide more accurate estimates. A more detailed statistical explanation can be found in Section 8.4 "Estimating Choice Models under Alternative Sampling Strategies" in the text *Discrete Choice Analysis* (MIT Press, 1985) by Moshe Ben-Akiva and Steven R. Lerman, which remains the most authoritative text on discrete choice modeling of travel behavior.
- Using weights in estimation is required only if the survey data is collected using "choice-based sampling," meaning that respondents are sampled because they have used specific travel modes or visited specific destinations. None of the household travel surveys used for this project used choice-based sampling. Instead, all used address-based sampling (ABS) with the goal of obtaining representative random samples and travel behavior within specific geographic strata.

Appendix D: Transportation Cost Flowchart with Sources



Appendix E: Rationale for Excluding Group Quarters Residents

Group quarters populations are not included in the synthetic population or in any of the models or tools for the project, and it is not recommended to add them in future versions. Key reasons are:

1. People in institutionalized group quarters do not travel, so transportation cost burden is not relevant.
2. That leaves 1.2% of the population in noninstitutionalized group quarters, mostly in on-campus student housing or on-base military quarters. (See the estimates from 2021 ACS below.) Those people have some off-campus or off-base travel, but there are key issues in trying to get an estimate of transportation cost burden for them:
    a. There is little travel survey data, as they are not included in the postal address databases for address-based surveys.
    b. There are some data available from university student surveys, but disposable income is difficult to measure in such surveys. Do they include only their own income (if any) or do they also include money they get from their parents, student loans, etc.?
    c. Universities and the military often provide subsidized or free transit-e.g. some colleges include transit passes as part of the cost of tuition.
    d. College attendance is seasonal. Much of the travel distance for college students is likely during vacations or traveling between home and college.

Below is an estimate of group quarters population by type from 2021 ACS.

Total:	331,893,745	100.0%
Household population	324,132,886	97.7%
Group quarters population:	7,760,859	2.3%
Institutionalized group quarters population:	3,619,502	1.1%
Adult correctional facilities	2,069,113	0.6%
Nursing facilities/skilled nursing facilities	1,332,200	0.4%
Juvenile facilities	141,321	0.0%
Noninstitutionalized group quarters population:	4,141,357	1.2%
College/university student housing	2,707,211	0.8%
Military quarters/military ships	361,692	0.1%

Appendix F: Creating Separate 2019 and 2021 Transportation Cost Burden Estimates

A key reason for creating separate outputs for 2019 and 2021 is that 2021 was still in the midst of the COVID-19 pandemic, as a vaccine was not available until mid-2021 and there was another major wave of the epidemic (the Omicron wave) toward the end of 2021. Compared to 2021, the travel behavior and cost burden during 2019 may be closer to that now encountered in 2024.

The main three reasons that one might expect travel behavior and transportation cost burden to change between 2019 and 2021 are:

1. As mentioned above, the COVID-19 pandemic caused shifts in travel, due to factors such as reduced use of public transit, increased telecommuting, and a possible overall reduction in employment levels.
2. Apart from what can be measured from ACS data on socio-demographics and commute patterns, there may other behavior shifts due to COVID, such as a shift toward hybrid commuting (part-time telecommuting) and a decrease in trip rates for purposes such as shopping, restaurants, and social visits.
3. Underlying demographic trends, such as people moving from areas with declining populations to areas with growing populations, continued between 2019 and 2021, although the COVID pandemic may have altered those trends somewhat.

Adjusting the use of ACS data for population synthesis

Factors 1 and 3 listed above are mainly captured in population synthesis, which uses ACS-based targets for the number of households and persons in each block group, as well as distributions along dimensions such as income, workers per household, auto ownership, age, race/ethnicity, employment status, and commute mode/working from home. Households are sampled from the ACS PUMS microdata using the PopulationSim software to match the various household- and person-level targets as closely as possible for each block group.

ACS data at the Census tract or block group level is only available from 5-year ACS data products. The population synthesis targets used thus far were based on the 2017-2021 5-year ACS tables, so the synthetic population is essentially an “average population” from those 5 years. Table F-1 shows national-level distributions for one key target variable distribution — commute mode share. Compared to 2019, the 2021 public transportation mode share is about 50% lower (1.2% vs. 2.4%), while the share of workers working mainly form home increased by about 200% (8.4% vs 2.7%). The percent of people not employed increased by only 2.7%, but this includes all non-workers including children, retirees and others not in the labor force. If it were calculated only across those in the labor force, the fraction not employed would be 5-6% higher in 2021 than in 2019. As one would expect, the percentages for the 5-year data fall between those for the 2019 and 2021 1-year data.

Table F-1: Commute mode shares – National totals from 5-year and 1-year ACS

Commute Mode	ACS 5 year 2017-2021	ACS 1-year 2019	ACS 1-year 2021	Percent change 2021 vs 2019
Car, truck, or van — drove alone	35.2%	37.1%	32.2%	-13.3%
Car, truck, or van — carpooled	4.1%	4.3%	3.7%	-14.6%
Public transportation	2.0%	2.4%	1.2%	-51.9%
Walked	1.0%	1.1%	0.9%	-19.8%
Taxi, ride-hailing, bicycle, or other	0.9%	0.9%	0.9%	-4.5%
Worked from home	4.6%	2.7%	8.4%	207.3%
Not employed	52.2%	51.4%	52.8%	2.7%

In addition to creating a new 2019 synthetic population, we also want to create a new 2021 synthetic population that more accurately captures these types of differences between the two specific years.

To create more accurate 1-year targets for population synthesis, we use the smallest unit of geography for which 1-year ACS tables are available for the entire U.S. That unit is the PUMA (Public Use Microdata Area). Using the 2012-2021 PUMA system, there are 2,351 PUMA in the U.S, with an average of about 35 tracts or 100 block groups per PUMA. PUMAs are specified to be quite consistent in population size, each with between 100,000 and

200,000 people. That is in contrast to counties, which can have populations ranging from a few thousand to several million people, and 1-year ACS data is not available for small counties with less than 50,000 people. Another advantage for using PUMAs is that it is the smallest unit of geography available in the ACS PUMS data that the synthetic population is drawn from—for each block group, households are sampled only from the PUMA that the block group is located in.

The method for using the 1-year ACS data at the PUMA level for a specific year is as follows:

For each target variable V and PUMA P…

1. Extract/calculate the 1-year ACS value: V(1y,P)
2. For each block group in the PUMA, extract/calculate the 5-year ACS value and sum them across the block groups: V(5y,P) = sum across all BG in P of V(5y,BG)
3. Calculate the adjustment factor F(1y,P) = V(1y,P) / V(5y,P)
4. To get 1-year target values, multiply the 5-year target values for all block groups by the 1-year adjustment factors: V(1y,BG) = V(5y,BG) * F(1y,P)

When the 1-year target values are summed across the block groups within a PUMA, they now match the ACS 1-year data. Thus, this method captures the differences between single years at the PUMA level, while maintaining the relative differences in distributions across block groups <u>within</u> each PUMA. Note that this method has been used by RSG in several other projects.

For a worked-out example, we selected a single PUMA # 5908 in Texas with label "San Antonio City (West)—Between Loop TX-1604 & Loop I-410". This PUMA contains 51 block groups. The target variables in this example are the number of households with 0, 1, 2, 3, and 4+ vehicles, extracted from ACS table B25044.

Table F-2 shows the sum of the target variables across the 51 block groups in the PUMA for the 5-year data, as well as the values for the PUMA in the 1-year ACS for 2019 and 2021. The adjustment factors for each year (in yellow) are calculated by simply dividing the 1-year values by the 5-year summed values. This PUMA happens to be an area of high population growth between 2019 and 2021, as the average factor for 2019 is 0.93 while the average factor for 2021 is 1.11 — particularly high for 0-vehicle HH (1.49) and 4+ vehicle HH (1.39).

Table F-2: Example adjustment factor calculation for PUMA 5098 in Texas

Synthetic population target variable	5-year ACS 2017-2021	1-year ACS 2019	Adjustment factor 2019	1-year ACS 2021	Adjustment factor 2021
0 vehicles available	1,161	1,092	0.94	1,733	1.49
1 vehicle available	12,896	12,167	0.94	13,788	1.07
2 vehicles available	17,585	17,465	0.99	18,522	1.05
3 vehicles available	8,120	6,106	0.75	9,182	1.13
4+ vehicles available	2,674	2,618	0.98	3,710	1.39
TOTAL	**42,436**	**39,448**	0.93	**46,935**	1.11

This adjustment process is repeated for all of the target dimensions and variables used in population synthesis, and for all PUMAs. The two new sets of targets can be used to synthesize separate 2019 and 2021 populations. A small additional advantage of this method is that we do not need to use 2010 Census block group geography to generate the 2019 population, since PUMA geography did not change between 2019 and 2021.

Note that although there is some error in the ACS estimates, we would expect the error to be lower for 1-year data at the PUMA level than for 5-year data at the block group level, given that there are 100 block groups per PUMA, on average.

Applying the household-level models to the 2019 and 2021 synthetic populations

Having separate synthesized populations for 2019 and 2021 as inputs to household-level models will reflect most of the differences in travel behavior discussed above, as variables such as income, auto ownership, household composition, employment status, and working from home are key explanatory variables in the models. Where possible, we can use different block group-level land use variables in the models for 2019 and 2021 (although the SLD does not provide different data for the two years).

A key point is that the household-level models were estimated on a combined dataset with data from 2016 through 2023. In each model, a set of residual year-specific variables was estimated for all years of data relative to 2021. All of the models showed a significant residual positive effect on miles traveled for 2019 (and for 2016, 2017 and 2018) relative to 2021. The effects are much stronger for transit miles traveled than for auto VMT and are also somewhat stronger for taxi & ride-hailing mileage than for auto VMT. For all modes, the effects are somewhat stronger for commute travel than for non-commute travel. These estimates are consistent with what one would expect with increased hybrid telecommuting and avoidance of public modes in 2021 relative to 2019.

Appendix G: Summary Comparison of 2019 and 2021 National Level Model Outputs

While preparing the separate single-year model results for 2019 and 2021, we have carried out summary analyses to confirm that the differences in the model outputs are as expected, at least at the national level. This brief memo contains the results of those analyses.

Differences in model inputs for 2021 and 2019

There are three main differences in the inputs for running the HTS model for the two years:

1- As described in an earlier memo, separate synthetic populations are created for the two years, using ACS 1-year data at the PUMA level as upper-level controls while still using ACS 5-year (2017-21) data at the block group and tract levels as lower-level controls. (ACS data at the block group- and tract levels is not available for 1-year intervals.)

2- Different auto cost ownership and operation costs are used for 2019 and 2021, including fixed ownership costs and fuel prices at the state level. (Fuel efficiency at the county level was not varied across the two years.)

3- When estimating the disaggregate models of household mileage by auto, transit, and taxi & ride-hailing, data was used from several surveys years (2016-2023), and residual trend variables were used to capture differences in mileage for each year relative to the specified base year of 2021, all else equal. The results showed higher mileage for all modes for the pre-COVID years relative to the years 2020 and later, particularly for transit. When running the models for 2019, the 2019-specific trend coefficients were added to the model utilities to capture the difference relative to 2021.

Comparison of national-level averages for 2019 and 2021 model output

Table G-1 shows the national-level averages across all households (profile 1) for all of the key output variables. The next-to-last columns shows the percentage difference between the outputs for two years. All of the 2019 results are in 2019 U.S. dollars while the 2021 results are in 2021 U.S. dollars. When calculated from July 2019 to July 2021, there was 6%

inflation in the U.S. Consumer Price Index (CPI). The last column shows the changes in any monetary outputs adjusted for this 6% inflation.

The items in blue are all from the synthetic population. These were checked against the relevant ACS means and distributions, nationally and for one selected state (South Dakota), and the results all match the ACS means and distributions for two years very closely.

The items in green are the vehicle costs that go into the vehicle cost calculations. Separate values were used for the two years for all variables except for fuel efficiency. In real dollars, adjusting for inflation, gas price increased by over 8% and finance costs by over 3%, while average depreciation and maintenance and repair costs stayed about the same and other fixed ownership costs went down by about 5%.

The items in yellow are the outputs of the disaggregate mileage models based on combined household travel survey (HTS) datasets. As expected, the average mileage per HH is substantially smaller for 2021 than in 2019, especially for transit. For at least the first half of 2021, due to COVID most of the largest U.S. cities were still practicing social distancing, working from home, purchasing goods online, and many on-site businesses were closed, resulting in reduced household travel.

The items in violet are calculated based on other values. These calculations were replicated and compared to the values in the output data and found to be the same, apart from minor rounding error. Compared to 2019, auto operation costs, transit spending, and taxi & ride-hailing spending went down substantially in 2021, while auto owner costs when down only slightly (in real dollars). This reduced the average national transportation cost burden for from 18.7% in 2019 to 17.8% in 2021 — 5.2% lower when calculated as a percentage of a percentage (17.76/18.73 - 1). The transportation cost in nominal dollars is about the same for the two years, but the median income is 5.4% higher, so the ratio is lower. In real dollars, mean transportation cost went down by 5.7% but the median income went down by 0.5%, so the ratio went down by 5.2%. Note that VMT went down by more than 20%, but vehicle ownership stayed about the same and gas price went up by about 15% (8.4% in real dollars), so auto costs didn't go down by as much as VMT did.

Table G-1: Comparison of National-level Averages for 2019 and 2021 Key Output Variables

Variable	2019	2021	Percent Change 2019-2021	Adjusted for 6% Inflation
population	328,229,240	331,923,936	1.1%	n/a
households (HH)	122,707,118	127,641,530	4.0%	n/a
mean income per HH	$86,597	$90,907	5.0%	-1.0%
median income per HH	$66,395	$69,998	5.4%	-0.5%
mean housing cost per HH	$16,012	$16,672	4.1%	-1.8%
mean vehicles per HH	1.84	1.84	0.0%	n/a
finance charges per vehicle	$138	$151	9.4%	3.2%
depreciation per vehicle	$3,623	$3,774	4.2%	-1.7%
fixed ownership costs per vehicle	$1,133	$1,137	0.4%	-5.3%
maintenance repair costs per HH	$727	$776	6.7%	0.7%
gasoline price per gallon	$2.68	$3.08	14.9%	8.4%
fuel efficiency (miles per gallon)	20	20	0.0%	n/a
auto vehicle miles per HH	17,443	13,652	-21.7%	n/a
transit miles per HH	717	299	-58.3%	n/a
taxi & ride-hailing miles per HH	103	90	-12.6%	n/a
fuel spending per HH	$2,316	$2,085	-10.0%	-15.1%
auto operation costs per HH	$3,043	$2,862	-5.9%	-11.3%
auto owner cost per HH	$9,006	$9,313	3.4%	-2.4%
transit spending per HH	$178	$74	-58.4%	-60.8%
taxi & ride-hailing spending per HH	$206	$181	-12.1%	-17.1%
total transportation cost per HH	$12,434	$12,432	0.0%	-5.7%
mean transportation cost burden	18.7%	17.8%	-5.2%	n/a

Table G-2: Comparison of National-level Mean Transportation and Housing Cost Burden Model Output by Household Profile

Household Profile	Mean Transportation Cost Burden			Mean Housing Cost Burden		
	2019	2021	Change	2019	2021	Change
1: Average Household	18.7%	17.8%	-1.0%	14.4%	13.7%	-0.7%
2: 1st income quintile	45.0%	46.6%	1.6%	46.1%	47.4%	1.3%
3: 2nd income quintile	24.9%	25.0%	0.1%	24.6%	24.7%	0.1%
4: 3rd income quintile	18.1%	17.9%	-0.2%	18.0%	17.9%	-0.1%
5: 4th income quintile	13.8%	13.6%	-0.3%	13.7%	13.4%	-0.3%
6: 5th income quintile	9.2%	8.6%	-0.6%	8.5%	8.0%	-0.5%
7: Income less than $24,999	49.0%	50.6%	1.6%	50.4%	52.0%	1.6%
8: Income $25,000 to $49,999	26.4%	26.7%	0.3%	26.3%	26.6%	0.2%
9: Income $50,000 to $99,999	18.1%	17.9%	-0.2%	17.7%	17.5%	-0.1%
10: Income $100,000 to $149,999	13.1%	12.8%	-0.3%	13.0%	12.7%	-0.3%
11: Income $150,000 or more	8.8%	8.3%	-0.5%	8.2%	7.7%	-0.5%
12: Below 100% of the poverty level	75.5%	76.8%	1.3%	65.4%	68.1%	2.7%
13: 100 to less than 150% of poverty level	43.7%	43.8%	0.1%	36.2%	36.7%	0.5%
14: 150 to less than 200% of poverty level	34.5%	34.3%	-0.1%	28.8%	29.3%	0.5%
15: 200% of poverty level or greater	15.2%	14.5%	-0.7%	12.4%	11.9%	-0.5%
16: Zero vehicles in household	5.2%	4.3%	-0.9%	2.7%	2.2%	-0.5%
17: Vehicles fewer than adults in household	17.0%	15.7%	-1.3%	13.2%	12.1%	-1.1%
18: One or more vehicles / adult in hh	19.7%	18.7%	-0.9%	15.4%	14.7%	-0.7%
19: No transit service	22.7%	22.0%	-0.7%	17.9%	17.2%	-0.7%
20: Limited transit service	19.2%	18.3%	-0.9%	15.2%	14.5%	-0.7%
21: Transit service available	17.2%	16.4%	-0.9%	13.1%	12.5%	-0.7%
22: Least walkable	21.7%	20.8%	-1.0%	17.2%	16.5%	-0.7%
23: Below average walkable	18.6%	17.7%	-0.9%	14.5%	13.8%	-0.7%
24: Above average walkable	17.8%	16.9%	-0.9%	13.4%	12.7%	-0.7%
25: Most walkable	15.2%	14.1%	-1.2%	11.1%	10.5%	-0.7%

Appendix H: Estimates of Household Automotive Costs by Income Group

Background

In 2012-13, the U.S. Department of Transportation (U.S. DOT) Office of the Secretary and U.S. Department of Housing and Urban Development (HUD) Office of Sustainable Housing and Communities commissioned the Manhattan Strategy Group (MSG) to conduct an analysis to measure the cost of automobile ownership and automobile usage in the United States. The results of this study were used in the HUD Location Affordability Index (LAI), a model which provides estimates of household housing and transportation costs at the neighborhood level as a share of household income.

Ten years later, to develop the most recent version of the U.S. DOT Transportation Insecurity Analysis Tool (TIAT), U.S. DOT sought to develop its own updated model that estimates the transportation cost burden among local households in the United States. The U.S. DOT team referred to the modeling and analysis performed previously by the HUD LAI, and specifically, the Manhattan Strategy Group study. The MSG study provided estimates of various household automotive costs by household income group, and also examined the pattern of depreciation in automotive values.

The results published by the Manhattan Strategy Group in 2013 have been used in the HUD LAI as well as a similar tool produced by the Center of Neighborhood Technology (CNT) Housing and Transportation (H+T) Affordability Index. As subsequent versions of both tools were released, the results originally published by MSG were inflation-adjusted using the Consumer Price Index, however those original MSG study results were never updated with newer data.

In 2023, the U.S. DOT Transportation Cost Burden model team aimed to replicate as well as update the analysis performed by MSG. With the help of the original authors, the U.S. DOT Transportation Cost Burden team replicated the original analysis, first with 2006-10 data, before then updating the analysis with data between years 2015-19 and 2017-21. The datasets used to develop the original MSG analysis and updated replication study were sourced from the Bureau of Labor Statistics (BLS) Consumer Expenditure Survey (CEX) Public-Use Microdata (PUMD) files.

Findings

In an effort to ensure the updated modeling analysis was consistent, the U.S. DOT Transportation Cost Burden Model team first replicated the 2013 results using data from years 2006-10 before updating the analysis with data between years 2015-19 and 2017-21. In comparing annual expenditures for the more recent years to the 2006-10 period, the results were surprising; while vehicle purchase costs increased, other costs, such as drivability and fuel costs, decreased. One reason for this decrease might be if newer cars are designed to be more fuel efficient. In addition, fixed ownership costs remained relatively stable over the 10-year period. In essence, the change in per vehicle expenditures over time was not uniform across categories.

A review of the top line results for all consumer units showed that in real dollar terms:

- Per vehicle purchase costs **increased 13%**
- Per vehicle fixed ownership costs **decreased 3%**
- Per vehicle drivability costs **decreased 23%**
- Per vehicle fuel costs **decreased 23%**

To help validate the results of the 10-year update above, i.e., the varying change in per vehicle expenditures by category, the U.S. DOT model team compared these results that were derived from the Public-Use Microdata (PUMD) files with the official data tables published from the Bureau of Labor Statistics (BLS) Consumer Expenditure Survey (CEX).

Updates to the Analysis

Between the completion of the replication study and the release of the model and tool in Fall 2024, the model team sought further variation for one of the national multipliers: specifically, fixed ownership costs. Fixed ownership costs are an aggregation of 8 line items, 93% of the total value coming from the following three categories:

- Vehicle insurance (71%)
- Vehicle Registration (7%)
- Amount of personal property tax on vehicle (15%)

- Sum of all other (7%)

Because vehicle insurance premiums and expenditures rates can vary so significantly from state-to-state, the model team developed an approach to adjust the 71% of the national fixed ownership cost multiplier for each state according to average premiums and insurance expenditures by state from the National Association of Insurance Commissioners (NAIC). This adjustment changed the fixed ownership cost multipliers for both the 2019 and 2021 datasets to be state-level by income range, as provided in final tables below.

In addition, some of the other more minor changes since the completion of the MSG replication study were:

- Renaming of "service flow costs" to "depreciation costs"
- Renaming of "drivability costs" to "maintenance and repair costs"
- Fuel cost multiplier replaced with this equation:

Average fuel spending = (VMT/MPG) * gasoline price per gallon

where:

- VMT: derived from household-level data
- MPG: derived from county-level data
- Gasoline price: derived from state-level data
- Drivability ratio: derived from national-level data table (below)

The model team found that the replacement equation allows for more regionally appropriate values, and when the results were compared nationally the figures were very close.

The final cost multipliers used in the U.S. DOT Transportation transportation cost burden estimates in the TIAT are shown below for 2021 and 2019.

Final multipliers, 2021 dataset

Real 2021 Dollars

USDOT estimate *Per-vehicle spending by income group among households with at least 1 vehicle* Per Vehicle Expenditures, 2017-21				
Income group number	Depreciation costs	Finance charges	Fixed ownership costs	Maintenance and repair costs
Overall average	3,766	149	x	418
Less than $24,999	4,082	86	SEE STATE TABLE	374
$25,000 to $49,999	3,810	118	SEE STATE TABLE	392
$50,000 to $99,999	3,695	162	SEE STATE TABLE	411
$100,000 to $149,999	3,652	192	SEE STATE TABLE	445
$150,000 or more	3,696	194	SEE STATE TABLE	496

Real 2021 Dollars

USDOT estimate *Per-vehicle fixed ownership costs by state* Per Vehicle Expenditures, 2017-21					
STATE	Less than $24,999	$25,000 to $49,999	$50,000 to $99,999	$100,000 to $149,999	$150,000 or more
Alabama	940.76	987.65	1,008.74	1,027.61	1,140.23
Alaska	982.19	1,031.15	1,053.16	1,072.86	1,190.45
Arizona	1,036.59	1,088.26	1,111.49	1,132.28	1,256.38
Arkansas	924.06	970.12	990.83	1,009.36	1,119.99
California	1,024.24	1,075.28	1,098.24	1,118.78	1,241.40
Colorado	1,105.59	1,160.69	1,185.47	1,207.64	1,340.00
Connecticut	1,161.21	1,219.08	1,245.11	1,268.40	1,407.42
Delaware	1,187.01	1,246.17	1,272.77	1,296.58	1,438.69
District of Columbia	1,290.10	1,354.40	1,383.32	1,409.19	1,563.64
Florida	1,282.93	1,346.87	1,375.62	1,401.35	1,554.94
Georgia	1,163.26	1,221.24	1,247.31	1,270.64	1,409.91
Hawaii	880.08	923.94	943.66	961.31	1,066.68
Idaho	811.84	852.31	870.50	886.78	983.98
Illinois	948.65	995.93	1,017.19	1,036.22	1,149.79
Indiana	843.20	885.23	904.12	921.04	1,021.99

Iowa	799.35	839.19	857.11	873.14	968.84
Kansas	866.61	909.80	929.22	946.60	1,050.35
Kentucky	952.54	1,000.01	1,021.36	1,040.47	1,154.50
Louisiana	1,359.94	1,427.72	1,458.20	1,485.47	1,648.28
Maine	792.65	832.15	849.92	865.82	960.71
Maryland	1,143.15	1,200.12	1,225.74	1,248.67	1,385.53
Massachusetts	1,119.27	1,175.06	1,200.14	1,222.59	1,356.59
Michigan	1,288.10	1,352.30	1,381.16	1,407.00	1,561.21
Minnesota	919.04	964.85	985.45	1,003.88	1,113.91
Mississippi	986.80	1,035.98	1,058.10	1,077.89	1,196.03
Missouri	944.36	991.43	1,012.59	1,031.53	1,144.59
Montana	884.11	928.18	947.99	965.72	1,071.57
Nebraska	864.28	907.36	926.73	944.06	1,047.53
Nevada	1,172.98	1,231.45	1,257.73	1,281.26	1,421.69
New Hampshire	899.40	944.22	964.38	982.42	1,090.09
New Jersey	1,257.76	1,320.44	1,348.63	1,373.86	1,524.44
New Mexico	938.01	984.76	1,005.78	1,024.60	1,136.90
New York	1,306.56	1,371.68	1,400.96	1,427.17	1,583.59
North Carolina	827.45	868.69	887.23	903.83	1,002.89
North Dakota	788.11	827.39	845.06	860.86	955.22
Ohio	859.37	902.20	921.46	938.70	1,041.58
Oklahoma	938.00	984.75	1,005.77	1,024.59	1,136.88
Oregon	983.23	1,032.24	1,054.27	1,074.00	1,191.71
Pennsylvania	990.69	1,040.07	1,062.27	1,082.14	1,200.75
Rhode Island	1,260.96	1,323.81	1,352.07	1,377.36	1,528.33
South Carolina	1,072.97	1,126.44	1,150.49	1,172.01	1,300.47
South Dakota	820.61	861.51	879.90	896.36	994.60
Tennessee	903.99	949.04	969.30	987.44	1,095.66
Texas	1,088.78	1,143.05	1,167.45	1,189.29	1,319.64
Utah	960.10	1,007.96	1,029.47	1,048.73	1,163.67
Vermont	850.97	893.38	912.45	929.52	1,031.40
Virginia	900.36	945.24	965.42	983.48	1,091.27
Washington	1,027.92	1,079.16	1,102.19	1,122.81	1,245.88
West Virginia	953.18	1,000.69	1,022.05	1,041.17	1,155.29
Wisconsin	833.17	874.70	893.37	910.08	1,009.83
Wyoming	845.12	887.24	906.18	923.13	1,024.31

Final multipliers, 2019 dataset

Real 2019 Dollars

USDOT estimate *Per-vehicle spending by income group among households with at least 1 vehicle* Per Vehicle Expenditures, 2015-19				
Income group number	Depreciation costs	Finance charges	Fixed ownership costs	Maintenance and repair costs
Overall average	3,608	137	x	393
Less than $24,999	3,907	77	SEE STATE TABLE	349
$25,000 to $49,999	3,633	110	SEE STATE TABLE	364
$50,000 to $99,999	3,529	151	SEE STATE TABLE	377
$100,000 to $149,999	3,489	172	SEE STATE TABLE	425
$150,000 or more	3,613	181	SEE STATE TABLE	489

Real 2019 Dollars

USDOT estimate *Per-vehicle fixed ownership costs by state* Per Vehicle Expenditures, 2015-19					
STATE	Less than $24,999	$25,000 to $49,999	$50,000 to $99,999	$100,000 to $149,999	$150,000 or more
Alabama	929.06	975.36	996.18	1,014.82	1,126.04
Alaska	1,000.56	1,050.43	1,072.85	1,092.92	1,212.71
Arizona	1,029.82	1,081.14	1,104.22	1,124.88	1,248.17
Arkansas	920.62	966.50	987.13	1,005.60	1,115.81
California	1,011.62	1,062.04	1,084.71	1,105.00	1,226.11
Colorado	1,068.19	1,121.43	1,145.37	1,166.80	1,294.68
Connecticut	1,160.76	1,218.62	1,244.63	1,267.91	1,406.88
Delaware	1,207.62	1,267.81	1,294.88	1,319.10	1,463.67
District of Columbia	1,287.48	1,351.64	1,380.50	1,406.32	1,560.46
Florida	1,284.01	1,348.00	1,376.78	1,402.53	1,556.25
Georgia	1,122.75	1,178.70	1,203.87	1,226.39	1,360.80
Hawaii	897.84	942.59	962.71	980.72	1,088.21
Idaho	803.59	843.64	861.65	877.77	973.98
Illinois	957.64	1,005.37	1,026.83	1,046.04	1,160.69
Indiana	847.74	889.99	908.99	925.99	1,027.48
Iowa	798.24	838.02	855.91	871.92	967.48
Kansas	869.73	913.08	932.57	950.01	1,054.14

Kentucky	960.28	1,008.14	1,029.66	1,048.92	1,163.88
Louisiana	1,356.22	1,423.81	1,454.21	1,481.41	1,643.77
Maine	795.64	835.29	853.12	869.08	964.34
Maryland	1,145.29	1,202.37	1,228.04	1,251.01	1,388.12
Massachusetts	1,136.46	1,193.10	1,218.57	1,241.36	1,377.42
Michigan	1,321.25	1,387.10	1,416.71	1,443.22	1,601.40
Minnesota	926.12	972.28	993.04	1,011.61	1,122.49
Mississippi	984.01	1,033.05	1,055.10	1,074.84	1,192.65
Missouri	937.46	984.18	1,005.19	1,024.00	1,136.23
Montana	878.28	922.06	941.74	959.36	1,064.51
Nebraska	864.75	907.85	927.23	944.58	1,048.11
Nevada	1,155.81	1,213.41	1,239.32	1,262.50	1,400.87
New Hampshire	912.49	957.97	978.42	996.72	1,105.96
New Jersey	1,293.64	1,358.11	1,387.10	1,413.05	1,567.93
New Mexico	944.64	991.72	1,012.89	1,031.84	1,144.94
New York	1,301.92	1,366.81	1,395.99	1,422.10	1,577.97
North Carolina	817.58	858.33	876.65	893.05	990.93
North Dakota	797.22	836.95	854.82	870.81	966.25
Ohio	873.27	916.80	936.37	953.88	1,058.43
Oklahoma	955.35	1,002.97	1,024.38	1,043.54	1,157.91
Oregon	993.03	1,042.52	1,064.77	1,084.69	1,203.58
Pennsylvania	1,006.04	1,056.18	1,078.73	1,098.91	1,219.35
Rhode Island	1,252.69	1,315.13	1,343.20	1,368.33	1,518.30
South Carolina	1,048.24	1,100.49	1,123.98	1,145.01	1,270.50
South Dakota	813.57	854.12	872.36	888.67	986.07
Tennessee	907.06	952.27	972.59	990.79	1,099.38
Texas	1,091.66	1,146.07	1,170.53	1,192.43	1,323.12
Utah	955.55	1,003.17	1,024.59	1,043.75	1,158.15
Vermont	860.40	903.28	922.56	939.82	1,042.83
Virginia	907.50	952.73	973.07	991.27	1,099.91
Washington	1,031.06	1,082.44	1,105.55	1,126.23	1,249.67
West Virginia	980.81	1,029.69	1,051.67	1,071.35	1,188.77
Wisconsin	841.19	883.11	901.96	918.84	1,019.54
Wyoming	850.65	893.04	912.11	929.17	1,031.01

Appendix I: Use of ACS Housing Cost Data for Developing Housing Cost Estimates

Housing Costs represent the average annual housing cost per household, which in the TIAT is calculated for all households as well as for various subsets of households (referred to in the TIAT as Household Profiles) for each Census tract. These housing costs are calculated across a synthetic population of all households in the U.S. using data from the 2017–2021 American Community Survey (ACS) Public Use Microdata Sample (PUMS) and ACS 5-year tables, processed through the PopulationSim tool using a technique known as population synthesis. This population synthesis technique is described earlier in the "Population Synthesis" section of this TIAT Technical Documentation, with additional details also provided in Appendix A.

The average housing costs presented in the TIAT are not those directly reported in the ACS or PUMS data but instead are derived through the population synthesis process which uses the ACS and PUMS as inputs. This process ensures that the synthesized population matches the real-world distributions of variables such as income, household size, and housing tenure (owner/renter).

The TIAT outputs the mean housing cost (not the median) across all households within each geography and household profile combination. While median housing costs could be calculated in a similar manner, they are not currently included in the tool. This option could be considered for future updates.

To add housing cost data to the TIAT, it was added to the synthetic population that is the basis for applying the household-level models. Since housing cost is correlated with other variables in the synthetic population such as income and household size, it was added as a control variable when drawing the population from the ACS PUMS data using the PopulationSim software. The population synthesis process uses block group level targets where available, as the household-level models use block group-level data. There are only two ACS tables (from the 5-year files) that contain housing cost distribution data at the block group level:

B25087: Selected owner costs per month (including mortgage, property tax, insurance, utilities)

B25063: Gross rent cost per month (including rent and utilities)

The first table only covers owner-occupied households, while the second table only covers renter-occupied households, so together they cover all households (occupied housing units).

For the PopulationSim controls, 8 ranges of housing cost were used for setting targets, 4 for owner costs and 4 for renter costs. Table I-1 below shows the ranges and percent of households in each range based on the 2017-2021 ACS tables. The tables in the Appendix show how the columns in the ACS tables are assigned to those 8 groups, using national totals from the 2022 ACS to divide renter costs and owner costs into 4 categories each of roughly similar size (larger in the middle groups).

Table I-1: Ranges and Percent of Households in Each Range Based on the 2017-2021 ACS

Group	% of 2021 Households
1 - own_0_800$	21.4%
2 - own_800_1500$	18.3%
3 - own_1500_2500$	15.5%
4 - own_over_2500$	9.5%
5 - rent_0_800$	9.6%
6 - rent_800_1250$	11.0%
7 - rent_1250_2000$	10.1%
8 - rent_over_2000$	4.7%

PopulationSim draws households from the ACS PUMS microdata, so the data for each household record must be used to categorize each household into one of the eight categories above. The variables used in the ACS PUMS are:

TEN	Tenure
1	Owned with mortgage or loan (include home equity loans)
2	Owned free and clear
3	Rented
4	Occupied without payment of rent

SMOCP	Selected monthly owner costs (use ADJHSG to adjust SMOCP to constant dollars)
0	None
1-99999	$1 - $99999 (Components are rounded)

GRNTP	Gross rent (monthly amount, use ADJHSG to adjust GRNTP to constant dollars)
1-99999	$1 - $99999 (Components are rounded)

ADJHSG	Adjustment factor for housing dollar amounts (6 implied decimal places)
1105263	2017 ACS data factor to 2021 $ = 1.105263
1078962	2018 ACS data factor to 2021 $ = 1.078962
1059761	2019 ACS data factor to 2021 $ = 1.059761
1046695	2020 ACS data factor to 2021 $ = 1.046695
1000000	2021 ACS data factor to 2021 $ = 1.000000

The following rules are used to set ownership and rental costs:

- If TEN = 1 or 2, OWNCOST = SMOCP * ADJHSG, else...
- If TEN = 3 or 4, OWNCOST= 0
- If TEN = 3, RENTCOST = GRNTP * ADJHSG, else
- If TEN = 1, 2 or 4, RENTCOST = 0

The values of TEN, OWNCOST and RENTCOST are used to assign the household to one of the 8 housing cost categories and are also written to the household record in the synthetic population, so that the average housing cost for each tract / profile combination can be calculated and written to the output record when the household-level models are applied.

Distribution of columns in ACS tables to control groups:

TABLE B25087	Selected owner costs per month (including mortgage, property tax, insurance, utilities)				
B25087_001	Total owner-occupied units	84,649,084	100.0%	Group	2022%
B25087_002	Housing units with a mortgage:	51,394,498	60.7%		
B25087_003	Less than $200	55,371	0.1%	1	31.4%
B25087_004	$200 to $299	75,151	0.1%		
B25087_005	$300 to $399	138,741	0.2%		
B25087_006	$400 to $499	254,741	0.3%		
B25087_007	$500 to $599	500,996	0.6%		
B25087_008	$600 to $699	860,717	1.0%		
B25087_009	$700 to $799	1,280,289	1.5%		
B25087_010	$800 to $899	1,665,195	2.0%	2	28.0%
B25087_011	$900 to $999	2,036,894	2.4%		
B25087_012	$1,000 to $1,249	5,975,343	7.1%		
B25087_013	$1,250 to $1,499	6,335,654	7.5%		
B25087_014	$1,500 to $1,999	11,198,800	13.2%	3	24.9%
B25087_015	$2,000 to $2,499	7,692,780	9.1%		
B25087_016	$2,500 to $2,999	4,859,660	5.7%		
B25087_017	$3,000 to $3,499	3,027,617	3.6%	4	15.7%
B25087_018	$3,500 to $3,999	1,807,083	2.1%		
B25087_019	$4,000 or more	3,629,466	4.3%		
B25087_020	Housing units without a mortgage:	33,254,586	39.3%		
B25087_021	Less than $100	243,649	0.3%	1	
B25087_022	$100 to $149	437,429	0.5%		
B25087_023	$150 to $199	805,035	1.0%		
B25087_024	$200 to $249	1,239,532	1.5%		
B25087_025	$250 to $299	1,658,746	2.0%		
B25087_026	$300 to $349	1,999,359	2.4%		
B25087_027	$350 to $399	2,192,830	2.6%		
B25087_028	$400 to $499	4,546,831	5.4%		
B25087_029	$500 to $599	4,131,164	4.9%		
B25087_030	$600 to $699	3,417,687	4.0%		
B25087_031	$700 to $799	2,719,708	3.2%		
B25087_032	$800 to $899	2,124,427	2.5%	2	
B25087_033	$900 to $999	1,618,063	1.9%		
B25087_034	$1000 to $1,099	1,233,182	1.5%		
B25087_035	$1100 to $1,199	941,351	1.1%		
B25087_036	$1200 to $1,299	719,377	0.8%		
B25087_037	$1300 to $1,399	576,118	0.7%		
B25087_038	$1400 to $1,499	456,115	0.5%		
B25087_039	$1,500 or more	2,193,983	2.6%	3	

TABLE B25063	Gross rent cost per month (including rent and utilities)				
B25063_001	Total renter-occupied units	44,238,593	100.0%	Group	2022%
B25063_002	Total: pay cash rent	42,085,857			
B25063_003	Less than $100	100,928	0.2%		
B25063_004	$100 to $149	75,234	0.2%		
B25063_005	$150 to $199	128,308	0.3%		
B25063_006	$200 to $249	339,824	0.8%		
B25063_007	$250 to $299	563,510	1.3%		
B25063_008	$300 to $349	458,067	1.0%		
B25063_009	$350 to $399	411,283	0.9%		
B25063_010	$400 to $449	409,714	0.9%		
B25063_011	$450 to $499	462,035	1.0%		
B25063_012	$500 to $549	538,212	1.2%	5	18.1%
B25063_013	$550 to $599	645,928	1.5%		
B25063_014	$600 to $649	769,390	1.7%		
B25063_015	$650 to $699	879,826	2.0%		
B25063_016	$700 to $749	1,070,103	2.4%		
B25063_017	$750 to $799	1,159,964	2.6%		
B25063_018	$800 to $899	2,609,401	5.9%		
B25063_019	$900 to $999	2,891,333	6.5%		
B25063_020	$1,000 to $1,249	7,113,587	16.1%	6	28.5%
B25063_021	$1,250 to $1,499	5,737,862	13.0%		
B25063_022	$1,500 to $1,999	8,006,332	18.1%	7	31.1%
B25063_023	$2,000 to $2,499	3,965,502	9.0%		
B25063_024	$2,500 to $2,999	1,704,480	3.9%	8	17.4%
B25063_025	$3,000 to $3,499	887,374	2.0%		
B25063_026	$3,500 or more	1,157,660	2.6%		
B25063_027	No cash rent	2,152,736	4.9%	5	