Title

Assessing the Quantification Methodology for the Affordable Housing and Sustainable Communities Program

Authors
Chatman, Daniel G.
Rodynansky, Seva
Boarnet, Marlon
et al.

Publication Date
2025-08-15

UC Berkeley
Institute of
Transportation Studies

Assessing the Quantification Methodology for the Affordable Housing and Sustainable Communities Program

Daniel G. Chatman, University of California, Berkeley
Seva Rodynansky, Occidental College
Marlon Boarnet, University of Southern California
Andre Comandon, University of Southern California
Breitling Snyder, University of California, Berkeley
Kieran Patel, University of California, Berkeley
Jon Atkins, University of California, Berkeley

August 15, 2025

Assessing the Quantification Methodology for the Affordable Housing and Sustainable Communities Program

Contract Number: 22STC009

Principal Investigators: Daniel G. Chatman, Seva Rodynansky, Marlon Boarnet, Andre Comandon. **Other Authors:** Breitling Snyder, Kieran Patel, Jon Atkins

University of California, Berkeley (with the participation of Occidental College and the University of Southern California)

August 15, 2025

Prepared for the California Air Resources Board and the California Environmental Protection Agency

Acknowledgements:

This Report was submitted in fulfillment of Contract # 22STC009: Assessing the Quantification Methodology for the Affordable Housing and Sustainable Communities Program by the University of California, Berkeley under the sponsorship of the California Air Resources Board.

Authorship was shared as follows: Chatman was lead coordinating author. Comandon and Rodynansky were lead authors for Task 1; Chatman and Snyder for Tasks 2 and 4 (with data and methods contributions from Patel and Atkins); Rodynansky for Task 3; Snyder for Task 5; Snyder, Comandon and Chatman for Task 6. Snyder was lead drafting author for Task A (with Patel) and all other portions of the text not mentioned above. Patel and Snyder carried out initial data assembly, trouble shooting, and quantitative analysis of travel data for Task 2. Atkins carried out additional regression modeling and graphic production for Tasks 2 and 4, as did Snyder. Chatman edited all portions of the report and was lead author (with Snyder) on executive summary, introduction and conclusions. Chatman, Comandon, Rodynansky, and Boarnet collaborated on research design and interpretation of findings for all parts of the analysis.

Table of Contents

Abstract

Do affordable housing projects in high-quality transit-oriented development areas reduce auto use? By how much? Under what conditions? These questions are complex but highly relevant for the state of California. Its Affordable Housing and Sustainable Communities (AHSC) program estimates reductions in vehicle miles traveled (VMT) associated with project applications, and scores applications partly on this basis. Building on a large set of existing empirical literature, we carried out a new analysis of how the built environment affects travel in California. We relied on several data sources including movement data from cell phones purchased from a private firm; travel diary data from the 2017 National Household Travel Survey (the most recent household travel survey for the state); data on housing characteristics and commuting from the 2017 American Housing Survey; and neighborhood, community, and regional built environment and public transportation data from Federal and local sources. Consistent with previous literature, we did not find evidence that parcel-level characteristics influenced auto use, but our study reinforced the evidence in existing empirical literature about the importance of larger-scale built environment factors influencing VMT, from the scale of the Census block group up to the radius of a 45-minute drive from home. We recommend that the AHSC calculator be modified to take these factors into account, in addition to including the availability of off-street parking, when calculating VMT reductions. We also recommend that the calculator use a counterfactual assumption about alternative development locations using our quantitative estimates and based on a more appropriate spatial scale.

Public Outreach Document

Since 2015, the Affordable Housing and Sustainable Communities (AHSC) program has invested more than $4 billion to fund over 22,000 affordable units and low-carbon transportation throughout California. Project proposals compete for funding according to multiple criteria, including greenhouse gas (GHG) reduction points which are allocated by a quantification method used to estimate reductions through avoided passenger vehicle miles traveled (VMT). Avoided VMT is calculated based on percentage reductions from the "Unmitigated VMT" for a project, or how much VMT would have been generated had that project not been funded and built. California Air Resources Board (CARB) sought to better inform project selection by updating the quantification methodology to capture other factors of importance and refine VMT reduction estimates. CARB also wished to test the accuracy of the existing methodology by determining whether stated reductions matched actual reductions for existing AHSC-funded projects. To address these areas of interest, the project team used a variety of statistical methods to analyze data from various sources, including location-based services data from Advan, the 2017 National Household Travel Survey (NHTS) California add-on, 2017 American Housing Survey (AHS), 2017 EPA Smart Location Database (SLD) and others.

The project team investigated four key questions: whether AHSC projects reduced VMT in the local neighborhood after development; the importance of built environment measures at the neighborhood and community level, including proximity to transit, destination accessibility, and other measures; whether and how housing unit size influences auto use; and whether there are differences in how senior housing should be assessed in comparison to the general population.

The team found that AHSC projects in low- or high-VMT neighborhoods result in travel behavior similar to the surrounding neighborhood, indicating that building-level characteristics do not cause VMT to vary significantly. In comparison with block groups in which market rate or non-AHSC affordable housing projects were developed over that same period, AHSC projects had no measurable influence on average activity levels as measured by mobile phone data. In contrast, NHTS data provided abundant evidence about the importance of larger-scale built environment factors influencing VMT, from the scale of the Census block group up to the radius of a 45-minute drive from home. Many factors are correlated with VMT reductions, including rail station access, network load density, regional employment access, and street density. There is no evidence that larger units are associated with lower total household VMT, though the AHS lacked comprehensive travel data, limiting analysis. Built environment factors do appear to affect travel differently for seniors, though small sample sizes in the NHTS limit the ability to confidently make recommendations for senior populations specifically.

Based on these research findings, the team presented a number of recommendations on how to update the quantification methodology. The team recommends adopting a different counterfactual “Unmitigated VMT” that reflects likely alternatives for households and likely alternative growth patterns for the state. The team also recommends measuring impacts with absolute VMT reductions, not percentage reductions, and to base the calculator primarily on neighborhood, community, and regional built environment measures, including walking distance to rail, network load density, employment accessibility, and street density. In doing so, it is recommended to remove most building-level density and land use measures, save off-street parking availability. As part of the technical documentation, the team has also provided resources for CARB to calculate VMT estimates by block group-based on the built environment factors of importance. These resources can be used by CARB to update the quantification methodology per the project team’s recommendations.

Assessing the Quantification Methodology for the Affordable Housing and Sustainable Communities Program was sponsored by California Air Resources Board. The authors are Daniel G. Chatman (University of California, Berkeley), Seva Rodynansky (Occidental College), Marlon Boarnet (University of Southern California), and Andre Comandon (University of Southern California), with assistance from Breitling Snyder, Kieran Patel, and Jon Atkins (all at the University of California, Berkeley). The final report will be posted on the CARB website. For more information, contact Daniel G. Chatman (dgc@berkeley.edu) or Julia Branco (Julia.Branco@arb.ca.gov).

Executive Summary

Background

Since 2015, the Affordable Housing and Sustainable Communities program has invested more than $4 billion to fund over 22,000 affordable units in California, as well as improvements to transit and active transportation infrastructure. Project proposals compete for funding according to multiple criteria, including quantitative policy, narrative policy, and GHG emissions reduction points. The California Air Resources Board (CARB) developed a quantification method to estimate GHG emissions reductions through avoided passenger VMT based on project-specific land use, housing, and transportation characteristics. This method has been used by the AHSC program for project selection. In sponsoring this research project, CARB sought to update the quantification methodology to capture other factors of importance and refine VMT reduction estimates. CARB also wished to test the accuracy of the existing methodology by determining whether stated reductions matched actual reductions for existing AHSC-funded projects.

Objectives & Methods

The objective of this research was to conduct up-to-date analysis of factors thought to influence auto use by residents of housing projects in California and use this information to suggest recommendations to improve CARB's method of estimating GHG emissions reductions in AHSC projects. We investigated four key questions: whether AHSC projects reduced VMT in the local neighborhood after development, distinguishing "low-VMT" and "high-VMT" areas; the importance of built environment measures at the neighborhood and community level, including proximity to transit, destination accessibility, and other measures; whether and how housing unit size influences auto use; and whether there are differences in how senior housing should be assessed in comparison to the general population.

To assess whether existing AHSC projects have resulted in a measurable reduction in travel distance, we constructed a database of AHSC projects and used mobile phone data to determine travel behavior of phone users within the Census-defined block groups in which the AHSC project were constructed, in comparison to a large set of control block groups. To test the importance of larger-scale neighborhood-, community-, and regional-scale factors and their impact on VMT, we compiled many spatial variables and used National Household Travel Survey (NHTS) data to conduct statistical analysis on auto use, while controlling for demographic and other control variables.

We were asked to additionally investigate two related questions. First, to determine whether larger housing units are associated with lower per capita auto use, we carried out a set of similar statistical analyses on American Housing Survey data. Second, to investigate the importance of senior housing, we used a subset of the NHTS data to conduct a similar statistical analysis for survey respondents over the age of 65 with and without an income restriction, comparing the effects of built environment characteristics with the population under the age of 65.

Results

Consistent with previous literature, we did not find evidence that parcel-level characteristics influenced auto use, because we did not find any changes in the average distance traveled (by all modes) in block groups where AHSC projects were developed between 2018 and 2023, as measured by mobile phone location data. This analysis was done in comparison with block groups in which market rate or non-AHSC affordable housing projects were developed over that same period.

In contrast, and consistent with previous literature, we found abundant evidence about the importance of larger-scale built environment factors influencing VMT, from the scale of the Census block group up to the radius of a 45-minute drive from home. We tested over 100 variables when accounting for both types of measure and scale. Based on statistical testing we arrived at a final pared-down model in which less than ten built environment variables were included.

The most important variables influencing household auto use include a measure we refer to as “network load density,” which is calculated by adding the number of workers and residents in an area and dividing by total street length, and is meant to measure the potential for congestion on the road network (Chatman 2008). Network load density measured at the two-mile radius around the home location was the most significant single variable associated with variation in auto mileage. Other important measures in our analysis included street density measured at the two-mile radius around home; bus stop density measured at a one mile radius around home; auto accessibility to jobs within a 45-minute drive; the ratio of transit to auto accessibility within 45 minutes travel time; and whether there was a rail station within a quarter mile of home.

Conclusions

We recommend that the AHSC calculator be modified to take the larger scale built environment factors mentioned above into account, in addition to continuing to include the availability of off-street parking, when calculating VMT reductions. We also recommend that the calculator use a regional (metropolitan), state, or block-group-radius-based counterfactual assumption about alternative development locations that is based on the results of our quantitative model, in lieu of the current assumption based on a county average trip length emerging from a state model.

Body of Report

Introduction

The broad purpose of this research project was to evaluate and suggest updates to the GHG emissions reduction quantification methodology for the Affordable Housing and Sustainable Communities program (AHSC). Our specific research purpose toward that end was to carry out several kinds of quantitative analysis using various data sources to estimate how vehicle miles traveled (VMT) is related to the funding and development of AHSC projects.

The AHSC program was established in 2014 with the primary goal of reducing greenhouse gas (GHG) emissions by financing development projects that include affordable housing, transportation, urban greening, and community programming. Projects compete for funding according to multiple criteria, including quantitative policy, narrative policy, and GHG emissions reduction points. The California Air Resources Board (CARB) is responsible for providing guidance on estimating the GHG emissions reductions and co-benefits from projects receiving monies from the Greenhouse Gas Reduction Fund, including the AHSC program.

The California Air Resources Board (CARB) created a method to calculate GHG emissions reductions through avoided passenger VMT based on project-specific land use, housing, and transportation characteristics. This method is used by the AHSC program to help guide project selection. Such methods are required to follow the California Climate Investments (CCI) program's implementing principles, which "ensure that the methodology applies at the project-level, provides uniform methods to be applied statewide, uses existing and proven tools and methods, uses project-level data, and results in GHG emissions reduction estimates that are conservative and supported by empirical literature." (AHSC Round 9 Guidelines, 2025) The current methodology uses information from the California Statewide Travel Demand Model, metropolitan planning organizations, the Institute of Transportation Engineers Trip Generation Manual and Parking Generation Manual, the California Air Pollution Control Officers Association (CAPCOA) "Quantifying Greenhouse Gas Mitigation Measures" report, and the California Emissions Estimator Model (CalEEMod) 6.

In current AHSC quantification methodology a large majority of factors included are building-level, contrary to empirical research which has focused on larger-scale spatial factors that are most influential. As a result, it was not clear whether the current methodology is accurate. Thus, CARB sought to update the method with input from this independent assessment.

This research was broken down into overlapping but discrete tasks:

1. **Estimate before-and-after VMT impacts of AHSC projects**: We carried out an evaluation of pre-development and post-development activity levels in Census block groups where AHSC projects were developed. We compared changes in those locations to control block groups consisting of locations that experienced both non-AHSC affordable housing projects and market rate housing projects, using regression discontinuity and difference-in-difference methods with mobile-phone-based location data.

2. **Estimate cross-sectional household VMT variation as a function of transit proximity and other spatial characteristics**: We conducted a detailed analysis of how neighborhood- and community-level built environment factors including distance to transit, employment density, and jobs accessibility are correlated with individual-level VMT, conducting regression analysis using data from the Nationwide Household Travel Survey (NHTS) and a number of other merged data sources measuring built environment characteristics.

3. **Investigate VMT as a function of housing unit size**: We estimated the relationship of drive-alone commuting duration as a function of housing unit size. Our analysis is based on the theory that larger housing units in dense, transit-proximate settings could enable increased vehicle sharing among larger households. We conducted this analysis using data from the American Housing Survey.

4. **VMT estimates for senior housing**: We investigated how senior housing compares to conventional housing in VMT generation as a function of the same built environment factors used in task 2, using the same data but focusing on individuals 65 years and older, as well as low-income individuals aged over 65.

In addition to the tasks listed above, we also conducted a “sensitivity analysis” showing how our recommended calculations based on Task 2 play out in terms of locations across the state (Task 5); and we described how the existing CARB calculator for VMT reductions is structured and identified areas for improving its accuracy (Task 6).

AHSC program & Quantitative Methodology overview

To gain a better understanding of the program and identify the current choices and assumptions in the VMT reduction calculator, we reviewed relevant AHSC documentation, including the Round 8 Program Guidelines (2023), AHSC Impact Report (2024), AHSC Quantification Methodology (2023), and the unlocked Benefits Calculator Tool (2023). For more in-depth evaluation and discussion of the existing methodology, see Task 6**.**

AHSC program requirements

Before we provide more detail about our analysis and the academic literature on which our analysis is based, we here describe aspects of the AHSC program to provide context for our analysis, including understanding program requirements in their current form.

To be eligible for AHSC funding, applicants must show that their projects align with project-specific requirements. These requirements can be understood as the aspects of a project which may be directly attributed to AHSC funding and therefore considered an aspect of VMT reduction within the Quantitative Methodology. Additionally, as a California Climate Investments (CCI) funding program, certain statutory investment minimums are also required to be met by the AHSC program itself toward the aim of ensuring that the program benefits disadvantaged and low-income communities.

Requirements for Affordable Housing Developments can vary based on whether the Project Area Type is defined as Transit-Oriented Development (TOD), Integrated Connectivity Project (ICP), or Rural Innovation Project Areas (RIPA). These classifications are based on the level of transit service for the project and the definition of the project area as Rural or non-Rural, as defined by California Health and Safety Code 50199.21. TOD projects are defined as those within 0.5 miles of High Quality Transit, not within a Rural Area. minimum of 35% of AHSC funding must be allocated to these project types. ICP projects are defined as those within 0.5 miles of Qualifying Transit, not within a Rural Area, and a minimum of 35% of AHSC funding must be allocated to this project type. RIPA projects are defined as those within 0.5 miles of planned or existing Qualifying Transit, and can be located within a Rural Area or not. These projects must receive a minimum of 10% of AHSC funding. For definitions of level of transit service, see the section on Transit Proximity.

Unit Affordability

Once the Affordable Housing Development (AHD) is operational, a certain amount of housing units provided must be restricted to households of a certain income bracket based on percent of the area median income (AMI) by county, as determined by the CA Tax Credit Allocation Committee. These units must be made available at an affordable rent, as determined by the CA Health and Safety Code 50053 – typically 30% of the upper end of the income bracket. An AHSC-funded housing project is required to have at least 20% of units as Affordable Units, defined as housing units restricted to a household earning no more than 60% AMI. However, the project must also have an overall average affordability of 50% AMI, meaning that among all income-restricted units the average income restriction bracket must be 50% AMI. In addition, all affordable housing units are subject to 55-year affordability covenants for rentals and 30-year covenants for ownership.

Housing Density

For residential-only projects, TOD, ICP, and RIPA projects must have at least 30, 20, or 15 units per acre, respectively. This density calculation may also be adjusted for unit size by multiplying units larger than one bedroom by a factor greater than one (see below for specific factors), though this is not required. In addition, mixed use TOD, ICP, and RIPA projects can demonstrate consistency with housing density requirements using a floor area ratio of 2.0, 1.5, and 0.75, respectively, to compensate for non-residential development.

Table 0.1: Density factor by number of bedrooms

Number of Bedrooms	Density Factor
0	0.7
1	0.9
2	1.5
3	1.6
4	1.8

Transit Proximity

All AHSC-funded projects require AHDs to be built no farther than 0.5 miles from at least one existing or planned transit station or stop, along a defined Pedestrian Access Route. Transit stations or stops near TOD projects must be served by existing High Quality Transit, defined as Qualifying Transit with high frequencies (seven days a week with headways no more than 15 minutes during peak periods) and permanent infrastructure (either a railway or bus rapid transit line). Stations or stops near ICP projects must be served by existing Qualifying Transit, defined as fixed or flexible transit service with arrivals no less than 2 times per hour during peak hours. Note that at the time of application, ICP projects may not be served by existing High Quality Transit.

RIPA projects must have existing or planned stops or stations, served by either High Quality or Qualifying Transit.

AHSC program scoring criteria

In addition to the baseline requirements for eligibility, AHSC applications are scored competitively based on a number of elements. Allocation of points is broken into three main areas: quantitative policy, narrative policy, and GHG emissions reduction points. The scoring of these factors is important to framing the counterfactual, as applicants are incentivized to include them to increase their likelihood of receiving funding, and may not have otherwise included them in the proposal.

Here, we mainly concern ourselves with the scoring of GHG emissions reduction points. These are allocated based on applicant inputs to the AHSC Benefits Calculator Tool, which uses the AHSC Quantitative Methodology to calculate GHG emissions reductions to projects based on a number of building-level and neighborhood-level characteristics for the project. Applications are then ranked and assigned points based on their relative ability to reduce GHGs.

Task A: Literature review

An ever-growing body of research has investigated how different built environment characteristics influence travel behavior and vehicle miles traveled (VMT). Built environment characteristics that have an impact on travel behavior have often been categorized using the “five D’s”: destination accessibility, distance to transit, design, density, and diversity (Cervero and Kockelman, 1997; Cervero et al, 2009). Destination accessibility quantifies the ease with which travelers can reach a location, given factors like travel time, distance, and/or mode of transportation. It is typically measured with a cumulative opportunity metric, which calculates the sum of destinations reachable within a certain distance or time, by a certain mode, or a gravity-based model, which accounts for the respective attractiveness of each possible destination. Distance to transit quantifies the ease with which people can take advantage of transit services, and typically takes into account the type of transit accessed. Design can refer to several aspects that make up the pedestrian experience, including street and sidewalk configuration as well as urban design aesthetics. This is typically measured empirically with street network density and connectivity but can also be observed through “walkability indices” that account for other design factors that augment the pedestrian experience as well. Diversity quantifies the amount of land use mix that exists within a given area, and density is typically measured by residential, employment, or commercial density. Collectively, this framework provides a commonly used structure for interpreting how various built environment elements are related to travel behavior. These studies use different methods to test variables in these categories and demonstrate that highly accessible, well-connected neighborhoods tend to have the largest negative impact on VMT, though the magnitude of these impacts tends to vary by study.

Among the five D’s, destination accessibility consistently emerges as one of the strongest predictors of travel behavior and VMT reduction. In their meta-analysis, Ewing and Cervero (2010) highlight that throughout the existing literature, accessibility to destinations—such as employment centers or commercial districts—is most strongly related to travel behavior, relative to population or job density. Salon et al. (2012) provide another meta-analysis of the existing literature and show that regional accessibility to jobs is highly influential in VMT reduction, in addition to network connectivity, though the results vary in terms of the magnitude of impact. A later study by Diao (2021) reinforced these conclusions to a certain extent while also highlighting the importance of street network design and proximity to transit, finding that inaccessibility to transit and jobs has a strong positive effect on VMT, while walkability and connectivity show significant negative associations. In a research project by Ding (2017), higher street connectivity was associated with lower VMT for both commute and non-commute trips. A study by Cervero and Murakami (2010) using data from 370 urbanized areas in the U.S. showed that while population density was positively associated with VMT per capita, denser road networks and local-retail accessibility moderated this effect. Though contradictory to other empirical findings, this study also found that employment accessibility has a comparatively small effect, relative to the other identified variables.

The role of density has been more debated. As Salon et al. (2012) highlighted in their meta-analysis, density is often highly correlated with other built environment variables, and the observed effect of density on travel may be primarily due to these other correlated factors. The meta-analysis by Ewing and Cervero (2010) found that population and employment densities were only weakly associated with travel behavior once controlling for other built environment factors. As seen in other research by Cervero and Murakami (2010), while population density can possibly have a positive effect on VMT, other built environment variables moderate this increase. Additional studies conclude otherwise, as in work conducted by Ding (2017) which finds that higher residential and employment density are negatively associated with VMT from commute trips.

Researchers have increasingly turned their attention to the spatial scale at which these variables are measured. Currently, the AHSC Quantification Methodology measures all characteristics at the building-level, with the exception of block group-level employment density. However, almost all the relevant literature measures the built environment at the local or neighborhood level. In addition, Nasri (2012) argued empirically that built environment variables measured at larger metropolitan/ regional levels can be influential. Research conducted by Hong et al. (2013) also concludes that the measured effects of various built environment characteristics vary depending on the geographic resolution at which the measurement is taken. More recent results from a study by Lee and Lee (2020) show evidence that urban spatial structure affects VMT as much as local land use factors.

Given the AHSC program’s emphasis on proximity to transit, the literature on distance to transit is particularly relevant. As noted previously, Diao (2021) found that inaccessibility to transit was one of the most strongly correlated built environment variables with increases in VMT. Chatman (2013) found that auto ownership and commuting were lower in households near rail stations, but that these decreases were not attributable to their proximity to rail. Salon et al. (2012) noted that while some studies find statistically significant elasticities of distance to transit, they are relatively small in comparison to other built environment factors.

Another important area of inquiry in relation to the AHSC program is how affordable housing and income level interact with built environment factors to influence VMT. As discussed further in the following sections, many studies show that among households and individuals, being low income is one of the strongest predictors of having lower VMT. However, the extent to which affordable housing itself is capable of reducing VMT is debated, especially in the context of transit-oriented developments. Using data from the 2010-2012 California Household Travel Survey, Howell et al. (2018) found significant reductions in vehicle trips with lower incomes and with increasing urbanization. A working paper by Newmark and Haas (2015) using the same data found that while income and location efficiency (measured by proxy with median commute distance) were statistically significant indicators of VMT individually, there were no statistically significant interactions between the two variables, suggesting that location efficiency reduces VMT similarly across all income levels. Zuk et al. (2015) surveyed residents of affordable and market-rate transit-oriented developments and found that surveyed residents in affordable developments took fewer private vehicle trips than market-rate residents, but they also found TOD factors to not be significantly associated with VMT (and were limited in their analysis by small sample sizes). Chatman et al. (2019) presented a more complex analysis of the issue. Analyzing 2009 National Household Travel Survey and 2010-2012 California Household Travel Survey data, they found that high-income households reduced VMT slightly more when living near rail stations than low-income households, suggesting that regional VMT is more likely to decrease than increase with a shift of high-income households closer to transit. They also noted that displacement of low-income households without densification near rail stations would be likely to increase VMT. Another study by Boarnet et al (2020) confirms these results, finding that rail TOD residence is associated with larger VMT reductions for high income individuals, especially if said residence has a high density of employment opportunities.

In addition to the basis of knowledge as described above, the remainder of this literature review focuses on more recent studies citing some of the studies above and using state-of-the-art data and methods.

Studies using travel diary data

The relationship between the built environment and travel behavior has historically most often utilized travel surveys or diary data that provides detailed information about the trips that respondents take over a day or a number of days. This is still one of the most common types of data used, though more recent research has utilized different methods of analysis, based on machine learning, to explore this data.

U.S. Studies
Newer research builds on previous studies of the built environment and VMT by accounting for self-selection by differentiating between neighborhood types, using these as options in a predictive model of neighborhood selection, and differentiating resulting effects on VMT by trip purpose and neighborhood type. In one study (Salon 2015), California travel survey data was collected from six different sources (three statewide surveys, three regional surveys from MTC, SCAG, and SANDAG) and results from OLS models showed that built environment factors had different effects on travel behavior, depending on neighborhood type and trip purpose (commute or non-work). Local job access had a strong negative impact on both work and non-work VMT, while regional job access had a positive relationship with both trip types. Home transit access was found to have a negative relationship with commute VMT, but little effect on non-work VMT. Workplace land use characteristics were more likely to impact commute VMT than home-based characteristics. For non-work VMT, road density had one of the largest negative elasticities across all neighborhood types, other than rural neighborhoods (Salon 2015).

Work by Miotti, Needell, and Jain (2023) used 2017 National Household Travel Survey data to analyze the relationship between urban form (destination density, land use entropy, transit availability, road network properties and road network capacity) and travel behavior (number of trips, travel distance, time, and energy use). They found that residents of the densest urban areas used 80% less energy for transportation than residents of rural areas and explored the causal factors influencing this relationship. They found that the primary driver of the density–energy use relationship was the larger number of nearby destinations available to urban dwellers, followed by differences in road network properties and public transit infrastructure. The energy benefits of urban dwelling were lower when there were more destinations available 10–100 km away. \The authors argued that urban form predominantly influences how and where people travel, rather than how often and how long (Miotti, Needell, and Jain 2023).

Work completed in 2024 by Sabouri et al. used structural equation modeling to analyze relationships between accessibility to transit, other built environment factors, socioeconomic characteristics, transit use, and vehicle miles traveled. They used household travel survey data collected from 31 MPOs throughout the U.S., resulting in a sample size of 81,573 households. They found that transit and automobile accessibility to jobs, activity (population and employment) density, land use entropy, and the proportion of four way intersections all had an inverse relationship with VMT. While these effects were small relative to socioeconomic impacts, land use entropy appeared to have the greatest impact (Sabouri, Ewing, and Kalantari 2024).

Within the past few years, the rise of machine learning techniques for data analysis have allowed researchers to investigate travel diary data in new ways, particularly to explore nonlinear effects of built environment characteristics, and interactions between them. One study by Guang Tian et al. used boosted regression trees to study the impacts of land use patterns on VMT and compared the results to a multilevel two-stage regression model. The dataset was comprised of household travel survey data from 36 regional surveys in the US, collected between 2008-2014, including 107,949 households. Results from the traditional statistical model indicate that increases in activity density, land-use entropy, 4-way intersection density, transit stop density, presence of a nearby rail station, and destination accessibility within 30 minutes by transit all have a negative effect on both generating any VMT and the amount of VMT thus generated. Destination accessibility by transit has the biggest impact on generating non-zero VMT, while destination accessibility (by auto and transit) as well as intersection density have the greatest impact on the amount of VMT generated. Results from machine learning approaches were presented in terms of the relative importance, or rank, of all variables considered. In the first stage model (testing for non zero VMT) the cumulative impact of all built environment variables outweighed socioeconomic variables 72% to 26% (2% attributed to miscellaneous factors). Activity density was identified as the most important factor of all considered factors, followed by income, transit stop density, and intersection density. In the second stage model (amount of VMT for non-zero) socioeconomic factors outweighed built environment factors, but at a lower level than in the first stage (54% socioeconomic to 38% built environment, with 8% for other factors). Household size was determined to be the most impactful characteristic, followed by the number of household workers, activity density, and income. Of the built environment factors, transit stop density and destination accessibility by transit were determined to be the next most important. Results from machine learning approaches also presented the nonlinear relationship between variables and amount of VMT generated. (Tian et al. 2024)

Another study used 2018-19 disaggregated travel survey data from the Twin Cities region to investigate nonlinear relationships between built environment characteristics and travel behavior by mode. They utilized gradient boosting decision trees, which accumulates the results of simple decision trees one by one into a stronger model. They found that the model predicting driving distance found regional job accessibility and distance to Minneapolis to be the most important indicators of driving distance. Demographics made up half of the relative importance of all variables for driving distance indicators, and local characteristics were found to be less important. For the models predicting transit and active travel distances, all three regional built environment characteristics (job accessibility, distance to Minneapolis, and distance to St. Paul) were the top three most important factors for travel behavior of both these types. (Tao and Cao 2023)

Follow up research to the previous study explored the interactive effects of built environment attributes and income on travel behavior, specifically looking at transportation carbon emissions (TCE). The interactive effects indicate that the interaction between household income and built environment attributes can lead to additional impacts on TCE. Positive interactive impacts are observed among lower-income residents living in areas with limited job accessibility, greater distances from downtown St. Paul, within downtown Minneapolis, and in locations with abundant transit options. Conversely, negative interactive impacts are noted for higher-income individuals residing in downtown Minneapolis and areas well-served by public transit. (Tao and Zhong 2024)

International studies

A 2024 study investigated travel in Toronto, Canada using the 2016 Transportation Tomorrow survey, including disaggregate survey data on non-work trips in Toronto for one day. The researchers use the R5R package to calculate local accessibility to different types of amenities by walking, biking, and transit, and run beta regressions to test the effects of these variables on proportion of non-work trips taken by car. They find that it is only when neighborhoods become “complete” in terms of sufficient access to 4–5 categories of amenities that we see rates of driving start to decrease. The strongest effects on reducing driving trips are in terms of walking accessibility. (Yu and Higgins 2024)

Another study from Dhaka, Bangladesh took another approach to exploring how accessibility influences travel behavior and related carbon emissions. Here the 5D variables were used in a regression analysis to explain travel-related CO2 emissions. Disaggregate travel survey data was used, sourced from the 2014 Household Interview Survey, and CO2 was calculated as the travel distance times fuel mileage times fuel type factor, divided by trip occupants. It is worth noting that “TOD” was defined by neighborhood, based on neighborhood characteristics, rather than by building. TOD was found to have independent effects on transport-related CO2 emissions, and TODs were likely to reduce CO2 emissions. However, these effects vary by trip purposes. (Ashik, Rahman, and Kamruzzaman 2022)

Research based in Berlin, Germany uses a Causal Forest-based double machine learning approach with travel diary data from over 15K households to better understand the relationship between the built environment and CO2 emissions, while accounting for residential self-selection and non-linear moderating effects. They find that the built environment's influence on travel-related CO2 emissions varies across population groups, impacting some households more than others. Analysis reveals that non-senior, high-income, car-owning, working households are most affected, while retirement-age households are least affected. They find that city-wide destination accessibility is the most important variable in determining travel-related CO2 emissions but note the importance of location specificity and differences. (Nachtigall et al. 2025)

Yet another recent study conducted in Germany aims to better understand how different metrics for public transportation opportunities provided to residents of a certain area affect their mode shift differently. The authors use the German National Household Travel Survey with nationwide timetable data to investigate two indicators of transit service: (1) travel time to the nearest central district and (2) cumulative opportunity accessibility, both calculated as the ratio of PT to car travel. Binary logit models indicate that the travel time ratio does not have a relevant influence on the choice of motorized transport mode, but the accessibility ratio does. (Kühnel et al. 2025)

Across geographies and methodologies, these studies demonstrate the continued utility of travel diary data in examining the built environment–travel behavior nexus. While traditional statistical models remain important, machine learning approaches are increasingly used to capture complex, nonlinear, and interactive effects. From reducing VMT and transportation energy use to lowering CO2 emissions, the evidence consistently points to the value of walkable, transit-accessible, mixed-use neighborhoods in achieving more sustainable travel patterns. However, the relative influence of built environment factors depends on socioeconomic context, trip purpose, and location, indicating that one-size-fits-all solutions may fall short. Future research should continue to explore how these relationships vary across population subgroups and urban typologies to better inform policy and planning.

Studies using location-based services (LBS), simulated data, or other sources

Other more recent studies have taken advantage of technological and commercial advances in big data collection and analytics, as well as the outputs from increasingly sophisticated simulations of travel behavior. One report of this nature from ITS Irvine and Davis is primarily concerned with estimating the impact of non-TOD housing developments on reducing VMT, in comparison to the highly prioritized densification of housing in rail-proximate areas specifically. The study aims to better understand the effects of developments in low VMT areas and near bus corridors. VMT estimates from (1) the California Statewide Travel Demand Model (CSTDM) and (2) Streetlight are used to estimate the impact of the built environment, though the model results from these two sources differ significantly. They find that the Streetlight data do not reflect conventional views that VMT can always be significantly reduced through transit and mixed-use development. This finding may indicate that more attention should be paid to the validation of the VMT estimates as well as the complex and context-dependent relationships between VMT, transit, and the built environment. (Kim et al. 2024)

Studies on parking

Though parking has long been identified as possibly one of the most important indicators among built environment characteristics, comprehensive data on parking is notoriously limited, making it challenging to incorporate into models attempting to explain variations in VMT or other types of travel behavior. The studies described below attempt to incorporate parking into their models in various ways, and though imperfect, typically identify significant relationships between parking supply, price, and travel behavior.

Travel diary studies

The most recent and relevant research in this area examines the ability of parking policy to influence travel behavior, including the decision to work from home, using data from the 2010-12 CHTS. The authors built the work parking variable from CHTS data about the price and availability of parking at various trip destinations. They estimated the home parking variable using AHS data to determine the probability of having home parking. Tobit regressions were used for models where VMT was the dependent variable. They find that most households have “free” parking at home, work, and most trips usually end with a free parking space. Regression models reveal strong associations with people choosing to drive and having free parking at home, and households with bundled parking are more likely to drive, drive more miles than other households, and use transit less. They also find that having free parking at work makes people more likely to commute via personal vehicle, less likely to use transit, and more likely to drive more miles. (Ding and Manville 2025)

Research by Currans et al. offer further evidence using the California add-on of the National Household Travel Survey to estimate residential off-street parking constraints for households in LA County. They take a two-level approach to their analysis, first estimating vehicle ownership and travel costs, and then regressing household-level VMT on predicted vehicle ownership. They then applied these models to develop a suite of scenarios for typical households. Results support the hypothesis that the built environment (including parking) influences VMT through travel costs. Findings also show that constrained on-site residential parking (less than one space per unit) accounts for an approximate 10-23% decrease in VMT within each place type, in the scenario testing step of these analyses. Parking estimates are calculated as follows: For single family households in the survey data, the number of parking spots is determined by the age of the house, relying on typical parking provision rates for the given time period. For multifamily households, off-street parking is estimated using the CoStar commercial real estate database. (Currans et al. 2022)

LBS studies

One study uses StreetLight Insight data on vehicle trip origination at the census tract-level before and after the emergence of COVID-19 (2019 and 2021) to assess the contribution of several built environment measures to VMT and to the share of short trips in the 6-county Southern California region. Despite concerns over COVID-era changes, built environment measures remain solidly associated with lower VMT, even in an array of multivariate models. Design-based factors of walkability and low-speed streets show amongst the strongest and most consistent associations with lower VMT and shorter trips. (Kane and Zheng 2025)

Other relevant studies

Using the public transportation module of the 2013 American Housing Survey, Manville and Pinksi investigate the relationship between bundled residential parking and travel behavior while controlling for vehicle ownership. They construct a bundled parking variable based on whether the survey respondent indicates having a garage or carport, though there are limitations in defining the parking variable in this manner. They find that households with bundled parking use transit less, spend more on gasoline, and—when they do take transit—are more likely to drive from their homes to the transit stop. (Manville and Pinski 2020)

Work by Millard-Ball et al. analyzes the impact of San Francisco's affordable housing lotteries, and directly surveys applicants to understand how random assignments to a new unit affect their travel behavior while controlling for self-selection, due to the random nature of the lottery. They analyzed 779 completed surveys that provided relatively coarse information on residents' travel behaviors, car ownership, employment, and interactions with neighbors. Their analysis considers how four primary measures of accessibility affect household behavior, with auto accessibility being measured as the ratio of parking spots per unit. Transit, walk, and bike accessibility were also measured using indices from AllTransit and WalkScore, respectively. They find that a building's parking ratio not only influences car ownership, vehicle travel, and transit use, but has a stronger effect than transit accessibility. They also find no evidence that having a lower parking ratio impacts employment opportunities for low-income residents. (Millard-Ball et al. 2022)

Studies using data sources of other types

Though not focused on parking, Barajas et al. also directly surveyed residents of affordable and market-rate housing in the Bay Area. Data collection for this research included surveying 613 residents living in one of 62 market-rate or affordable housing developments and conducting focus groups. All developments were within two miles of a BART station, and were classified as TOD (within one mile of BART) or non-TOD (between one and two miles of BART). Results indicated that there were higher shares of transit use and lower shares of driving at TODs in comparison to non-TODs. More TOD residents (29%) reported driving less often after moving compared to non-TOD residents (22%), with similar shares across both affordable and market-rate housing types. However, less affordable housing TOD residents take BART (in particular) than market rate, primarily due to costs. Affordable housing residents were also more likely to move due to housing costs and availability of subsidized housing, relative to accessibility of employment or proximity to transit. (Barajas, Frick, and Cervero 2020)

Another study based in Florida uses 13-year panel data on all counties statewide from 2001 to 2014 to calculate the Gini coefficients for each county, as well as the Gini coefficients for the central city, and present OLS and 2SLS estimates for all counties, as well as those defined as urban and rural, based on these GCs. The results show that within urban counties, densifying multifamily housing and industrial uses reduces VMT, while concentrating single family housing and government owned properties increases VMT. Results also show that increasing the spatial dispersal of parking decreases VMT, which the authors explain as being a proxy for more driver-oriented destinations (both private and public). (Ihlanfeldt 2020)

Task 1: Estimate before-and-after VMT impacts of AHSC projects

In this task we evaluated changes in average (per capita) movement activity before and after AHSC projects were developed. We compared before-and-after activity levels in Census block groups with AHSC projects to control block groups including those with non-AHSC affordable housing developments, market rate housing developments, and without new development. We used an activity-based travel data measure from a commercial provider, Advan Research (formerly Safegraph), based on satellite global positioning system (GPS) readings of cell phone movements. These data are a good proxy for VMT in the sense that they are collected frequently (monthly) and are geographically precise (aggregated to the level of the Census block group). The data include all trip purposes (for example, work and non-work travel) and all modes of travel (including public transportation and non-motorized travel).

We used sophisticated statistical modeling techniques to estimate the causal change in activity levels attributable to the opening of a new AHSC project. We found that there was no statistically significant change in activity associated with AHSC project development when compared to Census block groups within one mile of the projects. This included comparisons with and without completed transportation elements, and comparisons between high versus low VMT neighborhoods. In contrast, we did find that new market rate developments sized similarly to AHSC projects had an average reduction in distance traveled of 0.15 miles per capita per day. We also found that in low-VMT block groups, market rate and non-AHSC affordable projects were associated with a reduction in average distance traveled, while in high-VMT block groups, they were associated with an increase in average distance traveled.

Finally, we did find in cross-sectional analysis that AHSC projects with a completed transportation element had lower activity levels. The analysis is limited by the quality of the data and the relatively low number of AHSC projects available for evaluation.

These results support the argument that aligning the AHSC calculator to place projects in low-VMT neighborhoods would be more likely to yield VMT reductions, and AHSC projects do not appear to be distinguishable from other development occurring in the same block groups. As discussed later in Task 2, there is in fact strong evidence that AHSC projects will reduce VMT to the extent that they are placed in low-VMT neighborhoods. The results of Task 1 suggest that any reductions in VMT are not to do with reductions below the level of the existing neighborhood residents, because this analysis found that overall activity levels did not change on a per capita basis within neighborhoods in which AHSC projects were built. We discuss the caveats to do with this finding below, the most important of which is the fact that the mobile phone data do not allow us to distinguish auto travel from other kinds of travel, and so our analysis is based entirely on total travel by all modes and for all travel purposes.

Below we provide more details of our Task 1 data, analysis and findings.

Key Research Question & Basic Theory

The hypothesized mechanisms for VMT reduction dictate the appropriate analytical approach. In this chapter, the focus is on a mechanism that sees the AHSC project changing the travel behavior of residents within the neighborhood where the project was developed. This can happen through the addition of new residents who have lower or higher amounts of travel than existing residents, or because the AHSC investments in auxiliary transportation resources change the travel patterns of existing residents, or both. The only reliable way to test this hypothesis would be to have data on residents' travel activity before the completion of the project and the travel activity of the same residents after, in addition to data on the residents of the new project.

In the absence of data on existing and new residents, evaluations of the impact of a new project must rely on data that indirectly capture the effect of the change of behavior. The proxy nature of the data means that there is no way to ascribe a change in travel patterns to a specific mechanism. All we can know is the change in travel happening in conjunction with new local investments in residential and transportation resources. As we will discuss in the methods section below, this link is replete with intricacies that make uncovering a causal effect (i.e., the addition of AHSC-funded residential units caused average VMT to go down) challenging.

An alternative mechanism that we introduce here, and that is the focus of the next chapter (Task 2), is the possibility that there is a reduction in VMT that comes from the AHSC project being built in a neighborhood that enables residents to drive less than they would have in an alternative location. For example, a project located in central Oakland will likely generate less travel than if the same project had been sited in Dublin, despite both projects being located in the same county and with access to BART stations. That is because Oakland has a higher density of transit, amenities, and employment than Dublin, so residents in Dublin would likely still drive more than if they had been in Oakland.

The two mechanisms lead to different questions and different analytical methods. The first mechanism is tested by comparing a neighborhood only to itself. It focuses on whether the addition of an AHSC project to a neighborhood directly impacts how much people travel on average. For this question, it is essential to have travel/location before and after the completion of the AHSC project. The type of neighborhood is irrelevant because we are not asking whether the type of neighborhood has an impact, only whether the development itself does.

Neighborhood characteristics are essential for understanding the second mechanism. Under the second hypothesis, we examine whether an AHSC project is associated with larger reductions in VMT if it is in one type of neighborhood as opposed to another. The following section details the different data sources we used to conduct both types of analysis.

Materials & Methodology

Travel data and the methods used to analyze these data have advanced significantly in the 2020s. This section provides an overview of the steps involved in developing an approach to measuring changes in VMT as a result of changes in the built environment, specifically the addition of affordable housing. The Literature Review chapter highlighted cutting-edge research on the relationship between the built environment and travel behavior. Despite a diversity of approaches and focuses, this corpus of research concludes that the built environment is critical to understanding how much people drive. However, this research relies primarily on comparing different types of neighborhoods. It answers the questions: if a household lives in this kind of neighborhood instead of that one, will they drive more or less? There is less research to address the question: if the built environment changes locally, will residents of that neighborhood drive less? An important obstacle to examining the second question is the lack of data available and the complexity of the methods that can capture these kinds of changes.

Detecting how an event–in our case, the completion of an AHSC-funded project (that includes a housing element) – causes changes in VMT from one period to the next requires precise data on when the event happened, measures of VMT in the period preceding and following the event within a time frame that is long enough to pick up the effect but not so long that the timing of the event loses relevance. In the case of a new housing development, which would likely have an effect as soon as new occupants move in, data collected in one-year increments would be useless, because we would be unable to differentiate between a project completed at the beginning of the year and having a full year to have an effect before the next measurement, and a project completed later in the year.

Similarly, if the data were collected at too large a scale, the county for example, any impact the project might have would likely be undetectable due to the size of the population included. At the other extreme, comparing VMT at the level of the parcel would be infeasible in the case of the AHSC program because most parcels were vacant or non-residential before the completion of the projects. We selected data sources that strike a compromise in terms of temporal frequency and geographic scale and achieved state-wide coverage for as many years of the program as possible.

Affordable Housing and Sustainable Communities Program

The database of completed and ongoing AHSC projects includes precise locations based on the street address associated with the project (CARB provided the initial database; project team double-checked addresses). The database included the coordinates based on these addresses so that each project could be joined to the matching Census block group - our unit of choice for measuring changes in travel activity. According to the U.S. Census Bureau, a Census block group is a statistical division that contains on average between 600 and 3,000 people; it is smaller than a census tract and can be thought of as a small neighborhood. AHSC projects are in a unique block group, but three block groups contain two projects. Therefore, the unit of analysis is the project-block group (i.e., some block groups appear multiple times in the data but with different timings for the completion of the projects).

Our proxy measure of VMT, or travel activity, described in the Travel Activity Measure section below, is only available from January 2018 to December 2023. Therefore, we only included AHSC projects completed within that time frame. The project completion date was recorded in the database for some projects, but not in all cases. We reviewed every project thoroughly to ensure the date was as accurate as possible down to the month of completion. We assumed, because most projects have a waitlist of applicants, that there was little delay between completion and occupancy. The review process was completed in three steps.

First, we searched for project information on the internet. Some developers have dedicated pages for every project that include the date of completion. In some cases, local news outlets reported the grand opening date of the project. Where no information was available online, we contacted the developer or management company by email and phone to confirm directly the date of completion. We also matched the development with data from CoStar, a commercial real estate firm that collects data on real estate developments, including most large-scale (i.e., over 25 units) residential developments. The CoStar data include completion dates with varying degree of precision. Lastly, we sought to confirm the date by checking that it matched available photographic evidence on Google StreetView. For many streets, Google StreetView has multiple snapshots available for different months and years. For major streets, there are often multiple months available within the same year. In most cases, visual inspection allowed us to confirm that completion happened within a date range that matched what we recorded. Where we found discrepancies, we returned to the previous steps to reconcile the timeline.

AHSC includes transportation elements, or “Sustainable Transportation Improvements,” in addition to residential developments. Transportation element data and completion dates were provided by CARB for our analysis. Elements identified in the database include a wide variety of capital projects, such as installation of new pedestrian walkways, street crossing enhancements, bike sharing and infrastructure, improved transit service, and transit electrification. Though these transportation elements are diverse and may have varying impacts, we are unable to distinguish between types of elements given small sample size. Furthermore, the transportation elements may or may not be directly tied to the development and were not necessarily completed at the same time as the housing development.

We identified 85 AHSC projects with full information for use in the analysis, completed between January 2018 and December 2023 (Table 1.1 and Figure 1.1). 32 of these had completed transportation improvements. On average, AHSC projects had 89 units, with the number of units ranging from 23 to 326.

Table 1.1: AHSC, Market Rate, and non-AHSC Affordable Project Database Overview

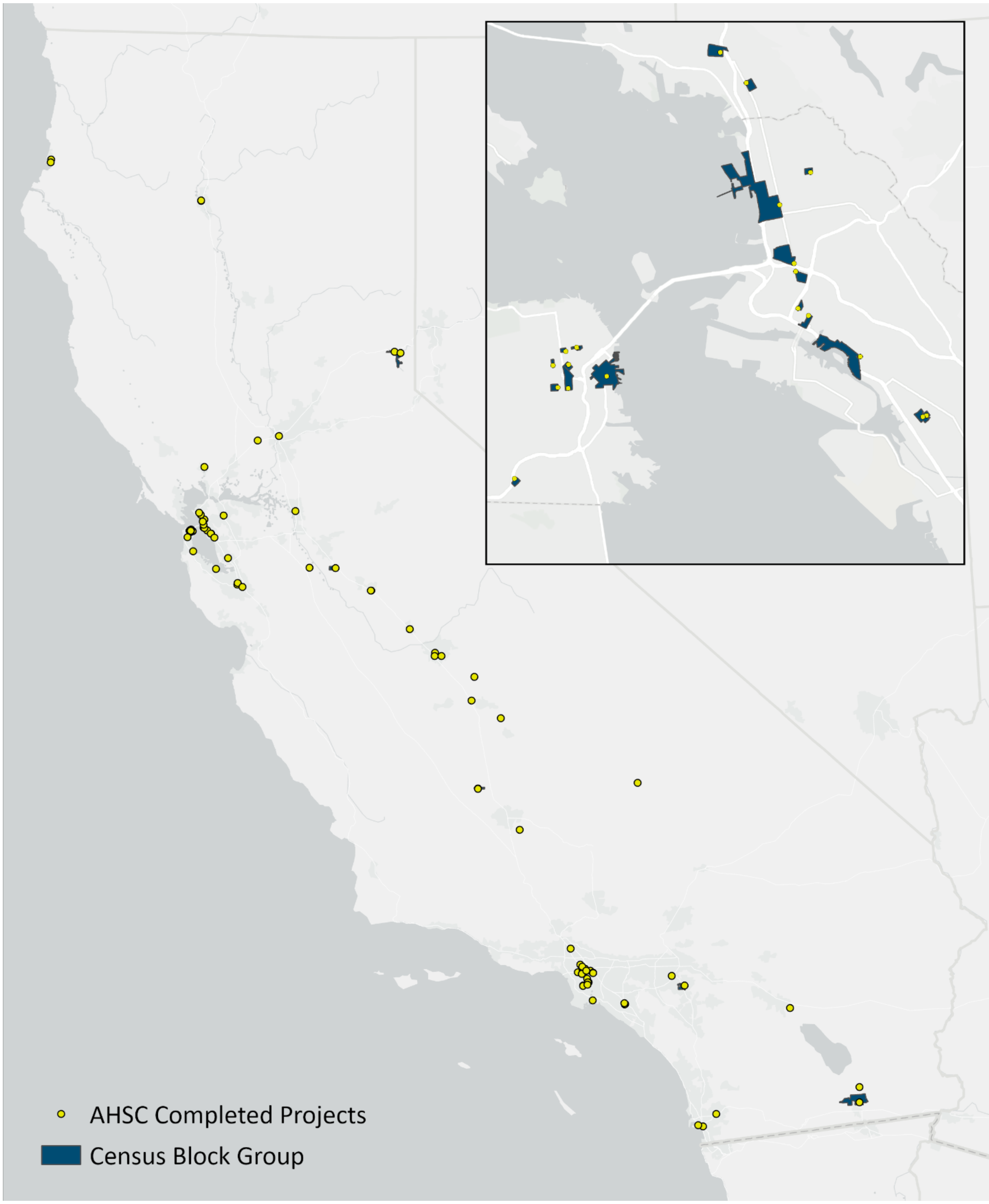
Project type	Number of completed projects	Smallest project (# of units)	Largest project (# of units)	Average number of units
AHSC residential	85	23	326	89
AHSC transportation	32	NA	NA	NA
Market rate*	1068	25	1840	168
Affordable	353	25	552	87

Notes: The category for market rate includes mixed buildings that include affordable units.

Non-AHSC Residential Developments

We include non-AHSC-funded residential developments in the analysis to test whether AHSC projects have a different effect than similar non-AHSC residential developments. The smallest AHSC-funded project in our database included 23 units, a size that matches the data CoStar collects on residential developments. We pulled data from the CoStar database for all residential developments of 25 units or more completed in California between January 2018 and December 2023.

Figure 1.1: Map of AHSC project locations



As with the AHSC projects, CoStar data include precise locations for all projects based on the address. Also, like the AHSC database, not all CoStar developments include a completion date, so some observations had to be dropped. CoStar codes developments as either rental or ownership properties and as market, affordable, or mixed market/affordable. We excluded ownership properties as not directly comparable with AHSC projects and retained all types of rental properties.

This approach yielded 1,421 non-AHSC projects as a basis for comparison (Table 1.1). Of these, 1,068 were market rate and 353 were non-AHSC affordable. The average project unit count was very similar for non-AHSC affordable (87 vs. 89 for AHSC) but was larger for market-rate developments (168).

Travel Activity Measure

This research relies on estimating travel activity at the Census block group level on a monthly basis. Modern methods of estimating travel activity generally, or miles driven specifically, include large surveys (e.g., the National Household Travel Survey), odometer readings, and mobile phone-derived data. Of these methods, only the latter can provide the geographic and temporal granularity required for this analysis.

All providers of mobile phone-derived data rely on data from location-based services (LBS). When users opt into an LBS, they agree to share their real-time location to access tailored results such as directions from a mapping service. The data providers buy these anonymized data and aggregate them in accordance with privacy guidelines before making them available to researchers.

Several data providers make LBS-derived data available to researchers. We evaluated providers based on cost, temporal and geographic coverage, and VMT estimation method. We selected Advan because it was the only provider to enable monthly statewide estimates at a reasonable cost. The main drawback of Advan travel data is that it is not designed as a measure of VMT, but other mobile phone data providers who provide VMT do so as part of a modeling process that introduces potentially large error and renders it potentially unusable for causal analysis of the kind required here.

Advan is an activity-based data aggregator. It samples devices stopping at points of interest (POI) such as coffee shops, retailers, parks, etc., to estimate how many people visit them, for how long, at what time of day, and from where. We use the information on where devices are visiting from to derive a measure of VMT. Advan samples every POI to average a rate between 2% and 10% of all devices. The Advan data are available at the POI level and aggregated to the block group level. We obtained Advan data from Dewey, a provider of datasets for academic research.

We used the data aggregated to the block group level (Advan neighborhood patterns) because it reduces the effect of Advan's privacy rules. Advan suppresses the block group of origin of any POI that received fewer than two visits from that block group. For a single POI, this means that at least two devices originating in the same block group have to visit for it to show up in the data. Considering the sampling rate, this implies that at least 20 devices from the same block group would have to visit the same POI within a month (10% sample with a minimum of 2 visits). This is a high bar for most locations that draw their visitors from a wide array of origins. The same rule applies to the aggregated data. So, all origin block groups where at least 20 devices visited any POI in the destination block group are included. The aggregation of all POI greatly increases the chances that an origin block group will be included.

We use the number of devices traveling between each pairing of origin and destination block groups and the straight-line distance between block group centroids to estimate the total distance traveled by devices. There is no way to recover the sampling rate for the origin block groups because the data are sampled at the destination, but we use the population of the origin block group from the 2015-2019 5-year American Community Survey to weight the distance traveled. We use the 2015-2019 ACS since they pre-date 2020 changes to census block group boundaries. See Census Data section below. The final measure is the device-weighted average distance between the home block group centroid and the destination block group centroid that people from that home block group have visited in a month.

There are few ways to validate the estimates because data that match the block group scale either rely on work commute distance (derived from LODES), some undisclosed source for workers (EPA), or are proprietary. The terms of our contract with Dewey Data, the provider of Advan data, forbids us from directly comparing the data with alternative sources.

Census Data

The Advan-derived measure of average travel distance is available at the block group level based on the 2016 vintage of Census boundaries. The 2016 boundaries are a refinement (correcting minor issues) of the boundaries drawn based on the 2010 decennial census. This limits us to using Census data pre-dating the 2020 boundary updates that were used for all data after 2020, including American Community Survey 5-year estimates that overlap the 2010s and 2020s (Census data can be standardized to have the same set of boundaries, but, especially for the ACS at small scales, where margins of error are often large, this introduces more potential distortions). We use the 2016-2021 American Community Survey (ACS) 5-year estimates to provide information on the block groups, with the assumption that neighborhoods are unlikely to have changed drastically in the following years. Exceptions are areas of rapid development like San Francisco's Mission Bay district, where residential development increased rapidly.

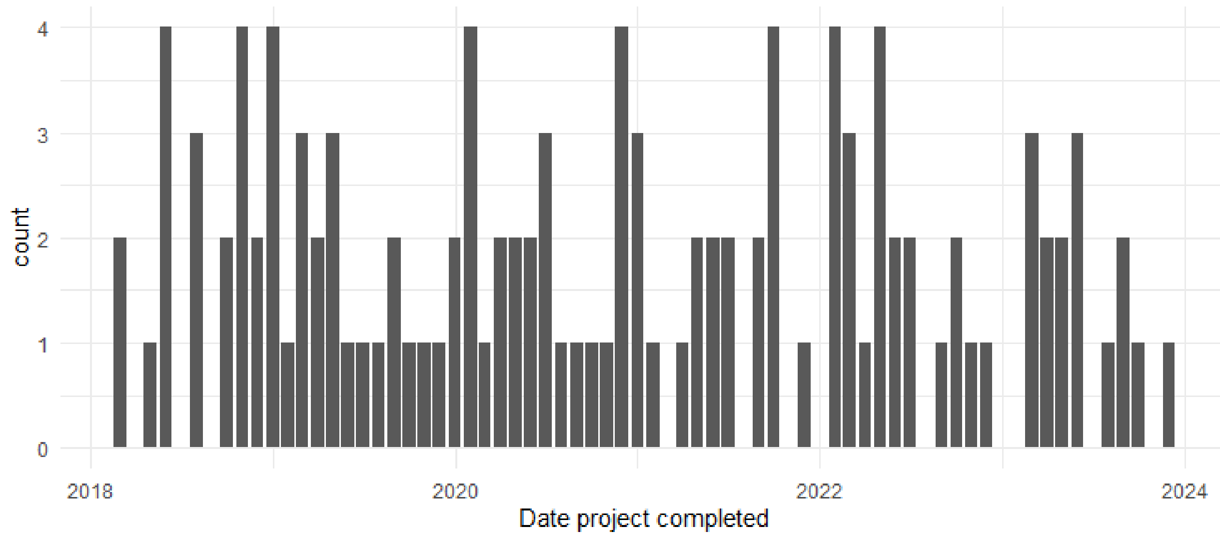
Statistical Models

We cannot answer the question of whether AHSC projects result in reduced travel simply by looking at travel distance before the project was completed and comparing the levels to travel distance after. The simple comparison is unreliable because it ignores the broader changes that may have caused the change and fails to account for neighborhood-specific characteristics. For example, if a project was completed in early March 2020, we would observe a dramatic decrease in travel distance in the following month. However, the decrease would be due primarily to the physical distancing mandates put in place to slow the spread of COVID-19.

The main model used in this analysis is designed to estimate the impact of an event on an outcome measured over time while removing the effect of broader changes like the COVID-19 pandemic and keeping neighborhood characteristics fixed. This type of model is called a difference-in-difference model because it estimates the impact of an event (the treatment) based on the difference with what would have happened in the absence of the development. The model draws from control cases to establish the baseline for what would have happened in the absence of the project.

The principle behind the difference-in-difference method is straightforward, but deviations from the ideal case quickly add complexity. This type of model was developed to evaluate the effect of a treatment that happened at the same time in a set of locations and not at all in the control locations. This would be the case if all AHSC projects had been completed at the same time. However, this is obviously not the case. Figure 1.2 shows project completion counts by month: there were projects completed in almost every month of the study period. These completions include the residential element and transportation element separately, which creates another complication.

Figure 1.2: Count of the number of projects completed in each month between January 2018 and December 2023



Notes: The data include residential and transportation elements separately.

The modeling approach works best when each location is treated once. This is especially relevant in cases like the AHSC program because the projects' effect is hypothesized to be permanent. Once a project is completed, the reduction in travel activity should not revert to the levels before the project existed over time. This means that in neighborhoods where a project has already been completed, the baseline for the next project is not comparable to other neighborhoods where there have been no projects. This is made worse by the addition of the transportation elements and the variation in project size. Projects range in size from 23 units to 326; some of these projects had transportation elements attached to them and others did not. These variations in project size and elements make the assumption that the treatment is the same in all cases implausible. It also makes the choice of control cases difficult. For a non-negligible number of projects, the control is not a neighborhood where there have been no projects. It should be a neighborhood where there has been a project of similar size, but no transportation element yet or no second project yet. Given the small number of projects, such comparisons are not possible, and we are forced to use the same control based on no prior development for all cases.

Methodological advances address some of these challenges, but not all of them. For example, there are models designed to address the staggered nature of the treatment. No method, however, can address all the issues at once. To summarize, the issues are:

1. Project completions are staggered – not all projects are completed at the same time.

2. Treatment intensity varies – the size of the development, nature of the transportation and other investments varies by project.
3. Possible variation in the magnitude of the effect depending on the location of the project – a project may have a different impact if it is in a low or high travel area.
4. Presence of related but different treatments - non-AHSC funded affordable housing, market rate housing, other transportation investments.

Our approach is to begin with a standard difference-in-difference methodology that affords greater flexibility to deal with the multiple treatment types (issues 2 through 4) and then apply some of the methodological innovations for a subset of the data to deal with issue 1.

In addition to issues associated with the nature of the treatment, the choice of control cases poses a different kind of challenge. The method for evaluating the impact of new developments assumes that we can make a plausible assumption about what would have happened in the absence of the project. Usually, this means selecting a set of control cases that are comparable in the period before the project is completed so that the trend in average travel in those neighborhoods serve as a baseline for what would have happened in the absence of a project. As mentioned above, there are issues with that assumption when projects are not completed at the same time and when we have differences in the types of treatment.

We use two sets of control cases. In the first approach, we focus on AHSC projects only. By restricting the data to only one kind of treatment, we establish a baseline for the impact of projects based on a narrower approach to control cases because AHSC projects are more specific in the kind of location where they are likely to be developed. Therefore, we use a set of control block groups where there were no recent residential developments of any kind and that were within one mile of the location of an AHSC. The proximity to an AHSC project ensures that the control block groups are in similar locations and share socioeconomic characteristics (see Table 1.2 for a sample of variables). For this data set, we have 1,152 control cases. The focus on a single type of treatment also allows for a more sophisticated approach to estimation that accounts for the staggered nature of the data. This approach cannot be applied widely to all relevant models and serves as an assessment of possible bias in the results.

Table 1.2: Comparison of treatment and control block groups

Block group demographics	Treatment – AHSC	Control (within 1 mile of AHSC, no projects)
Median income	$54,800	$65,776
% below poverty line	26	22

Block group demographics	Treatment – AHSC	Control (within 1 mile of AHSC, no projects)
% renters	73	70
% drive alone	57	55
% households without a vehicle	25	22
N	**85**	**1152**

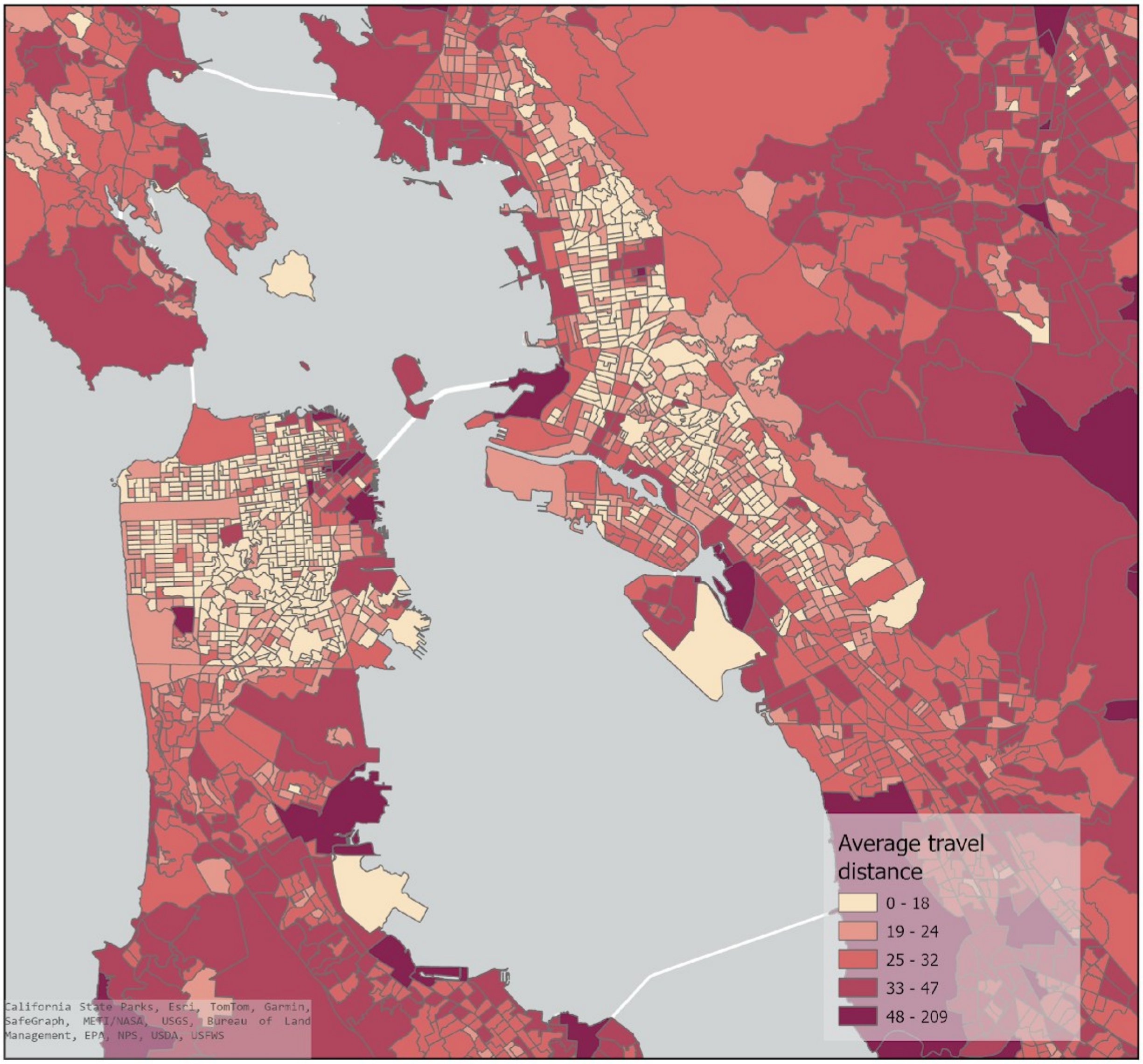
As a control for the second set of analyses that includes all possible treatment type (AHSC residential and transportation elements, market rate housing, and non-AHSC affordable housing, see Table 1.1 for a summary of each treatment type), we use block groups where no project of any kind was developed in the period of study. We also want block groups that represent a wide range of possible contexts to reflect the diversity of places in California where housing has been developed. We have the benefits of being able to draw from a large number of block groups which strengthens the model. For this second set of analyses, we randomly selected 3,000 block groups within 50 miles of an AHSC project as the control cases.

Descriptive Statistics

Affordable Housing and Sustainable Communities Program

The map in Figure 1.3 shows the estimated travel distance measure for the San Francisco-Oakland Metropolitan Area. The pattern of lower travel activity block groups matches with areas of high accessibility (e.g., central Oakland, Berkeley, central San Francisco), while block groups on the outskirts of the cities have higher travel activity. Like other mobile-phone-derived data, there is little ability to differentiate between modes and trip purposes. Therefore, an important limitation of these data is that they include all modes, including walking. The other major limitation is that the sampling rate is variable at the level of the POI, which affects how many people can be tied to an origin block group. As such the number of devices tied to an origin block group can vary widely. This is one reason we cannot reliably estimate total VMT since, in the absence of a sampling rate, we cannot adjust the total distance traveled. This is not material to our analysis, however, since our focus is on per capita VMT.

Figure 1.3: Average Travel Distance from Home block group to Destination block group for the San Francisco – Oakland Metropolitan Area



Source: Author calculations of Advan Neighborhood Patterns data

Results

We divide the results between the analysis focused on effects associated with the addition of an AHSC project in a neighborhood and the analysis focused on the types of neighborhoods within which AHSC projects are located.

Impact Within Neighborhoods

The results from the first set of models focused on AHSC projects return no significant coefficients (

Table 1.3). We cannot say with confidence that the estimates are different from zero, that is, we cannot rule out that the AHSC project has no effect on average travel activity within the neighborhood. With this caveat in mind, note that the coefficients are *positive* in all models and change little across specifications. The addition of an interaction with a transportation element being associated with the residential project suggests a negative effect, but, again, the estimate is not statistically significant and has no effect on average travel activity within the Census block group.

The third model in

Table 1.3 examines the interaction between the completion of an AHSC project and whether the project is in a high, medium, or low travel activity neighborhood. These neighborhood types were defined so that as many treated observations, about a quarter, were in the high and low categories, with the rest in the medium category. The high cutoff corresponds to neighborhood with average travel distance 60% or above that of the county average. The low cutoff corresponds to a neighborhood average of 40% or below the county average.

Most housing developments (of all types) were in block groups with average travel activity for the entire period lower than the county average (see Figure 1.3). The high shares are to be expected because our sample only includes multifamily projects with more than 23 units, which tend to be in higher density locations within urban areas, the kind of locations with lower travel. The cutoffs are well below the county average because most AHSC projects are in locations where travel distance tends to be significantly lower than the county average. The restriction of the sample to such locations means we are excluding most locations with higher travel activity and skewing the distribution downward. Based on these cutoffs, the results suggest that locations that are in lower travel neighborhoods are associated with a decrease in travel distance, while higher travel areas are associated with an increase.

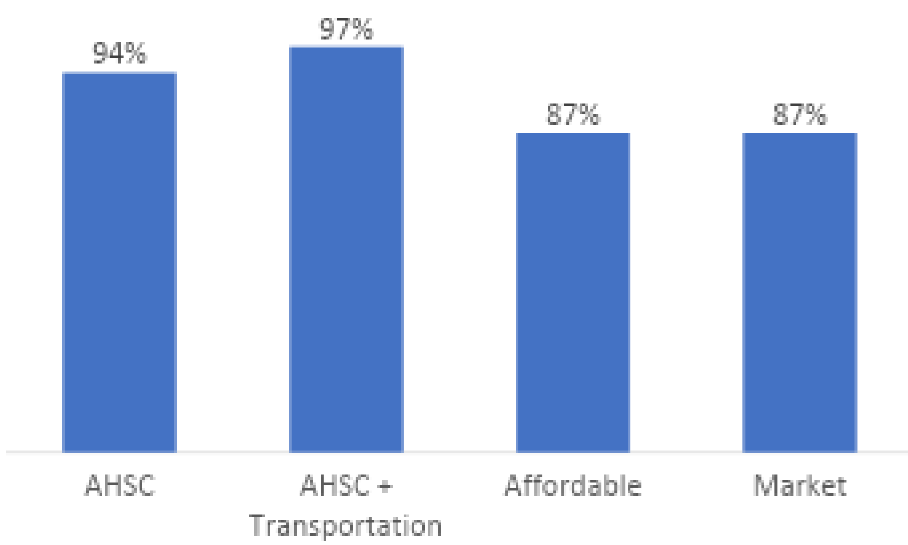
Table 1.3: Models based on AHSC projects only

Project type	AHSC project	AHSC project with transportation element	AHSC projects in high and low VMT areas	Staggered model
AHSC	0.251	0.266	0.372	-0.078
AHSCxTransportation	NA	-0.046	NA	NA
AHSCxLow travel	NA	NA	-0.846	NA
AHSCxHigh travel	NA	NA	0.538	NA

Notes: All models include a control for the number of devices from which the measure of travel distance was constructed. Standard errors are clustered at the project/neighborhood level. None of the coefficients are statistically significant.

The last model applies a method to account for the staggered nature of the data. Using this method, the results show a negative association, but the model is unreliable due to the small number of observations in each month (see Figure 1.4). In short, the estimates point to a consistent failure to estimate reliably the impact of AHSC projects. This may be due to the quality of the data or the small number of projects from which to derive an estimate.

Figure 1.4: Project Location Relative to County Average Travel Distance by Treatment Type



The second set of models uses a larger set of data to investigate how AHSC developments might compare to other types of residential developments. An advantage of this approach is that the larger sample size for non-AHSC projects can offer some insight about AHSC-funded projects, especially if we consider that non-AHSC funded residential projects are likely to have similar effects. Like the results in

Table 1.3, most estimates in Table 1.4 fail to reach statistical significance. The coefficient on “market rate” projects, which has the largest sample size, is the only one to be significant and indicates a negative association with travel distance. The coefficient suggests that the addition of a market rate building in a block group is associated with a decrease of 0.15 miles in average travel distance. The addition of interactions with the neighborhood travel distance level suggest that the addition of residential buildings in high VMT locations tend to increase average travel distance, and the opposite in low VMT areas. We find something similar for non-AHSC affordable housing projects: travel distance increases when projects are placed in high-travel locations. The final model in Table 1.4 (rightmost column, below) includes a measure of the difference in size across projects, and finds no statistically significant effect associated with project size.

Table 1.4: Models based on all types of developments

Project type	Project binary	Project binary by travel level	Number of units (10 unit increments)
AHSC – Residential	-0.162	0.017	-0.011
AHSC - Transportation	-0.263	0.139	0.002
Market	**-0.150**	**-0.540**	-0.001
Affordable	-0.088	**-0.431**	NA
AHSC – Residential x Low travel	NA	-0.647	NA
AHSC – Residential x High travel	NA	0.065	NA
AHSC – Transportation x Low travel	NA	-0.802	NA
AHSC – Transportation x High travel	NA	-0.625	NA
Market x Low travel	NA	-0.078	NA
Market x High travel	NA	**0.824**	NA
Affordable x Low travel	NA	0.330	NA
Affordable x High travel	NA	**0.697**	NA

Notes: All models include a control for the number of devices from which the measure of travel distance was constructed. Standard errors clustered at the project/ neighborhood level. Bolded coefficients are significant at the 95% level or above.

Impact Conditional on Neighborhood Type

The previous section focused on estimating the impact of new developments on neighborhood average travel distance by comparing each neighborhood to itself over time. This approach is useful for assessing whether the new development itself has an effect, but it says little about the type of neighborhood where change is taking place. This section takes a different approach that allows us to include neighborhood characteristics.

The results above suggest that the completion of an AHSC project has no discernible effect on average travel distance. However, this is not to say that all projects were in neighborhoods where average travel distance did not change before and after the project was completed. Table 1.5 breaks down projects based on how much per capita travel distance decreased or increased based on the difference of the average of the four months preceding the completion and the average for the four months following the completion. It shows that block groups at both extremes tend to have higher average travel distance. An important distinguishing factor is that block groups where large decreases happen have lower shares of people driving alone, more households with no vehicles, and higher walkability scores.

Table 1.5: Characteristics of block groups where AHSC projects were associated with the top 25% of increases and decreases in average per capita travel distance, and block groups with increases and decreases below these thresholds

Variable	Large increase (+1.2 mi)	Increase	Reduction	Large reduction (-1 mi)	California
Travel distance (in miles)	16.6	10.8	11.4	14.4	13.8
% of who drive alone	66.2	57.3	73.8	42.7	73.7
% whose commute is more than 30 minutes	37	45.6	46.9	50.2	44
% with no vehicle	20.7	28.8	15.5	31.7	7.1
Median household income (in $1,000s)	57.6	49.5	55	58.2	84.7
Share of projects with no transit access	28.6	18.2	10	4.8	26.3
Distance from to nearest transit stop (meters)	314	278	410	326	469
Jobs within 45 mins by car (in 1,000s)	139	253	279	167	150
Walkability index [0 (low) - 20 (high)]	13.3	14.8	14.3	16.1	12.2
N	**21**	**22**	**20**	**21**	**23,212**

We look at these differences with a statistical model that builds on the measure of difference in travel distance before and after the completion of a project. We constructed a variable that compared the average travel distance four months before and after the completion of a development relative to the county average.[1] Table 1.6 shows an example of the outcome of interest (travel distance) where the county increased by one mile while the block group increased by two miles. The adjusted difference (adjusting for county wide increase) is a one mile increase.

Table 1.6: Illustrative example of the county-adjusted change in travel distance

Variable	Before (B)	After (A)	Difference (A-B)
Block group travel distance (X)	10	12	+2
County travel distance (Y)	20	21	+1
County-adjusted difference (Y-X)	10	9	+1

The results in Table 1.7 are largely confirmatory of the results in the previous section. There is no significant association between AHSC projects and change in travel distance. However, this model shows a strong negative association for projects that included a transportation element. The estimates for housing development type are relative to market-rate developments, suggesting that affordable housing appears to be associated with decreases relative to market-rate housing, and that there does not seem to be a difference between AHSC-funded housing and other affordable housing.

None of the neighborhood variables are significant except the measure of job accessibility. The measure of accessibility, based on the number of jobs accessible within 45 minutes by car shows a significant and negative association with change, suggesting that locations that have high accessibility tend to experience decreases in travel distance.

Table 1.7: Regression results for county-adjusted change in travel distance for all types of projects. The category of reference of project type is market-rate housing

Variable	Estimate
Intercept	0.847
AHSC	-0.222
AHSC + Transportation	***-0.940**

[1] We tried different time bands up to 6 months before and after, and while results were consistent, the 4 month window provided greater level of significance.

Variable	Estimate
Affordable	-0.203
Block group average travel distance	-0.010
% who use transit	0.007
% who drive alone	-0.004
% without access to vehicle	-0.004
% renter	0.000
% below poverty line	0.009
Jobs accessible within 45 minutes by car (in 100,000s)	****-0.162**
Walkability index	0.000
N	**1358**
R^2	**0.072**

Notes: Significance levels: * 0.05 ** 0.01. The model uses year fixed effects

Task 2: Estimate cross-sectional household VMT variation as a function of transit proximity and other spatial characteristics

Key Research Question & Basic Theory

Current AHSC methodology limits measures of projects to the building level with rare exceptions. All AHSC projects are required to be within 0.5 miles of an existing or planned station or stop providing sufficiently frequent high quality transit service. But even in that case, transit proximity is not directly considered in the methodology–rather, it determines the project area type (TOD, ICP, or RIPA) which then groups projects together to be ranked comparatively to one another. Evidence suggests that within that radius, the closer to a transit stop the lower the VMT, for both low- and high-income households. We evaluated the impact of proximity closer than 0.5 miles on VMT reduction using household travel survey data. We also investigated other land use characteristics including other transit metrics, job accessibility measures, and neighborhood- and community-level characteristics. These would be new additions to the AHSC methodology which is focused on site-specific characteristics.

We used the most recent publicly available household travel survey in California—the National Household Travel Survey (NHTS) of 2017—to estimate household VMT as a function of built environment characteristics, as described in more detail below, including by distance to transit using discrete distance rings (e.g., at the 1/4 mile increment), and controlling for other household characteristics. Such an analysis had not been done for the NHTS in California. We obtained restricted, remote access to NHTS data with geocodes of home locations which were needed for the analysis.

In carrying out this analysis we also accounted for additional typically unmeasured factors that are also likely to play an important role in VMT reduction; as has been noted in previous research, a number of other factors are critical in influencing vehicle use, some of them substantially more important than access to transit. We took other steps to go beyond the scope of previous literature by accounting for some specific concerns in our model specifications, such as accounting for interactions between built environment measures, and accounting for the geographical scale of influences on household travel behavior by testing built environment measures at different spatial scales.

Materials & Methodology

VMT and the 2017 NHTS

We relied on the restricted, confidential version of the California add-on to the 2017 National Household Travel Survey (NHTS), with spatially precise household locations and disaggregated demographic characteristics. Disaggregate data were used in this analysis due to the nature of the regressions conducted to estimate individual person miles traveled. Our analysis is dependent on having person-level data, including demographics and household locations to determine neighborhood characteristics. The data were accessed via a remote system maintained by the Transportation Secure Data Center, with approval from the National Renewable Energy Laboratory (NREL). To access the restricted disaggregate data, NREL must approve an application submitted by the prospective user. Application instructions and materials can be accessed at the NREL website. The survey had a 13% response rate; the dataset consists of 26,095 households across the state, with more than 55,000 individuals represented.

To calculate daily VMT per survey respondent, we used the confidential version of the NHTS trip file. The trip file contains records of every trip reported by every survey respondent who is 5 years of age or older. Within the trip file, trips are reported by individuals, not by households. For instance, if two household members traveled together, it would be reported twice in the trips dataset, with identical travel mode, number of passengers, and distance traveled.

We began by identifying all trips that were taken in a private vehicle, which we defined as a trip taken by car, SUV, van, pickup truck, motorcycle/moped, RV, taxi/limousine (including Uber/ Lyft), or rental car. If there were any passengers on the private vehicle trip, the total VMT for the trip was divided evenly between all individuals on the trip. Individuals on the trip could include household members or others. The distance for each trip that was taken in a private vehicle was matched to any survey respondents on that trip and was summed up over all trips for those individuals to give the total daily VMT per survey respondent.

Changes in the process for reporting trip distance for the 2017 NHTS increased the average daily VMT for respondents significantly, when compared to previous survey years. As the documentation notes, “a web-based tool mapped the origin and destination of each reported trip while collecting other core data… Because the web-based travel day retrieval was primarily self-reported data, there are some very large values for some of the variables describing distances” (Federal Highway Administration, 2018). Due to this caveat, we removed the 109 respondents who reported VMT exceeding 300 miles on the survey day (0.2 percent). After filtering, we were left with 55,683 total observations, out of 55,792 total individuals in the dataset.

We faced some challenges with verifying the reliability and accuracy of the restricted NHTS dataset which we identified for NREL and NHTS staff and subsequently resolved with their cooperation and assistance (see Appendix Section B.1: NHTS data discrepancies**)**.

Explanatory variables for regression analysis

Our main explanatory variables of interest in regression analysis were the built environment characteristics (see Table 2.1 for a summary of all variables, and Appendix Table A.3 for a full list with descriptive statistics). We tested different scales of measurement for these measures, given that recent research has shown that larger scales play an important role in influencing auto use (this work is described in more detail within the next section). Variables that were calculated based on aggregating Census block groups (BGs) to larger spatial units are indicated in Table 2.1.

Raw spatial files for rail stations and bus stops came from transit agencies and metropolitan planning organizations across the state, collected by the authors in previous research projects. Unfortunately the data collected at this time did not include reliable statewide information on routes, peak hours, or waiting times at peak hours, so we were unable to construct any variables related directly to High Quality Transit (HQT) as is defined in the AHSC Program Guidelines. As of 2021, more comprehensive GTFS data on statewide transit stops is available via the California Open Data Portal and can be used to update the built environment characteristics in the future, but to conduct analysis using 2017 travel survey data we could not rely on data from 2021.

Distance to rail was calculated directly in the remote environment by measuring from the geocoded household location to the closest rail station along a pedestrian network defined as described in Appendix B.3. We tested a number of specifications of this variable but found that threshold values were most statistically significant. After testing a number of these in one-eighth-mile increments we found the most statistically robust measures were at the 0.25 mile and 0.5 mile limits. For more information on how the distance to rail variable was computed, see Appendix Section B.2: Network distance to rail.

We spent a great deal of effort assembling a statewide map of bus stops that was consistent across transit agencies in the state for a cross-sectional snapshot as close as possible to the NHTS dataset (ranging from 2016 to 2018). Once we had assembled the dataset we calculated both Census block group-based metrics and measures based on airline radii centered on rooftop geocodes for households (for an illustration of the latter see Figure 2.1). For more information on how the bus stop dataset was cleaned and assembled, see Appendix Section B.3: Cleaning bus stop data

Figure 2.1: Illustration of bus stop density measures, by radius, around household geocodes (Santa Cruz, CA)

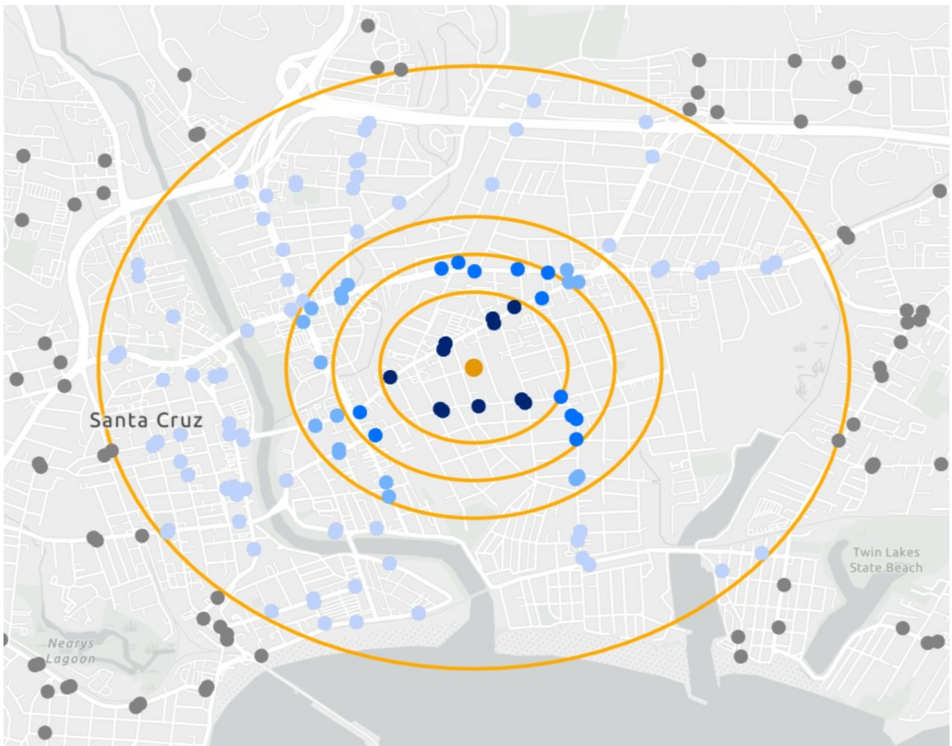


Table 2.1: Summary of calculated built environment variables for regression analysis

Name	Notes
Distance to rail (miles)	Network distance from household location, by drive or by walk. Represented in the model in both continuous distance and dummy variable form at one-eighth mile radii out to one mile
Bus stop density (stops/mi^2)	Calculated two ways: block group (BG-based) and at straight line radii from households (HH-based) (BG-based: block group, 0.5 mi, 1 mi, and 2 mi radius; HH-based: 0.25 mi, 0.375 mi, 0.5 mi, 1 mi)
Population density (residents/mi^2)	(block group, 0.5 mi, 1 mi, and 2 mi radius)
Employment density (workers/mi^2)	By wage and by industrial sector (wage: High, Medium, Low; sector: Retail, Entertainment, Office, Industry, Service, Education, Health, Public Service) (block group, 0.5 mi, 1 mi, and 2 mi radius)
Employment accessibility (# of jobs)	Number of jobs (workers) within a 45-minute trip via auto and transit
Ratio of transit to car accessibility (unitless)	Interaction variable; ratio of the number of jobs within a 45 minute commute by transit to the number of jobs within 45 minutes by car commute
Street density (mi/mi^2)	Calculated using OpenStreetMap data w/out private roads (for current 2024 data and archived 2018 data, see Appendix Section B.4: Network identification & density calculation for more detail**)** (block group, 0.5 mi, 1 mi, and 2 mi radius)
Intersection density (intersections/mi^2)	Calculated with OpenStreetMap data excluding private roads (for only current 2024 data, see Appendix Section B.4: Network identification & density calculation for more detail)

Name	Notes
	(block group, 0.5 mi, 1 mi, and 2 mi radius)
Network load density (people/road mi)	Interaction variable; sum of residents and workers, divided by the street length in a given area (block group, 0.5 mi, 1 mi, and 2 mi radius)
Core Based Statistical Area (dummy variable)	Dummy variables for each CBSA to indicate if the household is located in that respective CBSA including Riverside-San Bernardino-Ontario, Los Angeles-Long Beach-Anaheim, Sacramento-Roseville-Folsom, San Diego-Chula Vista-Carlsbad. San Francisco-Oakland-Fremont, and San Jose-Sunnyvale-Santa Clara.

Notes: Sources are described in the text.

Population densities, employment densities, and employment accessibilities were all sourced from the 2021 Environmental Protection Agency Smart Location Database (EPA SLD), which in turn used other data sources to construct their variables. The population density variable was calculated using 2018 American Community Survey (ACS) 5-year estimates and employment density variables were created using 2017 Longitudinal Employer-Household Dynamics (LEHD) Workplace Area Characteristics (WAC) data, both from the US Census Bureau. We tested workplace-area employment densities by industrial classification as a measure of mixed or "supportive" land uses. For employment accessibility by car and public transportation, EPA SLD used the 2020 Google TravelTime API and 2020 General Transit Feed Specification (GTFS) data, in addition to 2017 LEHD data. We specified an interactive variable, the "transit to car accessibility ratio," to capture the comparative ease one has in reaching employment via one mode over the other. The higher the value of the variable, the more competitive public transportation is with the auto in terms of travel time.

Street and intersection densities were calculated based on 2018 data from OpenStreetMap, archived by GeoFabrik, as described in Appendix Section B.4: Network identification & density calculation**.** Network load density is an interaction variable, indicating the potential level of road congestion; it is the sum of the total population and employment in a given area (such as a Census block group), divided by total street length. Note that using road lane miles in either the street or network load density calculations would provide a more granular metric than road length, but Geofabrik data did not include reliable information on lane miles for the state.

The core based statistical areas (CBSAs) of household residential locations were provided in the 2017 NHTS data.

While we had hoped to include a metric for parking in our model, it was not possible to do so. There is a significant lack of accurate parking data at the statewide level. While the literature has explored the impact of parking supply and price on travel behavior, as discussed in Task A, we were not able to use comparable metrics during our analysis.

We used the demographic variables listed in Table 2.2 as controls. (Note that the US Census and the NHTS define “ethnicity” as Hispanic status). We do not include a variable for car ownership, as car ownership is endogenous with driving. In other words, individuals who drive more are also more likely to own a car, and vice versa, making it impossible to understand the direction of causality between the two. Including car ownership as an explanatory variable would therefore bias our estimates.

Table 2.2: Summary of prepared demographic variables for regression analysis

Name	Notes
Age	Open-ended brackets: under 5, categorical variables increasing in intervals of 10 years, ending with 75 or older
Household income	Open-ended brackets: above $10,000, $15,000, $25,000, $35,000, $50,000, $75,000, $100,000, $125,000, $150,000, and $200,000
Education	Open-ended brackets: categories include some high school, high school, some college or community college degree, university degree, graduate school degree.
Household size	Number of people in each household
Presence of young children	Dummy variable indicating one or more young children (0-4) in household
Sex	Categories include female, male, and missing
Race	Categories include White, Black, Native American, Asian, Pacific Islander, Multiple races, Other race, and Missing
Ethnicity	Hispanic, non-Hispanic, or Missing
Housing tenure	Own, rent, other, and missing

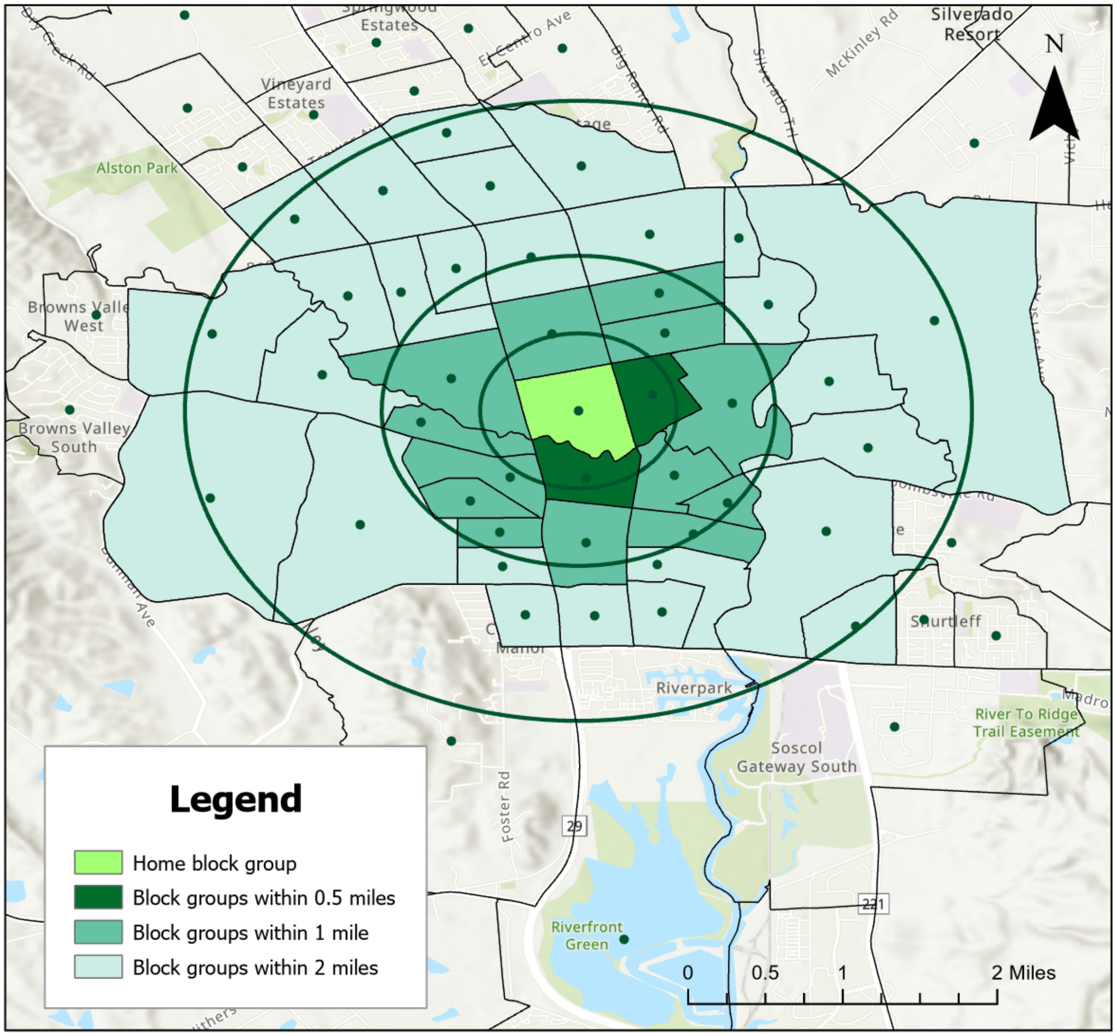
Notes: Source: National Household Travel Survey California add-on survey, 2017

Multiple levels of spatial measurement

In addition to measuring densities at the block group level, we also measured them at larger aggregated scales. For bus density, we were able to calculate measures using straight line radii from the household location. The remaining built environment densities were calculated at the block group level since those data (such as population and employment) were aggregated to zones.

We chose to build larger spatial areas for measurement by aggregating block groups that fell within a certain radius of the main block group of interest. To aggregate block groups we constructed a half mile, one mile, and two-mile geodesic radius around the centroid of each block group. For any given block group of interest, we identified all other block groups whose centroid fell within that radius. For a visualization of this process, see Figure 2.2. We then computed the density for that larger area by averaging the densities of those block groups; weighted by the area of each block group.

Figure 2.2: Aggregating block groups for different spatial scales (Napa, CA)



Statistical models

For our analysis, we used left-censored Tobit models to account for the significant proportion of individuals who reported no VMT on the given survey day (28.63% of the total sample). All variables were scaled and converted to Z-scores in preparation for regression analysis, except for the transit to car accessibility ratio (which is unitless). This was done for interpretability of results, enabling coefficients associated with a given variable to be interpreted as the amount of VMT associated with an increase of one standard deviation in that variable of interest. We used robust standard errors and clustered by block group to account for correlation among built environment variables for individuals in the same block group. Given the large number of built environment variables collected at multiple spatial scales and the strong collinearity between many, estimating a robust model required selecting a subset of built environment variables to include in the model, that would be the most predictive of spatial variations in VMT. We carried out a theory driven process in which we made several adjustments to the variables and scales of measurement included in our model to best fit the data before deciding on our final specifications.

Once we determined the variables of interest, we applied these to both an initial diagnostic model and to an interaction model, where additional variables were specified differently for observations with household income below $50,000 (N=18,398). This income threshold was selected by analyzing the population-weighted median Multifamily Tax Subsidy (MSTP) Income Limits for all California counties in 2017, as well as the 50% and 60% below median income thresholds. In doing so, we can more directly address the impact that affordable housing built under the AHSC program may have.

Our base model (with no interactions) is specified as follows:

Equation 2.1: Task 2 base full sample model specification

$$VMT_i = \beta_1 x_i^1 + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 + \beta_6 x_i^6 + \beta_7 x_i^7 + \beta_8 x_i^8 + \beta_9 C_i + \beta_{10} D_i + \varepsilon_i$$

Table 2.3: Variable definitions for Equation 2.1: Task 2 model specification

Variable	Variable definition	Units
VMT_i	Daily vehicle miles traveled of the ith observation in the sample.	Miles/day
β_n	The coefficient associated with the nth built environment characteristic for n= 1, 2, ..., 8. n=1: Network load density (2 mi radius) n=2: Street density (2 mi radius) n=3: Bus stop density (BG based, 1 mi) n=4: Population density (2 mi radius) n=5: Retail density (block group) n=6: Has rail access within 0.25 mi n=7: Ratio of transit:car job accessibility n=8: Jobs accessibility by car	Miles/day per standard deviation units; see Table 2.5.
x_i^n	The z score of the nth built environment variable, for the ith observation.	Standard deviation units; see

Variable	Variable definition	Units
		Table 2.5. For raw units by variable, see Table 2.1.
β_9	A vector of coefficients associated with the CBSA of residence.	NA
C_i	A vector of dummy variables indicating if the ith observation is in a CBSA. The sum of values in the vector is no greater than 1, as an observation can only reside in one CBSA.	NA
β_{10}	A vector of coefficients associated with the demographic control variables.	NA
D_i	The vector of demographic control variables.	NA For raw units by variable, see Table 2.2.
ε_i	The error term.	Miles/day

For the latter model, interaction terms were set to be their original value if the given respondent fell into the low income group and set to zero if the respondent did not. With this model, we are able to specifically test if coefficients differ between the groups and see the extent to which certain variables are statistically significant in comparison to the higher-income subset of population. Note that we only included interaction terms for built environment variables, and not the demographic control variables. Our base model with low income interaction terms is defined as follows:

Equation 2.2: Task 2 base full sample with low income interactions

$$VMT_i = \beta_1 x_i^1 + \beta_1^* x_i^{1*} + \beta_2 x_i^2 + \beta_2^* x_i^{2*} + \beta_3 x_i^3 + \beta_3^* x_i^{3*} + \beta_4 x_i^4 + \beta_4^* x_i^{4*} + \beta_5 x_i^5 + \beta_5^* x_i^{5*} + \beta_6 x_i^6 + \beta_6^* x_i^{6*}$$
$$\beta_7 x_i^7 + \beta_7^* x_i^{7*} + \beta_8 x_i^8 + \beta_8^* x_i^{8*} + \beta_9 C_i + \beta_{10} D_i + \varepsilon_i$$

Table 2.4: Variable definitions for Equation 2.2: Task 2 base full sample with low income interactions

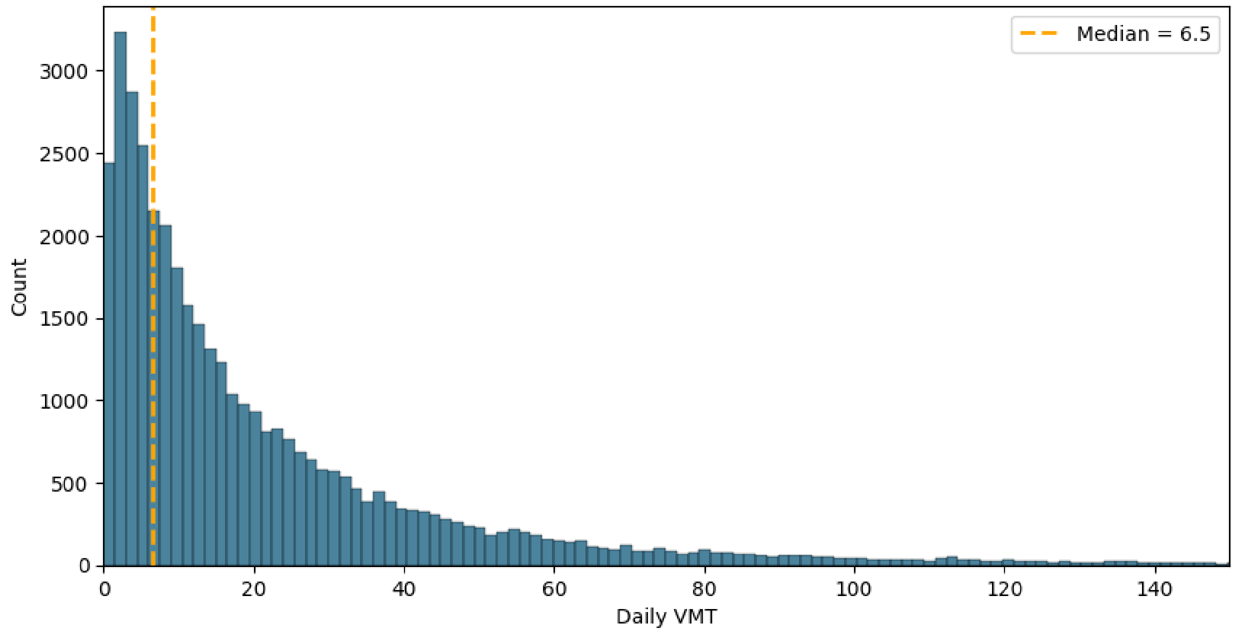
Variable	Variable definition	Units
VMT_i	Daily vehicle miles traveled of the ith observation in the sample.	miles/day
β_n	The coefficient associated with the nth built environment characteristic for n= 1, 2, ..., 8. n=1: Network load density (2 mi radius) n=2: Street density (2 mi radius) n=3: Bus stop density (BG based, 1 mi) n=4: Population density (2 mi radius) n=5: Retail density (block group) n=6: Has rail access within 0.25 mi n=7: Ratio of transit:car job accessibility n=8: Jobs accessibility by car	Miles/day per standard deviation units; see Table 2.5.
x_i^n	The z score of the nth built environment variable, for the ith observation	Standard deviation units; see Table 2.5. For raw units by variable, see Table 2.1.
β_n^*	The interaction coefficient associated with the nth built environment characteristic for $n = 1, 2, \ldots, 8$.	Miles/day per standard deviation units; see Table 2.5.
x_i^{n*}	The nth interactive built environment variable, for the ith observation. The variable is defined as follows: • $x_i^{n*} = x_i^n$ if the ith observation has household income less than $50,000. • $x_i^{n*} = 0$ otherwise.	Standard deviation units; see Table 2.5. For raw units by variable, see Table 2.1.

Variable	Variable definition	Units
β_9	A vector of coefficients associated with the CBSA of residence.	NA
C_i	A vector of dummy variables indicating if the ith observation is in a CBSA.	NA
β_{10}	A vector of coefficients associated with the demographic control variables.	NA
D_i	The vector of demographic control variables.	NA
ε_i	The error term.	Miles/day

Descriptive Statistics

Figure 2.3 shows the distribution of daily VMT values among all survey respondents with non-zero daily VMT. Many people did not travel on the survey day, making up 15,943 out of 55,683 total responses (29%). The mean daily VMT for the entire survey sample is 17.1 miles with a standard deviation of 29.1 miles. The median daily VMT is notably lower at 6.5 miles, indicating a significant right-skew in the data distribution (as seen in Figure 2.3). As can be seen below, most observations had relatively low VMT: 91% of all observations had daily VMT under 50 miles, 98% were under 100 miles, and 99% were under 150 miles. We do not present the distribution for observations that had VMT in the range (150, 300] given that they account for only 0.04% of observations, though they are included in the analysis.

Figure 2.3: Distribution of daily VMT variable for non-zero observations (N=39,215)



Notes: Source is 2017 NHTS CA-add on survey. Universe includes all observations with daily VMT with nonzero daily VMT less than or equal to 150 miles (N=39,215). Orange dotted line indicates the median VMT (6.5 miles). Total sample with daily VMT less than 300 miles is 55,683. The number of entries with no VMT logged on the travel day is 15,943 (28.63%). The number of entries with VMT in the range (150, 300] on the travel day is 525 (0.94%).

As seen in

Table 2.5 below, the built environment context for survey respondents shows substantial variability, with most variables skewing to the right. Variables that are more evenly distributed include street density, network load density, and population density. For some variables we see more skewed distributions, including retail density, bus stop density, transit to car jobs accessibility ratio, and jobs accessibility by car. This is especially true for retail density, where we see a mean over sixteen times as high as the median value. Bus stop density is similarly heavily skewed, with most observations having close to zero stops/mi^2, and only a few having significantly higher stop densities. We see the same pattern with the transit to car jobs accessibility ratio, where the median is less than a third of the mean. Within our sample there are very few respondents who live close to rail, with only 3% of the population living within a half mile of a rail station and 1% within a quarter mile.

Table 2.5: Descriptive statistics for all built environment variables in Task 2 (N =55,683)

Variable	Mean	Std dev	Q1	Median	Q3
Population density (2 mi radius)	4,375.00	4,276.86	1,344.63	3,721.36	6,001.28
Retail density (block group)	194.54	691.48	0	11.71	128.15
Rail within 0.25 mi (DV)	0.01	0.11	0	0	0
Rail within 0.5 mi (DV)	0.03	0.17	0	0	0
Bus stop density (BG-based, 1 mi radius)	10.84	19.67	0	3.60	14.03
Network load density (2 mi radius)	516.44	491.65	255.80	438.55	641.59
Transit to car jobs accessibility ratio	0.48	0.77	0	0.15	0.71
Jobs accessibility by car	76,305.26	98,141.01	11,252	38,470	105,027
Street density (2 mi radius)	10.22	4.84	6.64	11.11	13.94

Notes: Universe (N=55,683) does not include observations with daily VMT greater than 300. Source is 2017 Environmental Protections Agency Smart Location Database for population density, retail density, and jobs accessibility by car. OpenStreetMap data was used for street density and was accessed using the OSMnx Python package. 2017 rail station and bus stop data was sourced from transit agencies and regional planning agencies and was retained by the authors from previous work. Network load density and transit to car jobs accessibility ratio are interaction terms that were computed from other variables. Units for all variables can be found in Table 2.1.

As shown in Table 2.6, 20% of survey respondents had an annual income of less than $35,000, and 40% made $100,000 or more. While there is a much larger percentage of missing data on education, 68% of the sample has completed some form of higher education. Notable for the Task 4 work, 26% of our sample is 65 years of age or older. Racial distribution relatively matches the statewide distribution with slight overrepresentation of White individuals and underrepresentation of Asian and Black individuals. The same cannot be said for ethnicity, where only 15% of the respondents identified as Hispanic, compared to the statewide value of 40% (2017 ACS 5-year estimates, Table DP05). Homeowners are also overrepresented in this survey, as they make up only 56% of the statewide population (2017 ACS 5-year estimates, Table DP04). Some of these incongruencies are likely due to survey response bias. While some inconsistency exists in the data, the overall average household size for the survey is similar to the statewide mean (2017 ACS 5-year estimates, Table DP02). Fortunately, our analysis here does not rely on representativeness of the NHTS sample in comparison to the state, because we control for all demographic characteristics in regression analysis and the sample sizes for all subpopulations of interest are sufficiently large for robust statistical tests.

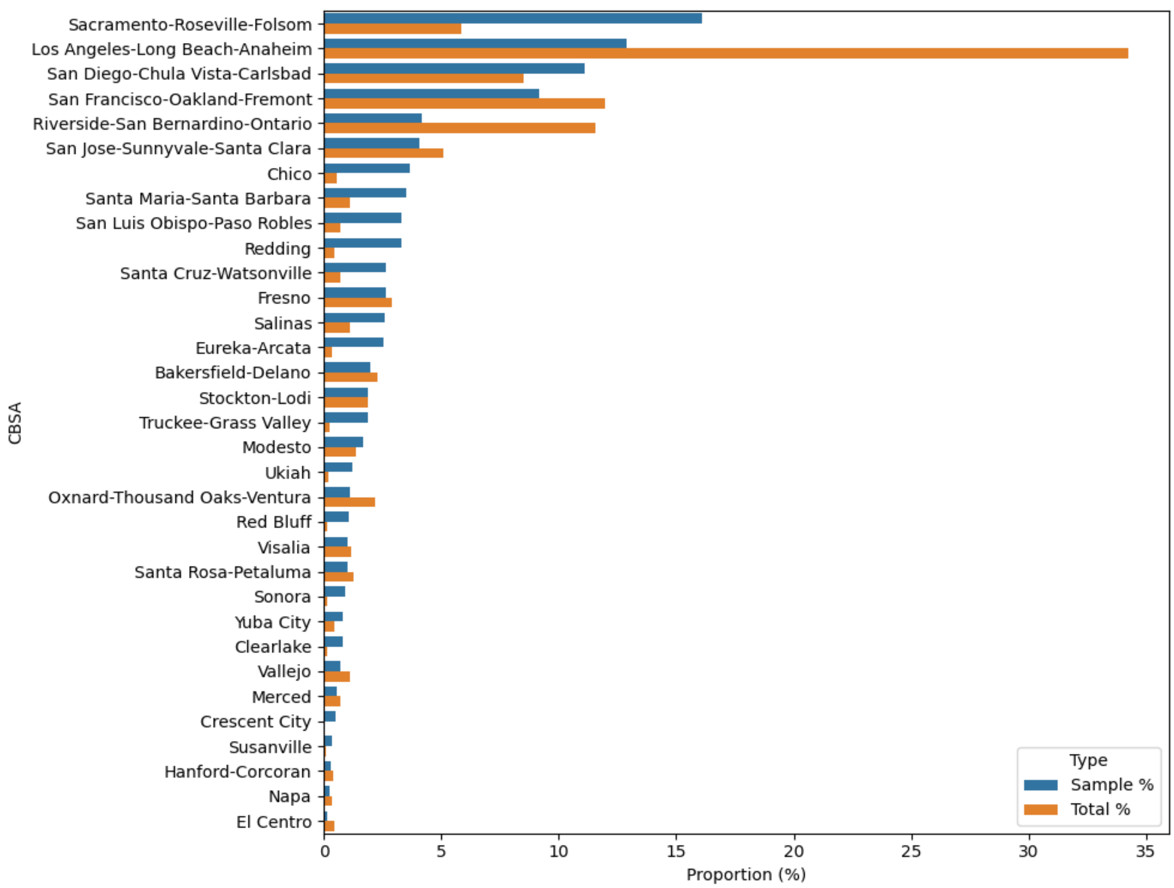
Table 2.6: Descriptive statistics for demographic variables used in Task 2 (N=55,683)

Variable	%	Variable	%
Income		**Race**	
Less than $10,000	3	White	74
$10,000 to 14,999	3	Asian	10
$15,000 to 24,999	7	Multiple responses selected	6
$25,000 to 34,999	7	Some other race	5
$35,000 to 49,999	9	Black or African American	3
$50,000 to 74,999	15	American Indian or Alaska Native	1
$75,000 to 99,999	14	Native Hawaiian or Pacific Islander	1
$100,000 to 124,999	12	Missing data	1
$125,000 to 149,999	7	**Ethnicity**	
$150,000 to 199,999	8	Non-Hispanic	85
$200,000 or more	10	Hispanic	15
Missing data	3	Missing data	0
Education		**Age**	
Less than high school	6	Under 5	4
Completed high school	13	5-14	9
Associates Degree	27	15-24	7
Bachelor's Degree	21	25-34	10
Graduate Degree	20	35-44	11
Missing data	12	45-54	13
Tenure		55-64	18
Own	72	65-74	16
Rent	27	75 or older	10
Other	1	Missing data	0
Missing data	0	**Sex**	
Young Children		Female	52
Has young children in household (DV)	13	Male	48
Variable	**Average**	Missing data	0
Household size	2.80		

Notes: Source is 2017 NHTS CA-add on survey. Universe does not include observations with daily VMT greater than 300 (N=55,683). Percentages showing 0 represent small non-zero percentages that have been rounded down. Due to rounding, some categories may not sum to 100%.

As demonstrated in Figure 2.4 below, we also see disparities when it comes to the distribution of people by geographical region. There is significant overrepresentation in the Sacramento-Roseville-Folsom area, some slight overrepresentation of the San Diego area and less urban CBSAs, and comparative underrepresentation in Los Angeles, the Inland Empire, and the greater Bay Area. It is not exactly specified why this may have occurred within the NHTS data documentation, although one or more CBSAs may have paid for additional sample collection. As with the demographic data noted above, the spatial representativeness of the data does not affect the controlled results here except to the extent that larger sample sizes in different CBSAs affects the statistical significance of CBSA-specific dummy variables.

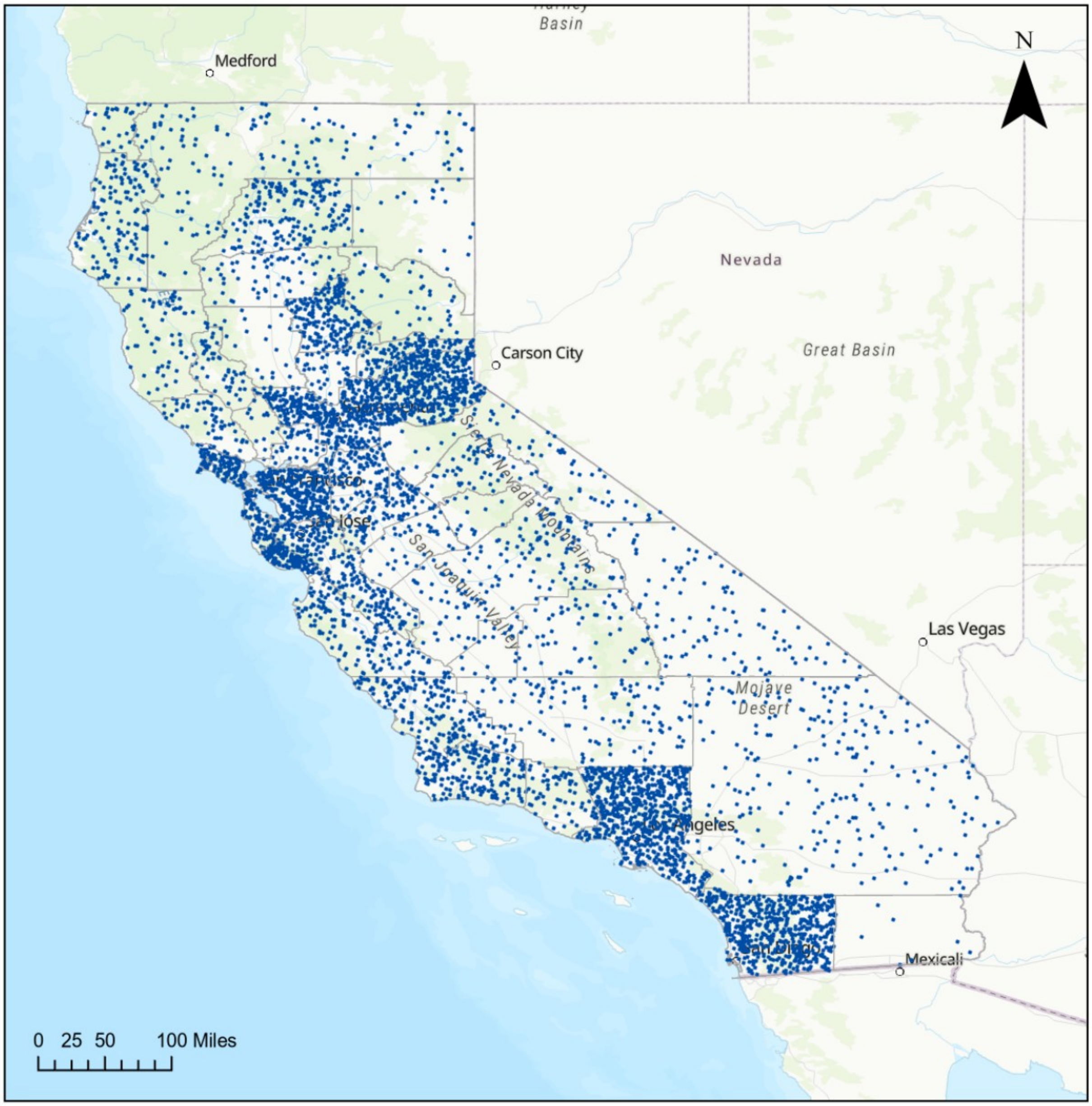
Figure 2.4: Survey respondent count by CBSA, compared to the general population (N=53,463)



Notes: Source is 2017 NHTS CA-add on survey for sample percentages and 2017 ACS 5-year estimates for total population percentages. Universe for sample percentages is observations with daily VMT of 300 miles or less that reside in a CBSA (N=53,463). 2,220 survey respondents (4% of the total sample) did not belong to a CBSA and are not shown in this chart. It is not specified in the 2017 ACS estimates what proportion of California residents do not reside in a CBSA.

The dot density map in Figure 2.5 is a visual representation of survey respondent locations throughout the state, based on their home CBSA. Note that these are not the exact home locations of respondents, but instead an approximate representation. Due to confidentiality we are not allowed to display the exact home locations of survey respondents. The actual distribution of survey respondents is focused on urbanized, denser neighborhoods within the CBSAs.

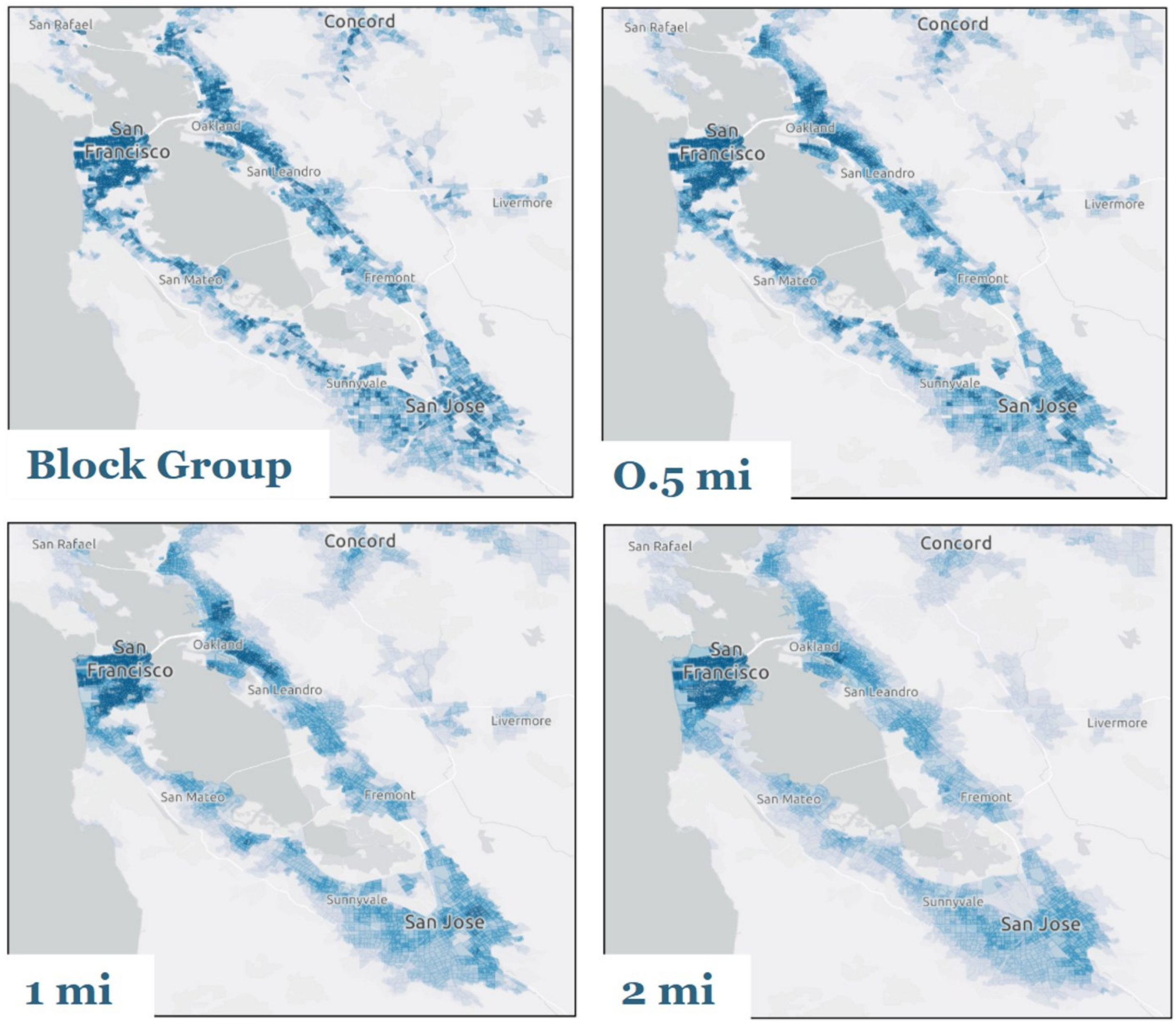
Figure 2.5: Dot density representation of survey respondents by CBSA (N=55,683)



Notes: Source is 2017 NHTS CA-add on survey. Universe includes observations with daily VMT equal to or less than 300 miles (N=55,683). Note that this is a depiction of individual survey respondents and not household locations.

To better understand how the aggregation of spatial densities functions in the context of our analysis, Figure 2.6 provides a visual representation of how measuring a single variable at multiple spatial scales can have varied effects**.** When variables are measured only at the block group level, there are greater differences between adjacent block groups. Successively larger scales of measurement effectively “smooth out” measurements in adjacent block groups, reducing heterogeneity among block groups that are close to one another or within the same region. In doing so, we can test to see the extent to which built environment characteristics matter at different scales. For example, having a high density of retail in block groups located within a mile of one's home block group may matter more than just what exists directly outside one's door.

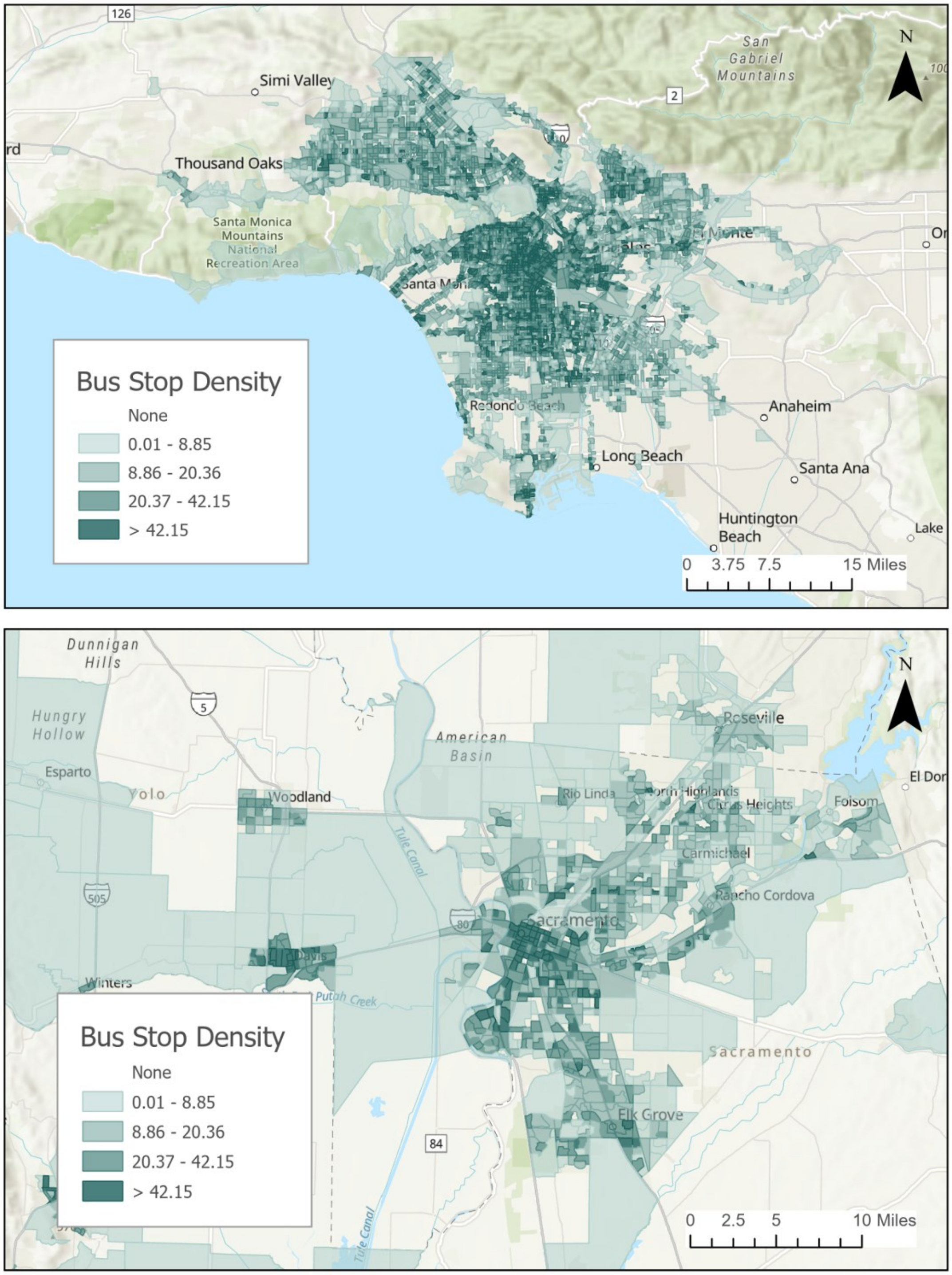
Figure 2.6: Population density at different spatial scales (Bay Area)



Notes: Source is 2017 Environmental Protection Agency Smart Location Database, using American Community Survey data. Note that these maps represent the homogenizing effect of increasing spatial scales of measurement on a qualitative level. They are not intended to present quantitative data and therefore no legend is included. Classification of colors is based on population density quintiles for the entire state, with darker colors indicating the highest density.

Figure 2.7 presents a visual representation of bus density across the Los Angeles and Sacramento regions by block group. Block groups that have a bus stop density of zero are left unshaded. As was noted previously, bus stop density was measured both by block group and by household.

Figure 2.7: Bus stop density (Los Angeles, Sacramento regions)



Notes: 2017 spatial data on statewide bus stops comes from multiple sources, including directly from transit agencies and regional planning agencies across the state. These data were collected and archived by the authors in 2017. See Appendix Section B.3: Cleaning bus stop data for more details.

Results

Base full sample model

The results of our initial diagnostic model can be interpreted by the coefficient associated with a given variable and understood as the variation in VMT associated with an increase of one standard deviation in that variable of interest. Dummy variables are simpler and can be interpreted as the VMT change associated with the presence of the attribute (e.g., rail stations within a quarter mile distance along the pedestrian network from home). Our model output in Table 2.7 shows that of all of the built environment characteristics considered, having rail access within a quarter mile and high network load density (2 mi radius) are the strongest and most significant correlates with VMT. Other significant and negatively associated built environment characteristics include jobs accessibility by car, the transit-to-car jobs accessibility ratio, street density (2 mi radius), bus stop density (1 mi radius, block group-based), and retail density in the home block group.

The CBSA dummy variable coefficients can be interpreted as the additional average daily VMT generated if an individual resides in one of these CBSAs, as opposed to living elsewhere in the state. Living in either the Inland Empire and Los Angeles regions appears to have a significant and strong positive relationship with VMT of 3.8 and 2.4 miles, respectively. Given that various features of the urban environment are controlled for in this model, it is likely that the CBSA effect seen here is due to other factors, which may cause more trips or longer trips, resulting in higher daily VMT. No other CBSAs tested here had a statistically significant relationship with daily VMT, which may be partly a function of sample size.

Table 2.7: Auto mileage regressed on built environment variables and demographic controls.

	Variable	Estimate
	Intercept	**-12.2648******
	Built environment variables	
	Population density (2 mi radius)	0.8129
	Retail density (block group)	**-0.4724*****
	Rail within 0.25 mi (DV)	**-3.6459***
	Bus stop density (1 mi radius, BG-based)	**-0.6509***
	Network load density (2 mi radius)	**-1.8895******
	Transit to car jobs accessibility ratio	**-0.8866*****
	Jobs accessibility by car	**-0.9754*****

	Variable	Estimate
	Street density (2 mi radius)	**-1.1500******
	CBSA	
	Riverside-San Bernardino-Ontario	**3.7575******
	Los Angeles-Long Beach-Anaheim	**2.4255****
	Sacramento-Roseville-Folsom	0.9761
	San Diego-Chula Vista-Carlsbad	0.6798
	San Francisco-Oakland-Fremont	-0.7047
	San Jose-Sunnyvale-Santa Clara	-0.0990
	Demographic Variables	
Age	15 and up	1.4274
(default: 5 and up)	**25 and up**	**6.9403******
	35 and up	-0.7478
	45 and up	0.5478
	55 and up	**-3.2395******
	65 and up	**-5.4039******
	75 and up	**-8.0396******
	Missing age	**-10.7989*****
Income	**$10,000 and up**	**3.1683****
(default: Less than	**$15,000 and up**	**3.6434*****
$10,000)	**$25,000 and up**	**2.5697****
	$35,000 and up	**2.4893*****
	$50,000 and up	-0.5201
	$75,000 and up	**3.0343******
	$100,000 and up	0.0122
	$125,000 and up	**1.4638***
	$150,000 and up	-1.4056
	$200,000 and up	1.0819
	Missing income	**7.9726******
Education	**High school or higher**	**6.9426******
(default: Less than	**Some college or higher**	**5.4236******
high school)	**Bachelor's degree or higher**	**1.5650*****
	Graduate degree or higher	**0.8410***
	Missing education	**-6.3749******
Sex	**Male**	**2.8822******
(default: Female)	Missing sex	3.5536
Race	Black	0.2425
(default: White)	Native	-0.1101
	Asian	**-2.6683******

	Variable	Estimate
	Pacific Islander	1.0841
	Multiple races	-0.0192
	Other race	0.3141
	Missing race	-3.0831
Ethnicity (default: non-Hisp)	**Hispanic**	**1.7227*** **
	Missing ethnicity	7.6438
Tenure (default: Own)	**Rent**	**-2.2635******
	Other	**-4.7288****
	Missing tenure	5.9263
Other	**Household size**	**-0.7825******
	Has young children (DV)	**-8.1824******
	N	**55,683**
	Log(scale)	**3.5589******

Notes: Bolded values represent variables with statistical significance. Stars indicate statistical significance: **** p<0.001, *** p<0.01, ** p<0.05, * p<0.10.

Interpreting demographic controls works somewhat differently than previously discussed variables. All categorical variables must be specified in reference to some default category, which is listed below the category title. Age, income, and education are ordinal variables that are specified as open-ended categories (e.g. if an observation belongs to the 45 and up group, they also belong to the 35 and up group) with the default as the lowest group within that category (e.g. the youngest age group, lowest income level, or least amount of educational attainment). When reading model outputs, this means that the coefficient for each group within a category can be thought of as the marginal VMT increase or decrease associated with each successive group relative to everyone younger than them, when controlling for all of the other demographic and built environment variables. For instance, the 6.94 coefficient associated with the 25 and up group means that individuals within this group typically drive 6.94 miles more than those in the groups younger than them (0-24 years), when controlling for all other variables. This makes sense given that this is the age by which people typically get their driver's license, own a car, and have a job to which they often will need to commute. The next age group for which we see a significant difference is 55 and up, at which point VMT generation begins to decrease with age. VMT continues to decrease at a higher rate with increased age. When looking at income, we see that those making more money typically generate more VMT, when controlling for all other variables. We see a similar phenomenon when looking at educational attainment, where the more education one receives, the higher the VMT they are likely to generate, all else equal. However, this effect is less strong with each successive level of educational attainment, as coefficients are still positive but smaller in magnitude.

For other categorical variables such as sex, race, age, and tenure we still set a default group but do not define open ended categories, given that these variables do not have an ordered structure. As can be seen in Table 2.7, when controlling for all of the other factors in the model, men on average drive 2.88 more miles per day than women, Asian people drive 2.67 miles less than White people, Hispanic people drive 1.72 miles more than non-Hispanic people, and renters drive 2.26 miles less than homeowners. We see other significant effects when investigating household makeup. For each standard deviation from the mean household size, we see a 0.78 mile decrease in daily VMT. Having a young child in the house is associated with one of the strongest and most significant decreases in daily VMT for the entire model.

Base full sample model with low income interactions

While the initial diagnostic model provides insight on how individuals travel by controlling for income, the next model gives a better understanding of how lower income individuals generate VMT relative to the rest of the population. Coefficients for built environment characteristics seen here can be interpreted as loss or gain in VMT for the general population, while the associated "low income" built environment variable coefficient can be interpreted as the marginal difference between the change in VMT for the general population associated with changes in the built environment variables, and the associated change for the low income population specifically.

As can be seen in
Table 2.8, the only built environment variable that appears to have a differential effect between the low- and higher-income groups is network load density (2 mi radius), for which low income individuals have an additional VMT decrease of 2.85 miles per day for every standard deviation increase in density. All other built environment variables included in the initial model do not have a statistically significant difference for low income individuals.

Table 2.8: Auto mileage regressed on built environment variables and demographic controls, with interaction terms for observations with household income less than $50,000

	Variable	Estimate
	Intercept	**-12.1828******
	Built environment variables	
	Population density (2 mi radius)	0.5197

	Variable	Estimate
	Low income (<$50,000)	0.9194
	Retail density (block group)	**-0.3880****
	Low income (<$50,000)	-0.3544
	Rail within 0.25 mi (DV)	**-4.5425***
	Low income (<$50,000)	2.6796
	Bus stop density (1 mi radius, BG-based)	**-0.6920***
	Low income (<$50,000)	-0.0628
	Network load density (2 mi radius)	**-1.2266****
	Low income (<$50,000)	**-2.8534*****
	Transit to car jobs accessibility ratio	**-0.7606***
	Low income (<$50,000)	-0.3493
	Jobs accessibility by car	**-1.1500*****
	Low income (<$50,000)	0.7420
	Street density (2 mi radius)	**-1.3124******
	Low income (<$50,000)	0.8560
	CBSA	
	Riverside-San Bernardino-Ontario	**3.7767******
	Los Angeles-Long Beach-Anaheim	**2.3955****
	Sacramento-Roseville-Folsom	0.8774
	San Diego-Chula Vista-Carlsbad	0.6686
	San Francisco-Oakland-Fremont	-0.7233
	San Jose-Sunnyvale-Santa Clara	-0.0770
	Demographic Variables	
Age (default: 5 and up)	15 and up	1.4680
	25 and up	**6.8684******
	35 and up	-0.7371
	45 and up	**0.5690**
	55 and up	**-3.2252******
	65 and up	**-5.4257******
	75 and up	**-8.0197******
	Missing age	**-10.8161*****
Income (default: Less than $10,000)	**$10,000 and up**	**3.1871****
	$15,000 and up	**3.5809*****
	$25,000 and up	**2.5473****
	$35,000 and up	**2.4760*****
	$50,000 and up	-0.5808
	$75,000 and up	**3.0289******
	$100,000 and up	0.0305

	Variable	Estimate
	$125,000 and up	**1.4664***
	$150,000 and up	-1.4209
	$200,000 and up	1.0651
	Missing income	**7.8211********
Education (default: Less than high school)	**High school or higher**	**6.9245********
	Some college or higher	**5.4450********
	Bachelor's degree or higher	**1.5628*******
	Graduate degree or higher	**0.8430***
	Missing education	**-6.3708********
Sex (default: Female)	**Male**	**2.8810********
	Missing sex	3.6345
Race (default: White)	Black	0.2176
	Native	-0.0115
	Asian	**-2.6015********
	Pacific Islander	1.1047
	Multiple races	-0.0121
	Other race	0.3360
	Missing race	-3.0102
Ethnicity (default: non-Hisp)	**Hispanic**	**1.7673*******
	Missing ethnicity	7.3913
Tenure (default: Own)	**Rent**	**-2.2103********
	Other	**-4.7344*****
	Missing tenure	6.0986
Other	**Household size**	**-0.7787********
	Has young children (DV)	**-8.1637********
	N	**55,683**
	Log(scale)	**3.5587********

Notes: Bolded values represent variables with statistical significance. Stars indicate statistical significance: **** p<0.001, *** p<0.01, ** p<0.05, * p<0.10.

We can more easily compare model outputs for our main built environment variables of interest by looking at Figure 2.8 and Figure 2.9.

Figure 2.8: Base full sample model outputs for built environment variables of significance

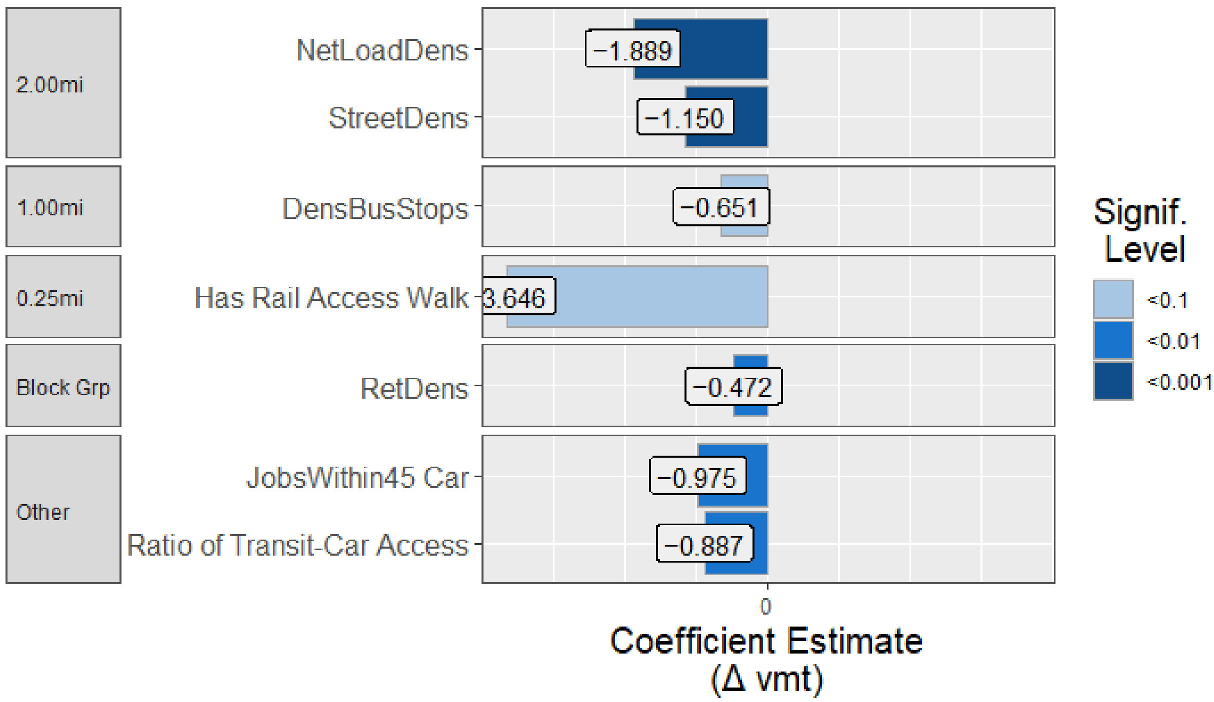
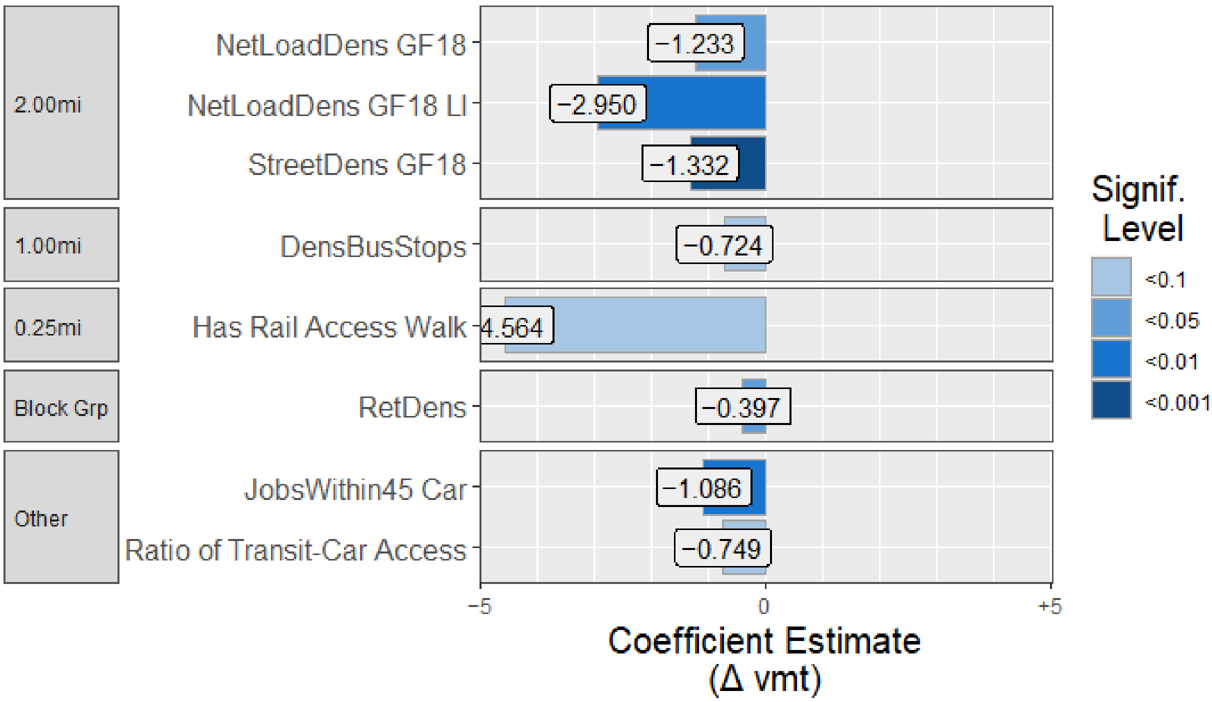


Figure 2.8 above provides a visual representation of the strength and significance of the statistically significant built environment variables on daily VMT Here we can see that while having rail access within a quarter mile has the largest negative coefficient, network load density has the most statistically significant impact on VMT generation out of all included built environment variables. While rail proximity may be influential, it may also be a proxy for off- and on-street parking availability (Chatman 2013), a measure we were unable to include in this model due to lack of data.

Figure 2.9: Base full sample model with low income interaction outputs for built environment variables of significance



Notes: "LI" indicates that the variable is the low income interaction version of the original variable.

Task 3: Investigate VMT as a function of housing unit size

Key Research Question & Basic Theory

Does project density have an impact on the VMT per capita of its inhabitants? More simply, do larger units lead to less per-capita driving for their inhabitants?

Theoretically, different living arrangements with multiple people in a unit could lead to lower need for driving, on a per capita basis. Such living arrangements could be a larger household, or a multigenerational household, or multiple unrelated persons living together. Each of these households would likely need a larger unit than smaller households, either in terms of unit size or room count. In terms of reduced demand for car use, perhaps carpooling among inhabitants, shared shopping, or more efficient route planning would lead to less car use or owning fewer cars.

Only a few papers have approached testing parts of this hypothesis. Wilbur Smith Associates, in a report for the city of San Diego in 2011, found that affordable housing residents tend to have more cars if they live in units with more bedrooms: 0.2 vehicle on average in a studio apartment, 0.5 in a one-bedroom, 1.25 in a two-bedroom, and 1.55 in a three-bedroom (Wilbur Smith 2011). They also found that there are more drivers per unit, and higher vehicle availability. They note that in senior housing developments, vehicle availability is lower, but so is the average number of bedrooms per property.

Newmark and Haas (2016) look at travel behavior by income and transit accessibility. They find that lower income families have fewer rooms per unit in transit rich areas than middle income families, for example. They conclude that “the building space measure is most pertinent to the consideration of the allocation of scarce location-efficient land. Lower-income groups simply use location-efficient space more efficiently. This efficiency extends the VMT reduction benefits of any parcel to more households” (Newmark & Haas 2016, p.20). They derive model predictions that suggest that developing parcels for lower-income households results in greater VMT reductions than a development on that parcel for higher-income households. Their results, however, rely on the unstated assumption that if given the option, lower-income households would sort into smaller units with fewer rooms. Yet, in a subsidized unit scenario, that is not necessarily the outcome. Moreover, program rules may dictate a certain occupancy or number or rooms per person ratio. Additionally, their comparison is low-income households in smaller units versus higher-income households in larger units, which is not necessarily apples to apples.

We have not located additional studies that examine either unit size or room count and its effect on VMT or travel behavior more broadly. The remainder of this task looks at empirically measuring these relationships.

Materials & Methodology

Measuring Travel and Unit Size

A primary limitation of this task is the dearth of data sources. Typical sources that measure travel variables do not include unit size or bedroom count, including various travel surveys or mobile data providers. Similarly, data sources focused on housing unit size do not generally concern themselves with travel.

We have identified one reasonable reliable source of data for this task: the American Housing Survey (AHS). The AHS is a national survey which asks unit inhabitants a series of questions about their lives and their homes. Unit size and bedroom count are included in every iteration of the survey (usually every two years). Travel variables are included only occasionally. We scanned every survey year since 2005 and only 2009 and 2017 had relevant variables. Both vintages had some measure of commute distance available (round-trip commute distance in 2017, single-direction commute distance in 2009). Note that commutes are just one portion of VMT, accounting for maybe fewer than half of all trips and miles.

For analysis, we looked at descriptive statistics of commute distance by bedroom count and unit size, as well as by tenure. We also estimated regression models of commute distance in terms of unit size, and other household and unit level variables. We estimated additional regressions looking at mode choice using the 2017 vintage as well.

Descriptive Statistics

Our first analysis looked at average commute distances by home size. We split up rental units and owner-occupied units when we noted a statistically significant difference in commute distance between occupants of each unit type. Rental unit occupants had a mean distance of 24.2 miles, while owned unit occupants had a mean distance of 29.6 miles. Bedroom count and unit size distributions also varied between owned and rented units, with studios, and 1- and 2- bedroom units accounting for ¾ of all rental units, but only about 1/3 of owned units. The unit size splits were also quite different by tenure.

Descriptive statistics suggest that commute distances increase as the number of bedrooms per unit increases (Figure 3.1). The increase is higher for owners than for renters, but the trend is similar for both. The AHS data shown here is the only available source of bedroom count-specific on travel activity. No VMT or travel data for AHSC residents is available so direct comparison is not possible. When looking at square footage, we see a similar trend for renters: commute distance increases with unit size (Figure 3.2). The pattern for owner-occupied units is less clear but the overall trend is still in line with renters. This could potentially be correlated with larger homes in certain suburban areas, which also tend to have longer commutes. Similarly, AHSC resident data on travel by apartment square footage is not available. These descriptive charts suggest possible correlations but do not indicate any causal effect.

Recall that the key outcome variable – round trip commute distance – is a poor proxy for VMT; but the only one available to directly measure unit size. Moreover, because the AHS only measures the commute and not other trips and only considers the head of household's commute but not the trips for the whole household, the measure is even less effective.

Figure 3.1: Mean Commute Distance by Bedroom Count (AHS 2017)

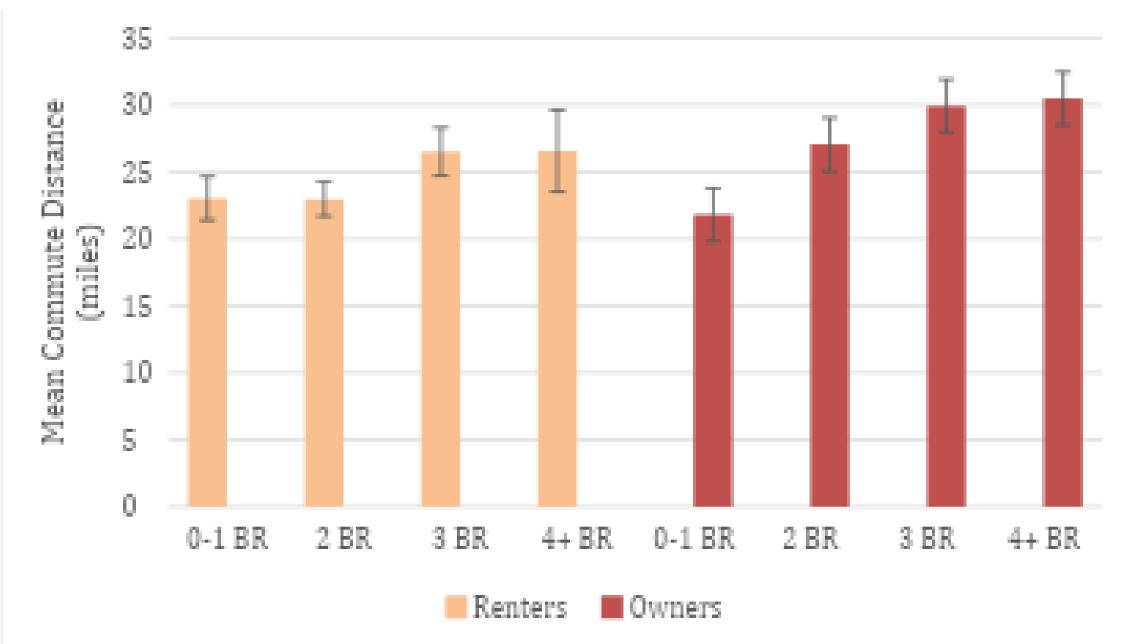
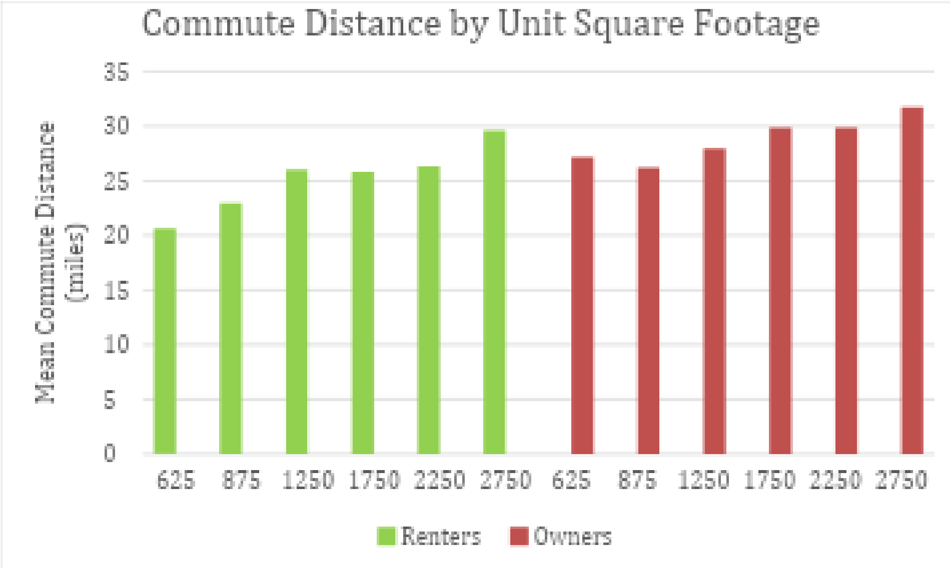


Figure 3.2: Mean Commute Distance by Unit Square Footage (AHS 2017)



Results

To better isolate the impact of home size on commute distance, we estimate regressions which control for home size, tenure, and household and unit control variables that may also affect commute distance. The dependent variable is round trip commute distance. Key variables of interest include bedroom count, household size, and their interaction as well as unit size, and unit size and household size interaction. Additional control variables include tenure, the number of bathrooms in the unit, total monthly housing costs (includes mortgage, rent, utilities, property taxes, insurance, HOA, lot fees, and maintenance), the number of elderly in the household, the number of children in the household, the head of household's age, total household income, whether any household members have a disability, whether head of household has a bachelor's degree, whether head of household is Black, Asian, Hispanic, or Other Race, whether head of household is female, days per week the commute is made, and total cost of the commute. Table 3.1 and Table 3.2 show regression results for variables of interest. We also ran specifications which control for commute mode (subway, bus, bike, walk, and carpool). The coefficient signs, statistical significance, and magnitude were nearly the same as the results shown below; those tables have been omitted for brevity.

The models are estimated using weighted least squares (WLS) in Stata using the svy functionality. We use household weights provided in the AHS 2017.

The non-interacted model of bedroom count (Model 1 in Table 3.1) shows that those living in 3 or 4 bedroom units tend to have 1.9 and 1.3 mile longer round trip commutes than those living in studios or 1 bedroom units, holding everything else equal. The interacted model attempts to more directly isolate the impact of bedroom count through household size on commute distance (Model 2 in Table 3.1). Nearly every interacted variable is not statistically significantly different from zero. This suggests that round trip commute distances largely even out by bedroom count when household size is accounted for.

Results for the square footage models are broadly similar. The non-interacted model (Table 3.2 Model 3) suggests that households living in 1500 – 2500 square foot units have a commute distance about 1.5 miles more round trip than those living in units with less than 625 square feet. The interacted model (Table 3.2 Model 4) similarly shows almost no statistically significant interactions between unit size and household size, again suggesting that the control variables and household size explain away most of the variation in commute distance by unit size.

Table 3.1: Regressions of Bedroom Count Impact on Commute Distance

Variables	Model 1	Model 2
Bedrooms (base = 1)		
2	0.17	-0.878
3	1.869***	1.95***
4	1.287*	1.026
Household Size (base = 1)		
2	0.486	-0.426
3	0.886*	-0.326
4	0.065	1.181
5	0.209	-9.415
Bedrooms x Household Size		
2 BR, 2 People		2.21*
2 BR, 3 People		2.329
2 BR, 4 People		-0.203
2 BR, 5 People		8.793
3 BR, 2 People		0.199
3 BR, 3 People		0.592
3 BR, 4 People		-1.103

Variables	Model 1	Model 2
3 BR, 5 People		8.961
4 BR, 2 People		0.925
4 BR, 3 People		1.259
4 BR, 4 People		-1.812
4 BR, 5 People		10.052
Controls	**yes**	**yes**
Sample Size	**10,854**	**10,854**
R2	**0.73**	**0.73**

Notes: Table displays regression model coefficients, stars indicate statistical significance: *** p<0.01, ** p<0.05, * p<0.10

Table 3.2: Regressions of Unit Size Impact on Commute Distance

Variables	Model 3	Model 4
Unit Size (base is <= 625 sq ft)		
625 - 1000 sq ft	0.154	0.071
1000 - 1500 sq ft	0.612	0.512
1500 - 2000 sq ft	1.494**	1.991*
2000 - 2500 sq ft	1.49**	3.501**
>2500 sq ft	1.285	2.605
Household Size (base = 1)		
2	0.575	1.293
3	1.091**	1.921
4	0.273	2.139
5	0.358	-6.367
Bedrooms x Household Size		
625 - 1000 sq ft, 2 people		-0.401
625 - 1000 sq ft, 3 People		0
625 - 1000 sq ft, 4 People		-0.401
625 - 1000 sq ft, 5 People		-0.471*
1000 - 1500 sq ft, 2 People		-1.033
1000 - 1500 sq ft, 3 People		9.584
1000 - 1500 sq ft, 4 People		-0.391
1000 - 1500 sq ft, 5 People		0.152
1500 - 2000 sq ft, 2 People		-1.303

Variables	Model 3	Model 4
1500 - 2000 sq ft, 3 People		6.823
1500 - 2000 sq ft, 4 People		-0.419
1500 - 2000 sq ft, 5 People		-2.234
2000 - 2500 sq ft, 2 People		-2.122
2000 - 2500 sq ft, 3 People		6.75
2000 - 2500 sq ft, 4 People		-2.569
2000 - 2500 sq ft, 5 People		-3.341
>2500 sq ft, 2 People		-3.317
>2500 sq ft, 3 People		4.207
>2500 sq ft, 4 People		-2.663
>2500 sq ft, 5 People		-0.630
Controls	**yes**	**yes**
Sample Size	**10,854**	**10,854**
R2	**0.72**	**0.72**

Notes: Table displays regression model coefficients. Stars indicate statistical significance: *** p<0.01, ** p<0.05, * p<0.10.

Our analyses have found that commute distance increases with unit size, contrary to the initial hypothesis. However, our outcome variable is far from perfect. It does not cover all trips, nor does it cover all households.

To have a stronger measure, instruments like travel surveys would need to have more robust questions about housing units. Conversely, housing surveys and administrative data would need to have better understanding about inhabitants' travel behavior.

One suggestion would be to administer short travel surveys or data collection of AHSC residents to more clearly understand travel behavior patterns among different unit types and sizes, with additional information about inhabitants' such as household size, income, age, access to a car, and other travel behavior predictors.

Task 4: VMT estimates for senior housing

Key Research Question & Basic Theory

The current methodology used by CARB to estimate the trip generation rates for senior housing relies on the Institute of Transportation Engineers (ITE) Trip Generation Manual to calculate the unmitigated VMT of different dwelling types. The ITE Trip Generation Manual relies on national survey data that may not fully reflect senior trip rates in the U.S. as a whole, or of course for California as a subset of the US–particularly for senior buildings, since it is based on a nationwide sample of six.

In Task 4 we carried out an independent investigation (using the NHTS data that were described in the Task 2 writeup above) of whether seniors as a group, or low-income seniors specifically, respond differently in terms of their use of autos in response to variations in the built environment and transit access. We describe the results below, but in summary, we did not find evidence to suggest that senior income-restricted housing should be treated differently from other kinds of low-income housing projects in estimating VMT reductions associated with AHSC projects.

Materials & Methodology

The same dataset from Task 2 was used here in Task 4. We analyzed the data differentiating between respondents under the age of 65 and respondents 65 or older, who comprised 15,177 respondents (27% of the state sample). For a detailed explanation of how the survey, built environment, and transit data were collected and processed, please see the Task 2 writeup above.

Statistical models

As seen in the results presented for Task 2, those aged over 65 drive less in comparison to the general population when controlling for other factors. But this does not answer the question of whether and how the built environment and transit proximity influence the VMT level of seniors (or of low-income seniors) differently than the rest of the population (or differently than other low-income individuals). We describe the analysis we conducted to answer that question below.

We began by applying the initial base model as specified in Task 2 to the subset of the NHTS sample with observations 65 years or older (N=15,177). The reduced sample sizes created challenges with finding statistically significant coefficients, and therefore we chose to specify another model with interaction terms for the over 65 population.

(There are some slight differences in the model specifications (e.g. these models include a dummy variable for half mile rail proximity, while the full sample model does not) though model results are still interpretable relative to one another.) To further investigate how affordable senior housing may have an impact in comparison to non-age-restricted housing, we subsetted the sample to the low income population (defined as observations with household income below $50,000, N=18,398) and added interaction terms to the built environment and transit variables for the over-65 population.

We began by investigating travel behavior of the entire senior population. For our first approach, we specified a diagnostic model based only on the subset of the data for which the respondent was age 65 or older (N = 15,177). In doing so we are able to initially identify the built environment characteristics that are important in determining VMT for seniors. For this model we used robust standard errors, but were unable due to small sample size to cluster errors on block groups, which has the effect of overestimating statistical significance. While this model can provide preliminary insights, the small sample size also leads to less statistical significance even before accounting for this overestimation. Only using the over 65 population subset in the model means that we cannot directly test whether the nominally different coefficients are statistically distinguishable from those in the population as a whole. The model is specified as follows:

Equation 4.1: Task 4 base over 65 model specification

$$VMT_i = \beta_1 x_i^1 + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 + \beta_6 x_i^6 + \beta_7 x_i^7 + \beta_8 x_i^8 + \beta_9 x_i^9 + \beta_{10} C_i + \beta_{11} D_i + \varepsilon_i$$

Table 4.1: Variable definitions for Equation 4.1: Task 4 base over 65 model specification

Variable	Variable definition	Units
VMT_i	Daily vehicle miles traveled of the ith observation in the sample.	miles/day
β_n	The coefficient associated with the nth built environment characteristic for n= 1, 2, ..., 9. n=1: Network load density (2 mi radius) n=2: Street density (2 mi radius) n=3: Bus stop density (BG based, 1 mi) n=4: Population density (2 mi radius)	Miles/day per standard deviation units; see Table 2.5.

Variable	Variable definition	Units
	n=5: Retail density (block group) n=6: Has rail access within 0.25 mi n=7: Has rail access within 0.5 mi n=8: Ratio of transit:car job accessibility n=9: Jobs accessibility by car	
x_i^n	The z score of the nth built environment variable, for the ith observation	Standard deviation units; see Table 2.5. For raw units by variable, see Table 2.1.
β_{10}	A vector of coefficients associated with the CBSA of residence.	NA
C_i	A vector of dummy variables indicating if the ith observation is in a CBSA. The sum of values in the vector is no greater than 1, as an observation can only reside in one CBSA.	NA
β_{11}	A vector of coefficients associated with the demographic control variables.	NA
D_i	The vector of demographic control variables.	NA
ε_i	The error term.	Miles/day

Due to the challenges the first model posed with statistical significance, we estimated a second model using all observations, regardless of age, and including variables interacting over-65 age status with each of the built environment characteristics included in the models reported in the Task 2 writeup. (Interaction terms were set to be their original value if the given respondent fell into the senior group and set to zero if the respondent was not in the senior population.) With this model, we are able to specifically test to see if coefficients differ between the groups and see the extent to which certain variables are statistically significant in comparison to the rest of the population. Note that we only included interaction terms for statistically significant built environment variables This specification is specified as follows:

Equation 4.2: Task 4 base full sample with senior interactions

$$VMT_i = \beta_1 x_i^1 + \beta_1^* x_i^{1*} + \beta_2 x_i^2 + \beta_2^* x_i^{2*} + \beta_3 x_i^3 + \beta_3^* x_i^{3*} + \beta_4 x_i^4 + \beta_4^* x_i^{4*} + \beta_5 x_i^5 + \beta_5^* x_i^{5*} + \beta_6 x_i^6 + \beta_6^* x_i^{6*}$$
$$\beta_7 x_i^7 + \beta_7^* x_i^{7*} + \beta_8 x_i^8 + \beta_8^* x_i^{8*} + \beta_9 x_i^9 + \beta_9^* x_i^{9*} + \beta_{10} C_i + \beta_{11} D_i + \varepsilon_i$$

Table 4.2: Variable definitions for Equation 4.2: Task 4 base full sample model with senior interactions

Variable	Variable definition	Units
VMT_i	Daily vehicle miles traveled of the ith observation in the sample.	miles/day
β_n	The coefficient associated with the nth built environment characteristic for n= 1, 2, ..., 9. n=1: Network load density (2 mi radius) n=2: Street density (2 mi radius) n=3: Bus stop density (BG based, 1 mi) n=4: Population density (2 mi radius) n=5: Retail density (block group) n=6: Has rail access within 0.25 mi n=7: Has rail access within 0.5 mi n=8: Ratio of transit:car job accessibility n=9: Jobs accessibility by car	Miles/day per standard deviation units; see Table 2.5.
x_i^n	The z score of the nth built environment variable, for the ith observation	Standard deviation units; see Table 2.5. For raw units by variable, see Table 2.1.
β_n^*	The interaction coefficient associated with the nth built environment characteristic for $n = 1, 2, \ldots, 9$.	Miles/day per standard deviation units; see

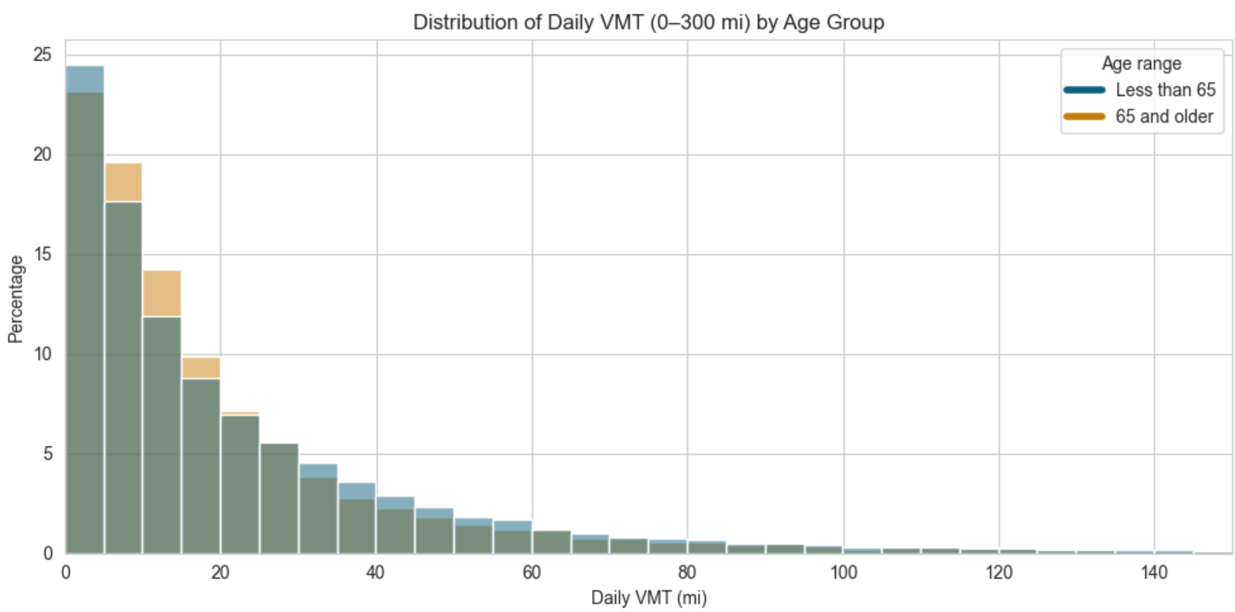
Variable	Variable definition	Units
		Table 2.5.
x_i^{n*}	The nth interactive built environment variable, for the ith observation. The variable is defined as follows: • $x_i^{n*} = x_i^n$ if the ith observation is age 65 or older. • $x_i^{n*} = 0$ otherwise.	Standard deviation units; see Table 2.5. For raw units by variable, see Table 2.1.
β_{10}	A vector of coefficients associated with the CBSA of residence.	NA
C_i	A vector of dummy variables indicating if the ith observation is in a CBSA. The sum of values in the vector is no greater than 1, as an observation can only reside in one CBSA.	NA
β_{11}	A vector of coefficients associated with the demographic control variables.	NA
D_i	The vector of demographic control variables.	NA For raw units by variable, see Table 2.2.
ε_i	The error term.	NA

In addition to including an over-65 interaction term for the full sample model, we specified a similar model for the subset of the sample with an annual household income of less than $50,000 (N = 18,398). The only difference in the way the first and second Task 4 interaction models are specified (other than the estimation sample) is that the low income model did not include the 0.5 mile rail dummy variable, more closely reflecting the model specifications from Task 2. Both of these interaction models used robust standard errors, clustered by block group.

Descriptive statistics

As is supported by the literature and seen here in Figure 4.1, people aged 65 years and older drive slightly less on average than those younger than 65. Not only do senior survey respondents drive less, but there is also a larger proportion of senior survey respondents who do not drive at all on the given survey day; 31.7% of seniors did not drive, compared to 27.4% of non-seniors. However, when we look at the distribution of daily VMT by age at a more granular level, a more subtle pattern appears. Overall there is a greater share of seniors who traveled 0-20 miles on the survey day when compared to non-seniors, and non-seniors were more likely to travel more than 40 miles. The one exception to this general trend is the greater relative share of non-seniors who traveled on the survey day, but no more than 5 miles, though we do not explore why here.

Figure 4.1: Daily VMT distribution and mean for those younger than 65 vs. 65 and older



Notes: Source is the 2017 NHTS CA-add on survey. The total sample with daily VMT in the range (0,150] is 28,934 for those under the age of 65 and 10,222 for those 65 and older. The number of entries with no VMT logged on the travel day is 11,079 for those under the age of 65 and 4,805 for those 65 and older. The median VMT was 6.8 miles/day for those under 65 and 5.9 miles/day for those 65 and older, though this difference is not statistically significant. Each bin on the chart represents an interval of 5 miles/day. Bars are layered so that green represents where bars overlap.

The general built environment characteristics near which seniors reside also differ somewhat from the general population. Comparing
Table 4.3 with descriptive results from

Table 2.5 in Task 2, we can see that seniors generally live in less dense areas with lower relative transit access and infrastructure density. For certain variables we see more significant differences between seniors and the general population than others. For instance, the median transit to car jobs accessibility ratio for seniors is one fifth that of the entire population. The median bus stop density for seniors is less than half of the median for the total sample. Senior median retail density is just over half that of the total sample. Other variables such as network load density and street density show less marked differences between both samples, and while senior median values are still lower, they are closer to that of the general population.

Table 4.3: Descriptive statistics for all built environment variables in Task 4 (N = 15,177)

Variable	Mean	Std dev	Q1	Median	Q3
Population Density (2 mi radius)	3,657.15	3,899.09	723.66	2,865.24	5,310.68
Retail Density (block group)	154.50	539.67	0	6.33	85.85
Rail within 0.25 mi (DV)	0.01	0.09	0	0	0
Rail within 0.5 mi (DV)	0.02	0.14	0	0	0
Bus stop density (BG-based, 1 mi radius)	8.84	17.48	0	1.72	11.24
Network load density (2 mi radius)	444.58	446.28	175.96	390.55	573.49
Transit to car jobs accessibility ratio	0.41	0.72	0	0.03	0.63
Jobs accessibility by car	62,090.59	89,173.79	6,399	23,235	85,512
Street density (2 mi radius)	9.36	4.98	5.24	9.93	13.38

Notes: Dataset restricted to respondents 65 or older with VMT less than 300 on the survey day. Source for population density, retail density, and jobs accessibility by car is the 2017 Environmental Protection Agency Smart Location Database. OpenStreetMap data were used for street density and accessed using the OSMnx Python package. Rail station and bus stop data were collected from transit and planning agencies across the state by the authors for previous research. Network load density and the transit-to-car jobs accessibility ratio are interaction terms that were computed from other variables.

As shown in
Table 4.4, the senior sample is generally more likely to be lower income, own their home, and live in a smaller household. The average income for seniors is $73,883, compared to $91,649 for the general population (using midpoints of income ranges

reported in the NHTS data). The senior sample is also more likely to be White and non-Hispanic.

Table 4.4: Descriptive statistics for demographic variables used in Task 4 (N = 15,177) compared to the full NHTS sample (N = 55,683)

	Sr. (%)	Total (%)		Sr. (%)	Total (%)
Income			**Race**		
Less than $10,000	3	3	White	85	74
$10,000 to 14,999	4	3	Asian	6	10
$15,000 to 24,999	9	7	Black or African American	2	3
$25,000 to 34,999	10	7	Multiple responses selected	2	6
$35,000 to 49,999	12	9	Some other race	2	5
$50,000 to 74,999	18	15	American Indian or Alaska Native	1	1
$75,000 to 99,999	13	14	Missing data	1	1
$100,000 to 124,999	10	12	Native Hawaiian or Pacific Islander	0	1
$125,000 to 149,999	5	7	**Ethnicity**		
$150,000 to 199,999	5	8	Non-Hispanic	93	85
$200,000 or more	5	10	Hispanic	7	15
Missing Data	5	3	Missing data	0	0
Education			**Age**		
Less than high school	4	6	65-74	62	NA
Completed High School	17	13	75 or older	38	NA
Associates Degree	33	27	**Sex**		
Bachelor's Degree	21	21	Female	53	52
Graduate Degree	25	20	Male	47	48
Missing Data	0	12	Missing Data	0	0
Tenure				**Sr. avg.**	**Total avg.**
Own	85	72	Household size	1.92	2.80
Rent	14	27			
Other	1	1			
Missing data	0	0			

Notes: Statistics represented here do not include observations with daily VMT greater than 300. Percentages showing 0 represent small non-zero percentages that have been rounded down. Due to rounding, some categories may not sum to 100%. Though there are age distributions for the total population, we do not present them here as they are not comparable. Source is 2017 NHTS CA-add on survey.

Results

Base over-65 model

The smaller sample size used in the subset model leads to less statistically significant results, especially for the built environment variables (Table 4.5). Only network load density (2 mi radius) and street density (2 mi radius) show up as statistically significant, at 1.66 and 1.32 mile daily decrease per standard deviation increase. No other built environment variables, including any variables relating to transit proximity or access, show up as significant in this model.

Table 4.5: Auto mileage regressed on built environment and demographic control variables for senior subset of the population (observations with age 65 years or older)

	Variable	Estimate
	Intercept	**-3.9618** ****
	Built Environment Variables	
	Population Density (2 mi radius)	0.7809
	Retail Density (block group)	0.2735
	Rail within 0.25 mi (DV)	0.7861
	Rail within 0.5 mi (DV)	-1.4623
	Bus stop density (1 mi radius, BG based)	-0.7430
	Network load density (2 mi radius)	**-1.6582** **
	Transit to car jobs accessibility ratio	-0.4045
	Jobs accessibility by car	0.2077
	Street density (2 mi radius)	**-1.3194** **
	CBSA	
	Riverside-San Bernardino-Ontario	2.6385
	Los Angeles-Long Beach-Anaheim	-1.3487
	Sacramento-Roseville-Folsom	0.4825
	San Diego-Chula Vista-Carlsbad	0.2500
	San Francisco-Oakland-Fremont	-0.3487
	San Jose-Sunnyvale-Santa Clara	-2.3833
	Demographic Variables	

	Variable	Estimate
Age (default: 65 and up)	**75 and up**	**-8.4714** ****
Income	$10,000 and up	1.9782
	$15,000 and up	**4.0789** **
(default: > $10,000)	**$25,000 and up**	**2.4400** *
	$35,000 and up	2.0538
	$50,000 and up	0.6829
	$75,000 and up	**3.9836** ****
	$100,000 and up	-2.0358
	$125,000 and up	**3.9731** **
	$150,000 and up	**-4.0426** *
	$200,000 and up	2.3069
	Missing income	**7.1889** ***
Education	**High school or higher**	**5.6957** ***
(default: Less than high school)	**Some college or higher**	**5.0074** ****
	Bachelor's degree or higher	0.9193
	Graduate degree or higher	**2.4261** ***
	Missing education	5.3277
Sex	**Male**	**4.7788** ****
(default: Female)	**Missing sex**	**62.4918** ****
Race	Black	-0.6921
(default: White)	Native	0.2105
	Asian	**-4.2611** ****
	Pacific Islander	2.2390
	Multiple races	0.5488
	Other race	**-5.2409** **
	Missing race	-1.8397
Ethnicity	Hispanic	1.6095
(default: non-Hisp)	Missing ethnicity	-6.9436
Tenure	**Rent**	**-5.6620** ****
(default: Own)	**Other**	**-4.9931** *
Other	**Household size**	**-4.1810** ****
	Has young children (DV)	2.1697
	N	**15,177**
	Log(scale)	**3.5993** ****

Notes: Bolded values represent variables with statistical significance. Stars indicate statistical significance: **** p<0.001, *** p<0.01, ** p<0.05, * p<0.10

Compared to Task 2, there are also fewer statistically significant demographic control variables. Being 75 years or older still shows up as having a significant and strong negative impact on daily VMT, relative to the population aged 65-74. While increased income is slightly positively associated with VMT, this relationship reverses when incomes are higher than $150,000 (though this relationship is not as significant). There is a significant and positive correlation of increased educational attainment for seniors with VMT, similar to results for the entire population in Task 2. There are also similar effects of sex, race, and tenure on VMT as was seen in Task 2. However, we see no significant correlation of ethnicity with VMT.

Base full sample model with senior interactions

We can observe the differential effects of the built environment on senior VMT relative to the general population by including interaction terms, as seen in Table 4.6**.** Note that this model does not look at the low income population, but instead the whole population. Coefficients for built environment characteristics in Table 4.4 can be interpreted as correlations with VMT for the general population, while the associated "Over 65" coefficient can be interpreted as the marginal difference for seniors compared to the rest of the population. For instance, while a one standard deviation increase in retail density from the mean results in 0.64 miles fewer daily VMT for the general population, seniors with the same retail density would generate about a half mile *greater* VMT (-0.65 + 1.14), although this result is of low statistical significance.

Table 4.6: Auto mileage regressed on built environment and demographic control variables with interaction terms for seniors (observations with age 65 years or older).

	Variable	Estimate
	Intercept	**-11.6025** ****
	Built Environment Variables	
	Population Density (2 mi radius)	0.8603
	Over 65 -	0.2405
	Retail Density (block group)	**-0.6447** ***
	Over 65 -	**1.1440** **
	Rail within 0.25 mi (DV)	**-5.694** **
	Over 65 -	6.1902
	Rail within 0.5 mi (DV)	**2.8861** *
	Over 65 -	**-6.2993** **

	Variable	Estimate
	Bus stop density (1 mi radius, BG-based)	**-1.1297** ***
	Over 65 -	**2.0525** ***
	Network load density (2 mi radius)	**-2.4948** ****
	Over 65 -	0.7454
	Transit to car jobs accessibility ratio	-0.0705
	Over 65 -	**-3.4576** ****
	Jobs accessibility by car	**-0.7837** **
	Over 65 -	-0.1483
	Street density (2 mi radius)	**-0.7388** **
	Over 65 -	-0.2224
	CBSA	
	Riverside-San Bernardino-Ontario	**3.6667** ****
	Los Angeles-Long Beach-Anaheim	**2.1046** ***
	Sacramento-Roseville-Folsom	0.696
	San Diego-Chula Vista-Carlsbad	0.4608
	San Francisco-Oakland-Fremont	-1.1149
	San Jose-Sunnyvale-Santa Clara	-0.611
	Demographic Variables	
Age (default: 65 and up)	**75 and up**	**-11.1043** ****
Income (default: Less than $10,000)	**$10,000 and up**	**3.1281** **
	$15,000 and up	**3.7451** ***
	$25,000 and up	**2.5096** **
	$35,000 and up	**2.5334** ***
	$50,000 and up	-0.3924
	$75,000 and up	**3.1195** ****
	$100,000 and up	0.1929
	$125,000 and up	**1.6719** *
	$150,000 and up	-1.4446
	$200,000 and up	1.3440
	Missing income	**7.7651** ****
Education (default: Less than high school)	**High school or higher**	**9.5298** ****
	Some college or higher	**5.5867** ****
	Bachelor's degree or higher	**2.1571** ****
	Graduate degree or higher	0.5439

	Variable	Estimate
	Missing education	**-9.6279** ****
Sex	**Male**	**2.8437** ****
(default: Female)	Missing sex	-0.3086
Race	Black	0.0619
(default: White)	Native	0.3820
	Asian	**-2.3040** ****
	Pacific Islander	1.5197
	Multiple races	0.1452
	Other race	0.5726
	Missing race	-2.4517
Ethnicity	**Hispanic**	**2.0400** ****
(default: non-Hisp)	Missing ethnicity	7.3318
Tenure	**Rent**	**-1.2488** ***
(default: Own)	**Other**	**-4.3405** **
	Missing tenure	**8.8211** *
Other	**Household size**	**-0.5327** ***
	Has young children (DV)	**-7.5430** ****
	N	**55,683**
	Log(scale)	**3.5993** ****

Notes: Bolded values represent variables with statistical significance. Stars indicate statistical significance: **** p<0.001, *** p<0.01, ** p<0.05, * p<0.10.

Low income sample model with senior interactions

For our final model we explored how low income seniors travel relative to the rest of the low income population by subsetting the sample to the low income population and including interaction variables for individuals over age 65, as done in the previous model. As can be seen in Table 4.7, subsetting our sample population once again resulted in decreased statistical significance. The results below can be quite briefly summarized: we find no evidence here that the built environment and transit proximity affect the VMT of low-income seniors differently than the rest of the low-income population.

Table 4.7: Auto mileage regressed on built environment and demographic control variables, subset to the low income population (<$50,000) with interaction terms for seniors (65+)

	Variable	Estimate
	Intercept	**-12.1922**** **
	Built environment variables	
	Population density (2 mi radius)	0.6979
	Over 65 -	2.0209
	Retail density (block group)	-0.7022
	Over 65 -	0.1984
	Rail within 0.25 mi (DV)	0.1081
	Over 65 -	-6.2430
	Bus stop density (1 mi radius, BG-based)	-0.5674
	Over 65 -	-0.0453
	Network load density (2 mi radius)	**-3.6262********
	Over 65 -	-1.6426
	Transit to car jobs accessibility ratio	**-1.3082****
	Over 65 -	0.5712
	Jobs accessibility by car	0.1970
	Over 65 -	-1.1006
	Street density (2 mi radius)	-0.1953
	Over 65 -	-0.0741
	CBSA	
	Riverside-San Bernardino-Ontario	**3.4969****
	Los Angeles-Long Beach-Anaheim	1.8114
	Sacramento-Roseville-Folsom	-0.2072
	San Diego-Chula Vista-Carlsbad	1.1062
	San Francisco-Oakland-Fremont	0.7055
	San Jose-Sunnyvale-Santa Clara	0.3165
	Demographic Variables	
Age	**15 and up**	**6.3997****
(default: 5 and up)	**25 and up**	**7.4452********
	35 and up	**-3.7223******
	45 and up	0.6819
	55 and up	**-3.3494******
	65 and up	**-5.8535********
	75 and up	**-6.6527********
	Missing age	-11.1591
Income	**$10,000 and up**	**3.3669****

	Variable	Estimate
(default: Less than $10,000)	**$15,000 and up**	**3.3408*** **
	$25,000 and up	**2.3255** **
	$35,000 and up	**2.5575*** **
	Missing income	**7.7929**** **
Education (default: Less than high school)	**High school or higher**	**5.5585**** **
	Some college or higher	**5.9934**** **
	Bachelor's degree or higher	0.7670
	Graduate degree or higher	-0.0363
	Missing education	-2.1717
Sex (default: Female)	Male	0.7886
	Missing sex	5.7155
Race (default: White)	**Black**	**-3.2843** **
	Native	-1.2219
	Asian	-0.6088
	Pacific Islander	**6.0196* **
	Multiple races	-0.8607
	Other race	-1.9541
	Missing race	-0.2824
Ethnicity (default: non-Hisp)	**Hispanic**	**3.3506**** **
	Missing ethnicity	4.2157
Tenure (default: Own)	**Rent**	**-3.7572**** **
	Other	**-5.7138** **
	Missing tenure	0.9724
Other	**Household size**	**-0.9014*** **
	Has young children (DV)	**-6.9342**** **
	N	**18,398**
	Log(scale)	**3.5495**** **

Notes: Bolded values represent variables with statistical significance. Stars indicate statistical significance: **** p<0.001, *** p<0.01, ** p<0.05, * p<0.10.

Task 5: Sensitivity analysis

In this section, we show how the analysis approach identified in Task 2 results in estimated VMT variation by block group in California as a function of built environment characteristics. Doing so may help to assess regional differences in VMT estimates and identify any unintended consequences to applying this method for the purpose of estimating VMT reductions associated with AHSC projects.

To estimate VMT for every block group in the state, we used the model coefficients for all of our statistically significant built environment characteristics in Task 2. We then used our block group level measurements for each of these built environment characteristics and multiplied them by their coefficients, summing them together and adding them to the intercept to give an estimate of overall VMT generation for individuals in that block group. Given that these built environment characteristics had originally been scaled by the mean and standard deviation values for our sample of observations for inclusion in the regression model, we used the same values to scale the measurements for each of the block groups in the state (see

Table 2.5 for the exact values used to scale).

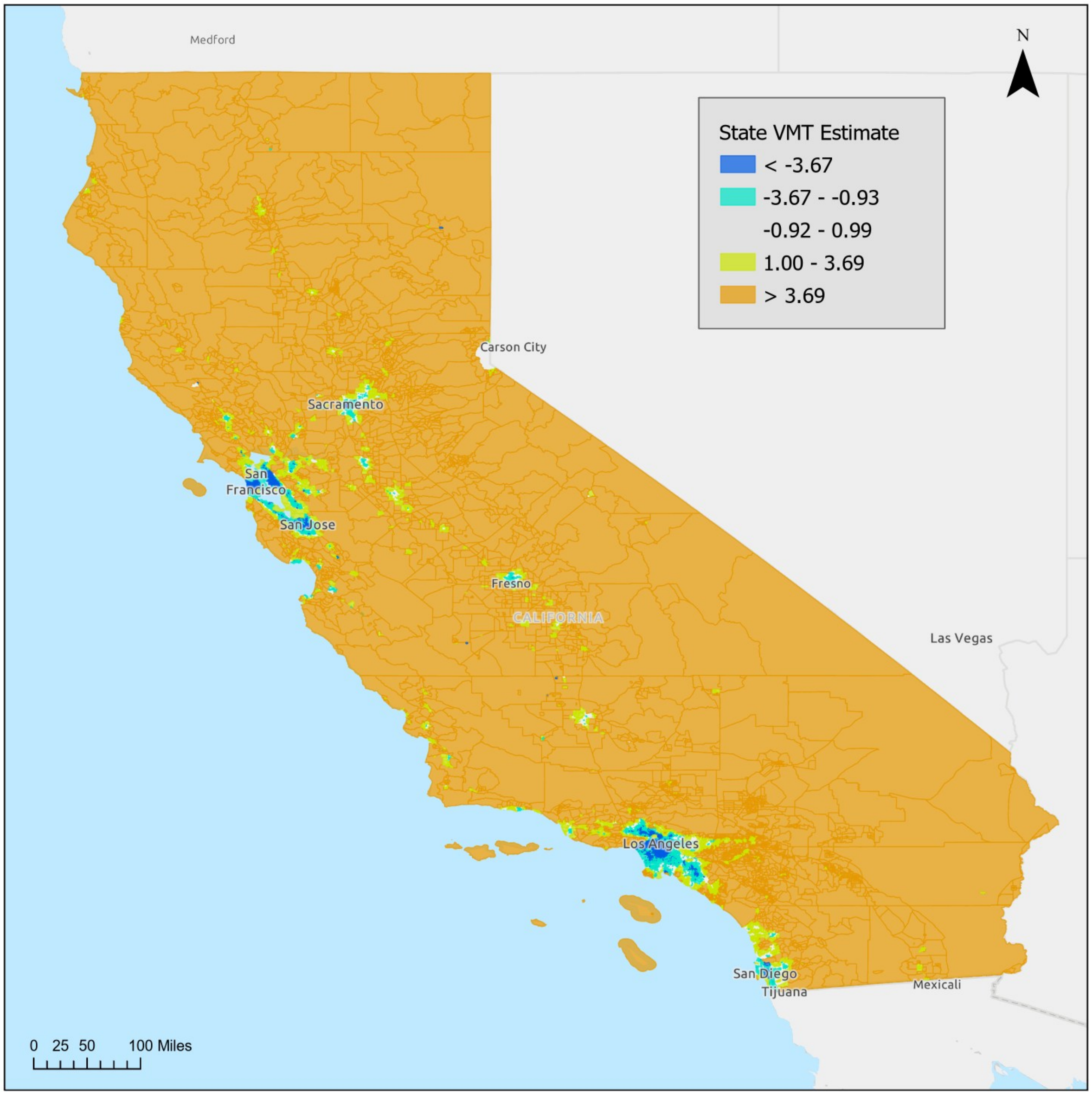
Because demographics have been controlled for, these estimates show the variation in VMT associated with built environment characteristics only. More specifically, the VMT estimates shown in these maps are based only upon the built environment characteristics measured at the block group level, meaning that walking distance to rail is not included. To visualize where transit proximity further reduces VMT estimates, we include in the maps rail station locations as of 2017-18 and half mile buffers around those.

In the context of our analysis, we are primarily concerned with relative differences between block group VMT estimates, rather than the absolute estimates themselves. How we calculate the relative difference in VMT is a function of how we determine our counterfactual. We describe two counterfactual scenarios that we addressed directly below; one based on the state average and one based on CBSA-specific averages.

State based counterfactual

The maps below display the estimated variation in VMT associated with built environment and transit characteristics across the state, at the block group level, based on the whole-sample estimates in Task 2 shown in Table 2.7. (Walking access to rail within a quarter mile is not included in these maps; it is not visible at the scale of the state.) The state average is calculated as the mean population-weighted VMT variation associated with built environment characteristics for all block groups in California. For a counterfactual in which the project would have been built somewhere else in the state, at some “average” statewide location, we subtract the block group VMT estimate from that state average to display how much more or less VMT is estimated to be generated as a function of the built environment characteristics of that block group, relative to the state (Figure 5.1). The choropleth map designates each block group depending on the VMT quantile to which it belongs, and thresholds for these quantiles are seen in the legend. Cooler colors indicate lower VMT estimates than the state average, and warmer colors indicate higher VMT.

Figure 5.1: Difference in VMT estimates from the state average by block group



Notes: Estimated VMT values by block group are represented as absolute differences from the California population-weighted mean VMT variation associated with built environment characteristics (-1.297 miles). Choropleth brackets are by quintile.

The maps show that a majority of the state by land area is above the state average. Block groups that are designated as high VMT are typically much larger (spatially) given their low population densities. Block groups with lower VMT contain a higher share of population per land area and are clustered in the more urbanized areas of the state. Smaller block groups in urban areas typically have higher population densities and higher levels of the built environment characteristics (like bus stop density and jobs accessibility) that result in lower VMT estimates. The Los Angeles and San Diego regions also have a lower share of block groups in the lowest VMT quantile when compared to the urbanized regions of Northern California, due to the CBSA coefficients (see Table 2.7) that identified the Inland Empire and Los Angeles regions as having significantly higher VMT than the rest of the state.

Metropolitan area CBSA based counterfactual

An alternative counterfactual is that an AHSC project would have been built at an average location elsewhere in the same metropolitan area. We follow a similar method as to the above, subtracting the block group VMT estimate from the mean population-weighted block group-based VMT estimate for the Core-Based Statistical Area (CBSA) to which the block group belongs. This relative VMT calculation estimates how much more or less daily VMT per capita would be generated in the block group of interest, compared to the average location elsewhere in the CBSA. We present the relative VMT estimate difference from the CBSA average for all block groups in the state in Figure 5.2. Population-weighted average CBSA values can be found in Table 5.1.

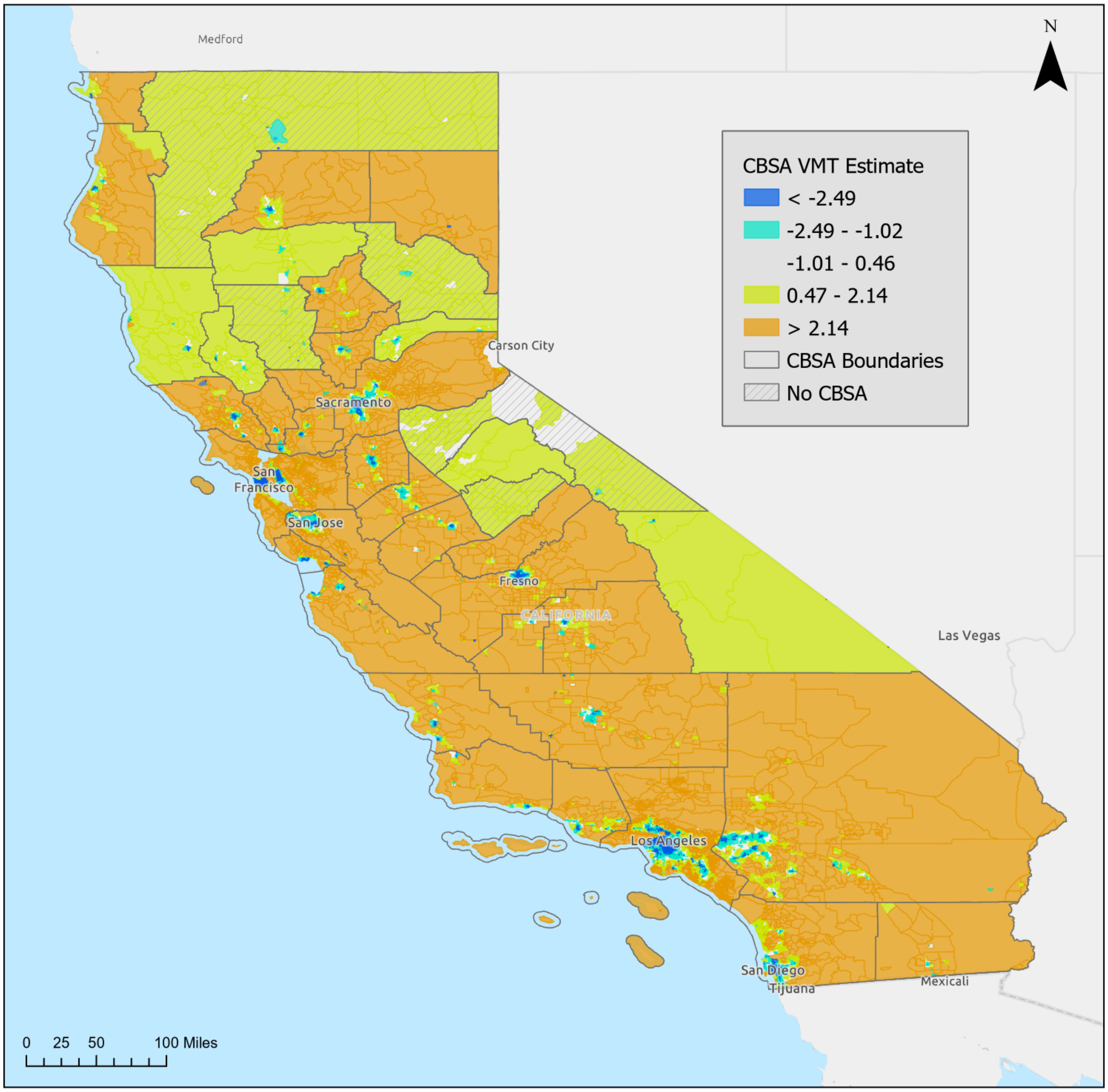
The main difference in calculating CBSA estimates is that we do not include CBSA coefficients when calculating absolute VMT estimates, prior to subtracting from the population-weighted average to get the relative difference. CBSA coefficients from the model are not relevant in this instance and are therefore excluded.

Table 5.1: Population-weighted average VMT estimates by CBSA

CBSA	VMT estimate
El Centro, CA	2.185
San Diego-Chula Vista-Carlsbad, CA	-1.244
Los Angeles-Long Beach-Anaheim, CA	-3.257
Hanford-Corcoran, CA	2.104
Visalia, CA	2.116
Modesto, CA	0.634
San Francisco-Oakland-Fremont, CA	-5.542
Riverside-San Bernardino-Ontario, CA	4.351
San Jose-Sunnyvale-Santa Clara, CA	-3.135
Truckee-Grass Valley, CA	3.694
Chico, CA	2.153
Yuba City, CA	2.002
Santa Maria-Santa Barbara, CA	0.121

CBSA	VMT estimate
San Luis Obispo-Paso Robles, CA	1.929
Eureka-Arcata, CA	2.710
Sacramento-Roseville-Folsom, CA	-0.360
Oxnard-Thousand Oaks-Ventura, CA	0.522
Susanville, CA	0.502
Salinas, CA	0.013
Santa Cruz-Watsonville, CA	-0.265
Stockton-Lodi, CA	0.407
Redding, CA	2.739
Vallejo, CA	0.176
Fresno, CA	0.359
Bishop, CA	3.907
Bakersfield-Delano, CA	0.914
Merced, CA	1.762
Clearlake, CA	3.681
Sonora, CA	3.681
Santa Rosa-Petaluma, CA	0.603
Crescent City, CA	2.985
Napa, CA	1.065
Ukiah, CA	3.434
Red Bluff, CA	3.806
Gardnerville Ranchos, NV-CA	5.532
No CBSA	4.325

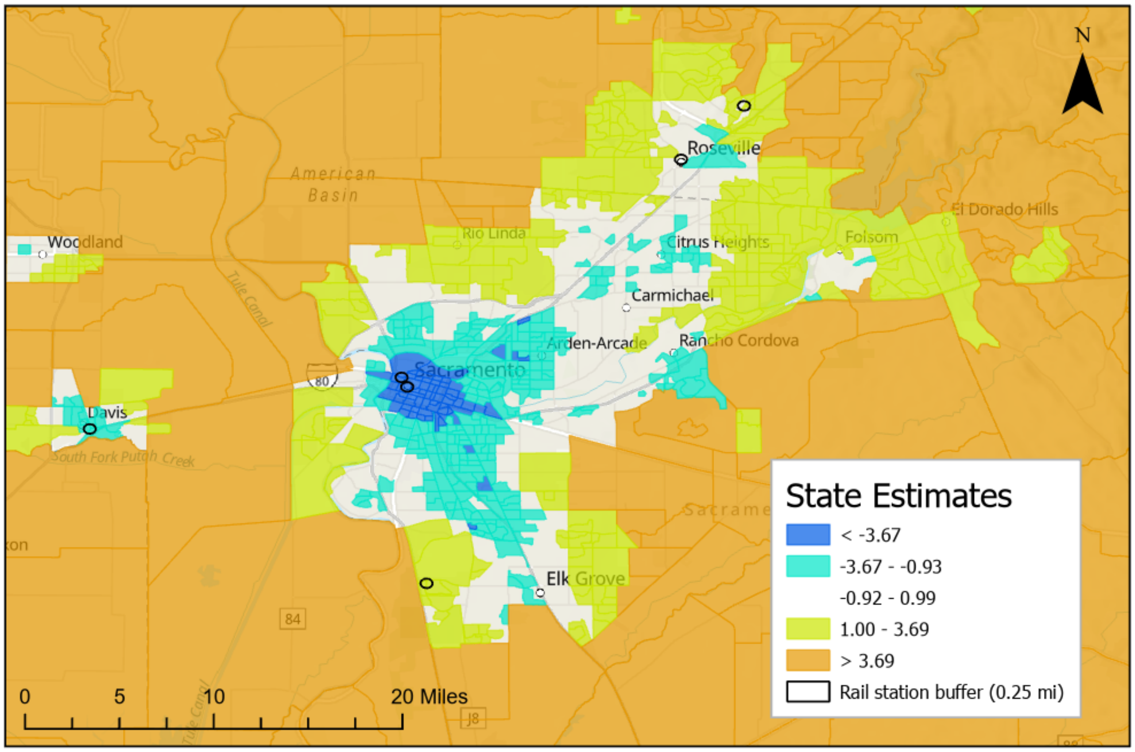
Figure 5.2: Difference in VMT estimates from home CBSA average, by block group

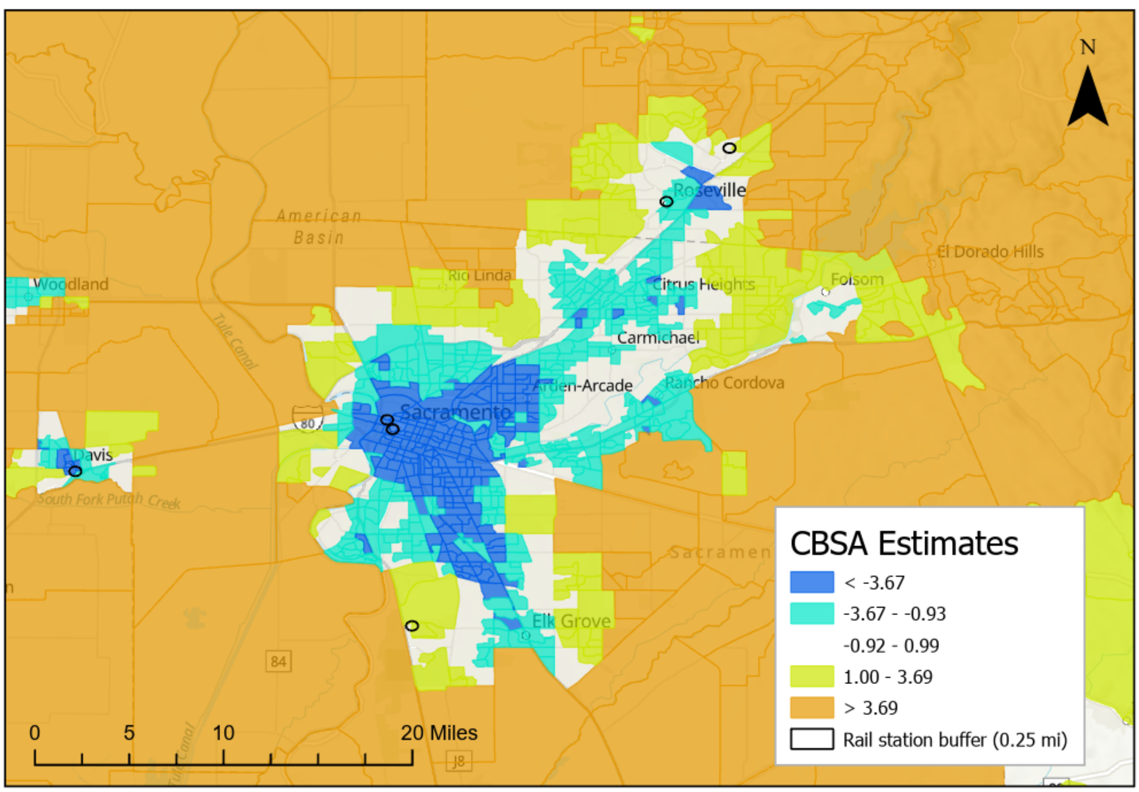


Note: Estimated VMT values by block group are represented as absolute differences from the population-weighted mean VMT variation associated with built environment characteristics for the home CBSA. Choropleth brackets are by quintile.

We can compare the VMT estimates, relative to the state and CBSA region, to see how different assumptions of the counterfactual may impact our estimates. Figure 5.3 shows an example of how block group-based VMT estimates differ in the Sacramento region. Note that when the VMT estimates are compared to the state average (top), fewer block groups fall into the lowest quintiles of VMT variation estimates. This is because the state average (-1.297) is lower than the CBSA average (-0.360)

Figure 5.3: State vs. CBSA VMT estimates (Sacramento region)





Notes: Estimated VMT variation values are generated using the base full sample model from Task 2. Choropleth brackets are by state-based VMT estimate quantiles.

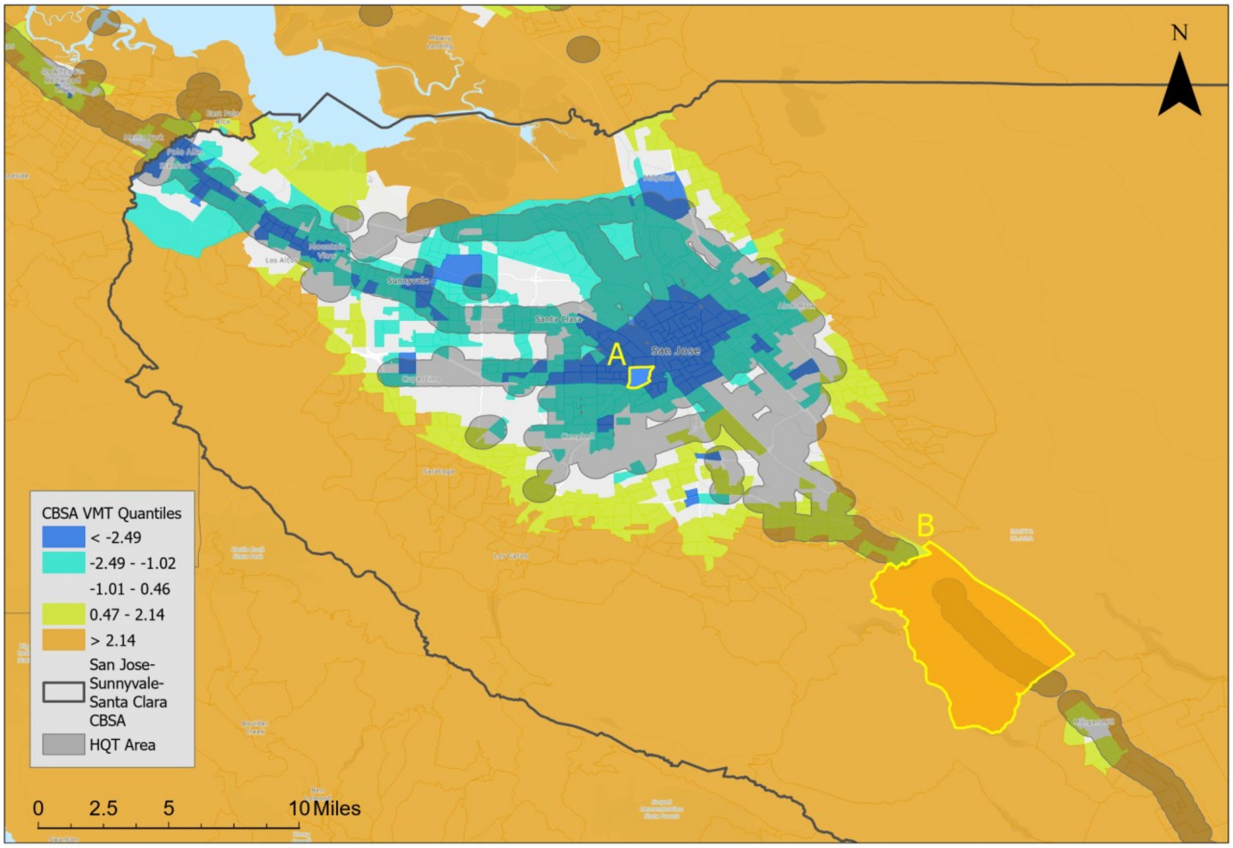
Block group VMT estimate examples

Finally, we provide an example of how block groups within a metro area can vary with respect to estimated VMT variation, even among those that are eligible for AHSC funding. We look specifically at block groups eligible for possible TOD projects, since projects are compared to other projects of the same project area type (TOD, ICP, and RIPA) in the application process. Although (as explained earlier) we do not have data on all "high quality transit" locations for 2017/2018, conforming with the year of our travel survey and built environment data and land use, we can for illustrative purposes use present-day high quality transit locations to identify eligible sites for possible TOD projects by drawing a 0.5 mile buffer (the "HQT Area") around these transit stops. Note, again, that the VMT estimates presented here are based on built environment characteristics that were measured in 2017/2018. For current use, estimates should be updated using present-day built environment data (as we discuss in our Recommendations section, below). While we focus on TOD projects specifically, ICP and RIPA projects are not subject to these siting constraints and could also be built in other locations throughout the state that are not identified here.

In the San Jose-Sunnyvale-Santa Clara CBSA, as can be seen in Figure 5.4, any block group that is intersected by the HQT Area is eligible for development proposals. While most block groups eligible for TOD project siting (according to the transit proximity criterion) have relatively low VMT estimates, there are some locations where block groups with high VMT estimates are eligible for the program as currently constituted. For example, two block groups of interest are Block Group A (FIPS: 060855019001) and Block Group B (FIPS: 060855121001), as labeled in the figure. Block Group A is located close to downtown San Jose and belongs to the lowest quintile of VMT estimates for the metro area. Block Group B is located further from the urban core, along the Highway 101 corridor, and belongs to the highest quintile of VMT estimates for the metro area.

In this case (and in the case of all our CBSA-based estimates) the counterfactual is embedded in the calculation. The counterfactual here is that the AHSC project would have been built somewhere else in the CBSA, and therefore by measuring the VMT variation from the CBSA average we are saying that this is how much daily VMT would increase or decrease if the project were built in this block group, compared to the case where it would have been built somewhere else in the CBSA.

Figure 5.4: CBSA based VMT estimates for block groups in the San Jose-Sunnyvale-Santa Clara CBSA, with highlighted sample "low-VMT" and "high-VMT" block groups.



We can observe the estimated effects of the built environment on VMT for these two specific block groups in

Table 5.2. Here we break down the VMT estimate by term to see how much each respective built environment characteristic contributes to the total VMT variation associated with built environment and transit characteristics, and how this absolute VMT value factors into the final VMT estimate relative to the CBSA average.

In Table 5.2, see that the San Jose average VMT estimate is noted in the last row, and that this value can be thought of as the counterfactual. It is then subtracted from the "total of all contributions" to give the amount of VMT per capita that would be saved or lost, compared to this counterfactual. As shown in the table, in Block Group A, the average individual would be estimated to generate 3.8 vehicle miles *less* than the average individual elsewhere in the San Jose-Sunnyvale-Santa Clara CBSA. The average individual in Block Group B, on the other hand, would be estimated to generate almost 6 miles *more* VMT per day than the average individual in the CBSA.

Table 5.2: Example VMT Variation Estimates with CBSA counterfactual.

VMT estimate contribution	Block Group A	Block Group B
Retail density (block group)	-0.381	0.132
Bus stop density (1 mi radius, BG-based)	-0.849	0.327
Network load density (2 mi radius)	-1.949	1.835
Transit to car jobs accessibility ratio	-1.322	-1.542
Jobs accessibility by car	-0.699	0.279
Street density (2 mi radius)	-1.690	1.763
Total of all contributions	-6.891	2.795
Total relative to CBSA (San Jose average: -3.135 miles/day)	**-3.756**	**5.929**

Notes: All numerical values in the table are in miles per day per capita.

Task 6: Analysis of Current Quantification Methodology and Data Needed for Improvements

Key Research Question & Basic Theory

The work in this section encompasses two main objectives that support findings from Task 1 through Task 5,and provide context for our findings and recommendations in relation to the AHSC Benefits Calculator Tool.

The first goal was to document and evaluate existing methods and data sources for calculating unmitigated and mitigated VMT. As part of this task, we discuss how to estimate the VMT that likely would have been generated had the project not been funded. We use the term “counterfactual” to refer to the fact that this calculation is speculative, and requires careful documentation. The term “baseline” might be used, except that that term implies some certain comparator. Here, we are referring instead to what would have happened in the absence of the program funding being awarded to an AHSC project.

Our second goal was to provide the data and directions necessary to estimate VMT reductions associated with AHSC project proposals according to the model results described in the Task 2 writeup. This work is provided in the format of an Excel Workbook (See Appendix C:). The Excel workbook provides example calculations of VMT estimates by block group, throughout the state. Note that this is a sample Workbook and the calculations made here should not be applied as they exist currently. Model outputs and variable coefficients must be updated.

While the first step of Task 6 can be understood as an evaluation of methods embedded in the Benefits Calculator Tool, the second can be understood as identifying data needed to update the Benefits Calculator Tool according to our research findings.

AHSC Quantitative Methodology & Benefits Calculator Tool overview

The AHSC Quantification Methodology was developed by CARB to estimate the potential benefits of proposed projects The AHSC Benefits Calculator Tool automates the Quantitative Methodology for applicants, along with a step-by-step guide intended to assist applicants with filling out the Tool and providing adequate documentation. Once the calculator has been fully filled out, applicants can see what the estimated GHG emissions reduction outcomes and co benefits are for their proposed project, as well as the estimated total project GHG emissions reductions per dollar of funds requested.

Various sheets within the Benefits Calculator Tool exist for different project Inputs, including Affordable Housing, Active Transportation, Transit, Shared Mobility, and Solar PV. The Affordable Housing Inputs are of main interest here, given the objectives of this research, though our results may also apply to the Transit Inputs tab as well. A full breakdown of the Methodology and variable definitions can be found in the AHSC Quantification Methodology (2023) document. The process is summarized as follows: first, total unmitigated VMT is calculated according to the number of proposed units and the assumed amount of annual VMT for these units based on state and county averages for number of trips and trip lengths, respectively. The trip lengths are reported by Metropolitan Planning Organizations or derived from the Caltrans California Statewide Travel Demand Model. As defined in Table 6.1, VMT reduction contributions are then calculated in the areas of Land Use, Parking, Traffic Calming, and Subsidized Transit. For more detail on how these reductions are calculated using the Quantitative Methodology, see Appendix Table A.2. Note that there are certain elasticities for each contribution and maximum reduction percentages allowed both for each contribution and in each area, based on project type (TOD, ICP, or RIPA). The maximums are cited from the California Air Pollution Control Officers Association (CAPCOA) “Quantifying Greenhouse Gas Mitigation Measures” report (2010), and other various sources, as noted in Appendix Table A.2. The reductions in each area are added together to find a total reduction percentage, which is then multiplied by the total unmitigated VMT to calculate the “avoided VMT” of a project. This avoided VMT amount contributes to the estimated GHG emissions reductions for the project, in addition to estimated reductions associated with investments in other Input areas. Notably, the Quantitative Methodology only allows for non-negative reduction values, and therefore no reduction area can cause a VMT increase from the baseline.

Table 6.1: Components of VMT reduction percentage calculation

Component	Description
Land Use Reductions (Max % - TOD: 65%, ICP: 30%, RIPA: 5%)	
Increase Density	Based on the increase in building-level project density (units per acre) above the requirement, based on project type (TOD, ICP, RIPA), with an elasticity of 7% and maximum of 30%.
Increase Diversity	Based on building-level percentages of residential and mixed-use space (by square foot), contributing

Component	Description
	to a “land use index), with an elasticity of 9% and maximum of 30%. The land use diversity index is applied to the share of the project area devoted to residential and public uses and compared to a single-use building index (represented by 0.15).
Increase Destination Accessibility	Based on comparing the employment density (jobs per square mile) of the block group to the county average (sourced from EPA SLD), with an elasticity of 7% and a maximum of 30%. Job density of the block group is identified by the applicant using the third-party Policy Map tool, where the applicant can enter the address of the project and the tool returns the employment density for the related block group. Applicants must submit a screenshot to verify. *Note that this is the only reduction contribution that takes into account neighborhood characteristics.
Integrate Affordable Housing	Based on the percentage of affordable units out of total dwelling units, with an elasticity of 4%
Parking Reductions (Max % - with TC, TOD: 50%, ICP: 35%, RIPA: 10%)	
Limit Parking Supply	Based on a reduction in total parking offered from the number of default spaces (defined by land use subtype- apartments, condos, low/mid/high rise, etc), with a maximum of 12.5%
Unbundle Parking Costs	Based on cost of an unbundled parking spot, with a maximum of 20%
Other Reductions	
Traffic Calming Measures (TC)	If traffic calming is included, 1% is given
Subsidized Transit Programs (ST)	Based on urban/rural and bracket of annual transit pass value (from Trip Generation sheet), with a maximum of 20%

Component	Description
Total = Land Use + Parking + TC + ST (Max % - TOD: 55%, ICP: 40%, RIPA: 15%)	

Evaluation of Quantitative Methodology

With this context we now identify some issues with the Quantitative Methodology and Benefits Calculator, and how they are informed by this research project and previous empirical research.

Establishing the counterfactual: Baseline VMT

To estimate any possible VMT reduction associated with an AHSC-funded development proposal, one must also estimate what the VMT level would have been without the funding decision. In this report we refer to this as a "counterfactual." In the CARB calculator, the nearest analogous comparison is "unmitigated VMT." One way to estimate this counterfactual or "unmitigated" VMT could be to estimate the VMT associated with a development in the same location that is not funded by AHSC. Based on this counterfactual, the VMT reduction would be estimated based on, for example, the estimated changes in VMT due to any incentives AHSC provided to incentivize the developer to increase density, encourage mixed uses, and reduce parking at the site. As noted previously in this report, our research and the previous literature suggests that in this case, if the proposed AHSC project were a dense, mixed use building with the same amount of parking as a conventional building, there would likely be *no* impact on VMT. This is because, as shown above, while built environment characteristics at a larger spatial scale than the development site are important, there is little to no evidence that building level characteristics other than parking are influential. (One exception would be a very large project that materially changes the measured characteristics of the area surrounding the project, at the Census block level and higher.) We note here, furthermore, that there is no *a priori* reason to believe that every AHSC project would necessarily reduce VMT. There might be instances where in fact VMT could be increased by providing funding and altering the characteristics of the project, if the location is altered substantially from where it might have been. For example, there are very likely instances in which proximity to high quality transit does not offset the VMT impacts of the location's built environment characteristics other than transit availability.

Setting aside the question of *development*, the counterfactual with respect to VMT generation should be understood more broadly as: "What will happen to the VMT of the population if AHSC does not fund this project" (or, if looking back, "What would have happened if AHSC had not funded the project?") There are many plausible possibilities in addition to the example discussed directly above. One possibility is that the building would have been developed exactly in the same location, with exactly the same number of units, and all of the other characteristics being identical to what was funded; in that case, the AHSC funding would be understood to have no effect on VMT generation. Another plausible counterfactual is that the AHSC funding is enabling (or enabled) the housing to be developed in this location, and that absent the funding, the residents of the AHSC project would have had to find housing elsewhere—perhaps somewhere else in the state, somewhere else in the metropolitan area, or somewhere else nearby. It is this latter counterfactual that the CARB calculator comes closest to using, because in that calculator, the assumption is that projects start with the average VMT of the county in which they are developed, and then are given credits for reduced VMT from that county average.

We do not conclusively resolve the dilemma of identifying the appropriate counterfactual as a part of this research project; it is impossible to identify alternative futures (or histories) with certainty. But as noted in the Task 5 writeup, we calculated a counterfactual VMT estimate for each metropolitan area in the state along with a counterfactual consisting of the weighted state average. We view both of these as plausible counterfactuals for an upper bound of potential VMT impact of AHSC projects. They assume that in the absence of the funding, the residents of the AHSC project would instead live (or would have instead lived) in a location with average built environment characteristics, either for the state as a whole, or for the specific metropolitan area within the state. It is outside the scope of our current research project to provide additional counterfactual calculations, but another possibility is to compare the average built-environment-related VMT variation for a subset of Census block groups within some radius of the AHSC project location representing a residential search area for those residents.

To reiterate, the current counterfactual assumption in the CARB calculator is to take the county average trip length estimates and revise those downward under the assumption that building level characteristics reduce VMT. We instead recommend comparing the VMT variation associated with the built environment for the proposed project (from our Task 2 model) with the estimated VMT variation for the metropolitan area, the state, or some other area that is seen as the appropriate comparator. Since housing markets are known to be regional, it seems quite plausible to us that the metropolitan area (CBSA) average should be used as a default.

Impact of Building- vs Neighborhood-Level Characteristics

To what extent are VMT reductions associated with the characteristics of the building itself, versus the surrounding environment? As shown previously in this report the neighborhood characteristics are far more likely to have significant impact on VMT generation than site characteristics. While it may be a programmatic priority to incentivize building-level improvements, we do not find evidence in our study or in empirical literature that they affect VMT, with the important exception of off-street parking provision.

Most specific reduction calculations in the Quantitative Methodology are dependent on building-level inputs, rather than neighborhood-level factors. Reductions associated with increasing density, mixed use development, and parking are in reference to building-level requirements based on project type (TOD, ICP, or RIPA), rather than the conditions of the neighborhood or larger surrounding area. The only exception to this rule is the VMT reduction area “Increase Destination Accessibility,” which factors in the employment density in the block group relative to the county average. As shown previously, we found that employment density in the block group was not statistically significantly associated the variation in VMT. However, one block-group level measure--retail employment density--was, and we recommend that this effect be included in estimating the VMT for proposed projects. In addition, though proximity to transit is a requirement for the program, it is not considered as an aspect of VMT reduction. This is one of the main concerns CARB identified for our research to address. As noted above, we did find that having a rail station within a quarter mile of home was associated with significantly lower VMT, and we recommend that this effect be included as well, subject to the caveat that is it quite possible that rail proximity within a quarter mile is partly capturing site-level parking availability.

Regional Specificity and Comparison

With consideration of neighborhood specificity in VMT reductions, new questions arise. At what spatial scale are neighborhood-level factors most significant? Once a level of spatial measurement for a neighborhood-level factor is chosen, how should this factor be compared? Relative to a county average, some other regional average, a standard based on project type (TOD, ICP, or RIPA), or another way? In addition, how will applicants be able to access this information publicly?

Calculating neighborhood-level characteristics limited only to the level of the Census block group, as is done in the VMT reduction area “Increase Destination Accessibility,” is not consistent with research showing that larger spatial scales are influential, and with the results from our analysis as shown above. Also, this VMT reduction estimate is calculated by comparing block group employment density to the county average. As far as we are able to infer, this is not based on empirical analysis. When comparing block group to county density, a proportional increase in density is scored the same. For example, a development in a block group with 250 jobs/sq mi and a county average of

500 jobs/sq mi would be assigned the same VMT reduction estimate as a development in a block group that had 8000 jobs/sq mi and a county average of 4000 jobs/sq mi, though the actual effect on VMT generation would be quite different. Further, two block groups with the same density in counties with widely varying county averages would be awarded very different VMT reduction percentages as a result. This assumption is not consistent with previous research on VMT variation with the built environment or with our analysis results described above.

The current structure of the calculator also poses challenges for the accessibility of neighborhood-level data for applicants. Reliance on third-party sites for up-to-date data limits how much neighborhood-level information contributes to VMT reductions, and at what spatial scales the data is available. We discuss further in this section how publicly available built environment data could be obtained from various sources by CARB (as we did for this project), and included within a revised calculator tool to make this information and VMT reduction calculations accessible to applicants. See the technical methodology for recommendations section, as well as Appendix C:.

Additivity of Elasticities

The calculator currently sums percentage reductions in each area to find a total VMT reduction percentage, with maximums set as previously described. The VMT elasticities and reduction caps used here are cited from the 2010 CAPCOA "Quantifying Greenhouse Gas Mitigation Measures" report, with TOD projects corresponding to the "urban" project setting, ICP to "compact infill," and RIPA to "suburban."

The issue here is twofold. First, the (somewhat outdated) literature to which reference is made in the guidance defines these elasticities and maximums based on neighborhood-level characteristics, but in the Benefits Calculator Tool they are applied at the project level (as discussed above). Second, assuming the elasticities are additive neglects interactions between built environment characteristics. For example, VMT may vary widely between two developments that have the same level of retail density but very different amounts of parking supply when all other factors controlled for. We tested a number of interaction terms in our analysis described above and found two highly significant interactive factors: "network load density" and the ratio of transit-to-auto job accessibility within a 45-minute travel distance.

Technical methodology for recommendations

The AHSC Quantitative Methodology for VMT reductions from Affordable Housing Inputs can be updated according to our research findings as follows. VMT estimates can be calculated for every block group in the state using the coefficients for built environment variables from the appropriate model. We have provided an example using the Task 2 base full sample model (without interactions), which we showed in a map in the Task 5 writeup. Note that in doing so we applied the model coefficients to the built environment measures from 2017/18. To update block group VMT estimates for more recent years, we recommend that CARB apply model coefficients to updated built environment measures reflecting more recent changes in population, employment by industry, street networks, and public transportation services. Built environment metrics should always be updated to reflect more frequent changes in urban form over time. We can assume for the purposes of VMT estimation that the ways in which the built environment influences travel behavior among individuals are less likely to change significantly over time, but it would also be helpful in future to re-estimate the statistical model when travel survey data are again collected in the state for a year later than 2017.

The data and calculations necessary to implement the recommendations and estimate VMT by block group are included in the Excel Workbook located in Appendix C:.This includes relevant built environment and transit data; the means and standard deviations with which to scale variables to Z-scores; block-group level VMT calculations and the relative contributions of each built environment characteristic to the absolute estimate; and relative VMT estimates compared to two counterfactuals: the state and CBSA estimated VMT contribution associated with built environment variation. The first sheet, containing the built environment variables by block group, can be updated to develop new VMT estimates that reflect changes in the built environment. Note that while many built environment variables were originally sourced from the 2017 Environmental Protection Agency Smart Location Database (EPA SLD) for our analysis, it is unknown at this time when this data will be updated. Table 6.2 provides a description of all of the built environment data needed to calculate updated VMT estimates by block group and from what source they can be updated from in the future if previously collected from the EPA SLD.

Table 6.2: Sources for built environment variables used in VMT estimation

Variable	Source
Retail density	Longitudinal Employer Household Dynamics (LEHD) workplace area characteristics (WAC)
Bus stop density	General Transit Feed Specification (GTFS) data, processed according to Appendix Section B.3: Cleaning bus stop data
Jobs accessibility by car	LEHD WAC, 2020 Google TravelTime API

Variable	Source
Transit to car jobs accessibility ratio	LEHD WAC, 2020 Google TravelTime API, 2020 GTFS
Street density	OpenStreetMap via OSMnx
Network Load Density	American Community Survey 5-year estimates, LEHD WAC, OpenStreetMap via OSMnx
Proximity to transit	Input directly by project applicant

Summary of findings

In this report we carried out a series of data analyses using multiple sources of data to investigate influences of demographic, housing and built environment factors upon vehicle miles traveled (VMT) or travel activity levels, with controlled quantitative statistical methods. Here we summarize those results and describe what they mean for correctly estimating the net VMT impacts associated with AHSC project proposals.

In Task 1, we examined travel distance data from satellite measurements of personal cell phone movements, aggregated to Census block groups, measured before and after the development of funded AHSC housing projects between 2018 and 2023. When controlling for other factors and looking at differences in travel distance levels before and after development, we found no evidence of changes in per capita travel distance in block groups that received such development in comparison to block groups that did not. This was even when controlling for factors such as the size of the developments, and even when looking for differences using varying time lags and other tests of robustness.

Although our measure of travel distance in Task 1 did not allow us to control for differences in travel by mode, our Task 1 findings are consistent with the previous literature on the built environment and travel, which has largely focused on larger-scale built environment measures. Our Task 1 results suggest that AHSC projects in low- or high-VMT neighborhoods result in similar travel behavior to the existing residents of those Census block groups. This is consistent with the hypothesis that project-level development characteristics do not cause VMT to vary significantly, or that the building-level characteristics of AHSC projects have typically been consistent with those of surrounding development. Interestingly, in slightly different (cross-sectional) analysis of the activity data, we do find that larger projects in low-VMT areas have a reduction in VMT in comparison with the surrounding block group, thus increasing the rationale for prioritizing projects located in low-VMT areas.

In Task 2, we analyzed cross-sectional travel diary data collected across the state of California in 2017by the National Household Travel Survey (NHTS). We compiled contemporary built environment and transit data and calculated a large set of built environment and transit measures thought to influence travel, including transit access (as measured by the walking distance to the nearest rail station and the number of transit stops nearby), population density, employment density, street density, and the accessibility of jobs by different modes. We tested the significance of these characteristics at different scales, from the local environment (as measured at the level of the Census block group) to as far as a two-mile radius around the home block group.

Consistent with previous literature, we found that the built environment measures most often strongly correlated with vehicle mileage traveled in the NHTS data were at a scale greater than the local Census block group, underlining the importance of larger-scale characteristics in estimating VMT variation associated with housing development proposals. Those factors most correlated with VMT reductions were rail station proximity to home (particularly within a one-quarter-mile walking distance), network load density measured at the two-mile radius around home (defined as the sum of residents and workers divided by street length), regional employment access (within 45 minutes of home), street density (within 2 miles of home), and bus stop density (within one mile of home).

In summary, while Task 1 showed that residents of new AHSC projects travel distances very similar to that of surrounding residents, Task 2 showed which neighborhood-, community- and regional-level characteristics are the most highly correlated with lower levels of auto use.

Regarding Tasks 3 and 4, we found no evidence that larger housing units are associated with lower VMT per capita, suggesting that unit size should not be considered in a VMT calculator. And in our most tailored analysis of how the built environment affects the VMT of low-income seniors, we did not find any statistically significant evidence of a differential effect for them, suggesting that a calculator for the low-income population is sufficient for use with age-restricted housing projects.

Recommendations for the CARB VMT reduction calculator

We recommend that the CARB calculator be revised based on the findings of this analysis, consistent with previous scholarly empirical literature on the topic, via the following primary recommendations.

1. **Remove all building-level density and land use measures**. Measures such as building level population density, currently included in the calculator, represent applications of neighborhood built environment findings from the voluminous empirical literature on this topic that were incorrectly used at the site (proposal) level. There is little to no theory supporting the equivalency of neighborhoods to buildings in terms of their influence upon household travel choices. There is little to no evidence in empirical literature, nor in our Task 1 results, that characteristics of AHSC projects distinguish their per capita VMT from that of surrounding housing developments in the Census block group. (The possible exception to this would be an instance of a large project significantly altering the characteristics of the entire Census block group, particularly its retail employment density; or a project with reduced off-street parking, as noted below.)

2. **Base the calculator primarily on neighborhood, community, and regional built environment measures, including walking distance to rail.** The significant built environment measures in our research, consistent with previous empirical literature, include the walking distance to rail; bus stop density (measured within one mile of the home Census block group centroid); network load density (within two miles of home); street density (within two miles of home); and employment accessibility (measured within a 45-minute travel distance by both auto and transit). The calculator should be based upon applying the regression coefficients from Task 2 to present-day measures of those built environment features. In other words, the expected VMT variation for project proposals is a function primarily of which Census block group they are developed in, in addition to the walking distance from the proposed project to the nearest rail station. In Task 6 we provided instructions for updating the built environment measures to current data.

3. **Of the building-level measures, retain only off-street parking availability**. Per the scholarly empirical literature cited above, the evidence is strong that the amount of off-street parking available to tenants in multifamily buildings significantly influences their auto use by affecting the cost of owning and parking a personal vehicle. The current CARB calculator includes a parking measure. We would suggest that this be retained, though modified to reflect quantitative estimates available in several studies. It is also possible that if a measure of off-street parking had been available for use in our Task 2 model, this would have

reduced the coefficients on the other built environment measures, particularly the measure of walking distance to rail.

4. **Adopt a more accurate counterfactual reflecting likely alternatives for households and/or likely alternative growth patterns for the state.** The current calculator assumes that proposed projects reduce VMT from a county average. (In fact, any given project proposal could either increase *or* decrease VMT from that average.) Instead the counterfactual should be defined as a likely alternative housing option that would have been chosen by a housing developer, or by a relocating household, in the absence of the program subsidy. While in some cases the county average might be the appropriate alternative, in most, the best measure of the counterfactual development alternative may be the housing market for developers, since that is more likely to relate to what might have happened in the absence of the AHSC subsidy. One measure of the average counterfactual in the housing market would be the VMT average for the metropolitan area. There is also an argument for using the state average VMT as the counterfactual, insofar as the spatial pattern of state population growth is driven by municipal policies. A third alternative, highly defensible but more difficult to calculate and implement, would be a radius-based counterfactual measure such as the cluster of Census block groups within 10 to 20 miles of the project, which would reflect the possible search process of a household seeking a residence.

A final note is that we were recently asked about the impact of transit subsidies on VMT and how transit subsidies (e.g., reduced transit fare for AHSC residents or for other persons) might be approached in the calculator. Addressing transit subsidies was outside of the scope of our work, and hence the results in this research do not directly inform how transit subsidies may affect VMT. That said, our understanding is that the current calculator uses an approach that draws from the CalEEMod which uses estimates of transit fare elasticities from a 2012-2014 CARB research project. CARB is currently updating that research. The CalEEMod approach relied on estimates of transit price elasticity from the literature in the early 2000s, and it used an estimate of -0.4, meaning that a 10% transit fare reduction is assumed to increase transit ridership by 4%. But to estimate how transit fare reductions would reduce VMT (as opposed to increase transit use), one needs to understand how transit trips replace car trips (generally not one for one) and the trip distance of any reduced transit trips in relation to total VMT. The CalEEMod approach used an estimate that 50% of new transit trips would replace a car trip. Note that the policy brief included estimates of 2% and 16%, but the calculator implements the 50% estimate, the highest estimate cited. It is not clear that the 50% replacement of transit for car trips is more reliable than 2% or 16% (see p7). Finally, the CalEEMod approach did not describe how to convert from reduced trips to changes in VMT.

References

Ashik, F. R., M. H. Rahman, and M. Kamruzzaman. "Investigating the Impacts of Transit-Oriented Development on Transport-Related CO2 Emissions." *Transportation Research Part D: Transport and Environment* 105 (April 1, 2022): 103227. https://doi.org/10.1016/j.trd.2022.103227.

Barajas, Jesus, Karen Frick, and Robert Cervero. "Travel of TOD Residents in the San Francisco Bay Area: Examining the Impact of Affordable Housing," June 30, 2020. https://escholarship.org/uc/item/1r20w0tv.

Boarnet, Marlon G., Raphael W. Bostic, Seva Rodnyansky, et al. "Do High Income Households Reduce Driving More When Living near Rail Transit?" Transportation Research Part D: Transport and Environment 80 (March 2020): 102244. https://doi.org/10.1016/j.trd.2020.102244.

Buehler, Ralph, John Pucher, Rico Wittwer, and Regine Gerike. "Travel Behavior of Older Adults in the USA, 2001-2017." *Travel Behaviour and Society* 36 (July 1, 2024): 100783. https://doi.org/10.1016/j.tbs.2024.100783.

Cervero, Robert, and Kara Kockelman. "Travel Demand and the 3Ds: Density, Diversity, and Design." *Transportation Research Part D: Transport and Environment* 2, no. 3 (September 1, 1997): 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6.

Cervero, Robert, and Jin Murakami. "Effects of Built Environments on Vehicle Miles Traveled: Evidence from 370 US Urbanized Areas." *Environment and Planning A* 42, no. 2 (February 1, 2010): 400–418. https://doi.org/10.1068/a4236.

Chatman, Daniel G. "Deconstructing Development Density: Quality, Quantity and Price Effects on Household Non-Work Travel." *Transportation Research Part A: Policy and Practice* 42, no. 7 (2008): 1008–30. https://ideas.repec.org/a/eee/transa/v42y2008i7p1008-1030.html

Chatman, Daniel G. "Does TOD Need the T?" *Journal of the American Planning Association* 79, no. 1 (January 2, 2013): 17–31. https://doi.org/10.1080/01944363.2013.791008.

Chatman, Daniel G., Ruoying Xu, Janice Park, and Anne Spevack. "Does Transit-Oriented Gentrification Increase Driving?" *Journal of Planning Education and Research* 39, no. 4 (December 1, 2019): 482–95. https://doi.org/10.1177/0739456X19872255.

Currans, Kristina M., Gabriella Abou-Zeid, Chris McCahill, Nicole Iroz-Elardo, Kelly J. Clifton, Susan Handy, and Irene Pineda. "Households with Constrained Off-Street Parking Drive Fewer Miles." *Transportation*, July 29, 2022, 1–26. https://doi.org/10.1007/s11116-022-10306-8.

Diao, M I, and Joseph Ferreira. "Vehicle Miles Traveled and the Built Environment: Evidence from Vehicle Safety Inspection Data." *Environment and Planning A: Economy and Space* 46, no. 12 (December 2014): 2991–3009. https://doi.org/10.1068/a140039p.

Ding, Chuan, Chao Liu, Yi Zhang, Jiawen Yang, and Yunpeng Wang. "Investigating the Impacts of Built Environment on Vehicle Miles Traveled and Energy Consumption: Differences between Commuting and Non-Commuting Trips." *Cities* 68 (August 2017): 25–36. https://doi.org/10.1016/j.cities.2017.05.005.

Ding, Hao, and Michael Manville. "Parking, Travel Behavior, and Working from Home." *Journal of Transport and Land Use* 18, no. 1 (January 23, 2025): 29–55. https://doi.org/10.5198/jtlu.2025.2501.

Ewing, Reid, and Robert Cervero. "Travel and the Built Environment: A Meta-Analysis." *Journal of the American Planning Association* 76 (June 21, 2010): 265–94. https://doi.org/10.1080/01944361003766766.

Hong, Jinhyun, Qing Shen, and Lei Zhang. "How Do Built-Environment Factors Affect Travel Behavior? A Spatial Analysis at Different Geographic Scales." *Transportation* 41, no. 3 (May 2014): 419–40. https://doi.org/10.1007/s11116-013-9462-9.

Howell, Amanda, Kristina Currans, Steven Gehrke, Gregory Norton, and Kelly Clifton. "Transportation Impacts of Affordable Housing: Informing Development Review with Travel Behavior Analysis." *Journal of Transport and Land Use* 11, no. 1 (January 5, 2018). https://doi.org/10.5198/jtlu.2018.1129.

Ihlanfeldt, Keith. "Vehicle Miles Traveled and the Built Environment: New Evidence from Panel Data." *Journal of Transport and Land Use* 13, no. 1 (2020): 23–48. https://doi.org/10.5198/jtlu.2020.1647

Kane, Kevin, and Huixin Zheng. "COVID-Era Built Environment and Travel: Insights from Location-Based Services Data." *Journal of Transport Geography* 126 (June 1, 2025): 104218. https://doi.org/10.1016/j.jtrangeo.2025.104218.

Kim, Jae Hong, Jesus M. Barajas, Nicholas J. Marantz, Douglas Houston, Veronica Herrera, Alex Okashita, and Maxwell B. Cabello. “Assessing the Potential for Densification and VMT Reduction in Areas Without Rail Transit Access,” September 1, 2024. https://doi.org/10.7922/G22V2DGJ.

Kühnel, Fabian, Michael Schrömbges, Nora Braun, and Tobias Kuhnimhof. “Public Transport Accessibility and Its Effect on Mode Choice.” *Urban Science* 9, no. 2 (February 2025): 49. https://doi.org/10.3390/urbansci9020049.

Lee, Sungwon, and Bumsoo Lee. “*Comparing the Impacts of Local Land Use and Urban Spatial Structure on Household VMT and GHG Emissions.” *Journal of Transport Geography* 84 (April 2020): 102694. https://doi.org/10.1016/j.jtrangeo.2020.102694.

Manville, Michael, and Miriam Pinski. “Parking Behaviour: Bundled Parking and Travel Behavior in American Cities.” *Land Use Policy* 91 (February 1, 2020): 103853. https://doi.org/10.1016/j.landusepol.2019.02.012.

Millard-Ball, Adam, Jeremy West, Nazanin Rezaei, and Garima Desai. “What Do Residential Lotteries Show Us about Transportation Choices?” *Urban Studies* 59, no. 2 (February 2022): 434–52. https://doi.org/10.1177/0042098021995139.

Miotti, Marco, Zachary A. Needell, and Rishee K. Jain. “The Impact of Urban Form on Daily Mobility Demand and Energy Use: Evidence from the United States,” June 1, 2023. https://doi.org/10.1016/j.apenergy.2023.120883.

Nachtigall, Florian, Felix Wagner, Peter Berrill, and Felix Creutzig. “Built Environment and Travel: Tackling Non-Linear Residential Self-Selection with Double Machine Learning.” *Transportation Research Part D: Transport and Environment* 140 (March 1, 2025): 104593. https://doi.org/10.1016/j.trd.2025.104593.

Nasri, Arefeh, and Lei Zhang. “Impact of Metropolitan-Level Built Environment on Travel Behavior.” *Transportation Research Record: Journal of the Transportation Research Board* 2323, no. 1 (January 2012): 75–79. https://doi.org/10.3141/2323-09.

Newmark, Gregory, and Peter Haas. “Income, Location Efficiency, and VMT: Affordable Housing as a Climate Strategy.” The California Housing Partnership, December 16, 2015. https://cnt.org/sites/default/files/publications/CNT%20Working%20Paper%20revised%202015-12-18.pdf.

Sabouri, Sadegh, Reid Ewing, and Hannaneh Abdollahzadeh Kalantari. “Estimating Transit’s Land-Use Multiplier: Direct and Indirect Effects on Vehicle Miles Traveled.” *Transportation*, October 17, 2024. https://doi.org/10.1007/s11116-024-10542-0.

Salon, Deborah. “Heterogeneity in the Relationship between the Built Environment and Driving: Focus on Neighborhood Type and Travel Purpose.” *Research in Transportation Economics*, Sustainable Transportation, 52 (October 1, 2015): 34–45. https://doi.org/10.1016/j.retrec.2015.10.008.

Salon, Deborah, Marlon G. Boarnet, Susan Handy, Steven Spears, and Gil Tal. “How Do Local Actions Affect VMT? A Critical Review of the Empirical Evidence.” *Transportation Research Part D: Transport and Environment* 17, no. 7 (October 2012): 495–508. https://doi.org/10.1016/j.trd.2012.05.006.

Tao, Tao, and Haotian Zhong. “Income Moderates the Nonlinear Influence of Built Environment Attributes on Travel-Related Carbon Emissions.” *Journal of Transport Geography* 120, no. C (2024). https://ideas.repec.org//a/eee/jotrge/v120y2024ics0966692324001947.html.

Yu, Anton, and Christopher D. Higgins. “Travel Behaviour and the 15-Min City: Access Intensity, Sufficiency, and Non-Work Car Use in Toronto.” *Travel Behaviour and Society* 36 (July 1, 2024): 100786. https://doi.org/10.1016/j.tbs.2024.100786.

Zuk, Miriam, Karen Chapple, Daniel G. Chatman, Elizabeth Deakin, and Carol Galante. “Assessing the Travel Demand and Co-Benefit Impacts of Affordable Transit Oriented Developments,” 2019. https://ww2.arb.ca.gov/sites/default/files/classic/research/apr/past/16rd003-1.pdf

Appendices

Appendix A: Tables

Table A.1: Summary of Baseline VMT Equations from AHSC Quantitative Methodology

Name	Equation	Notes
1: Average Daily Trips per Dwelling Unit	$\frac{(Weekday\ Trips * 5) + Saturday\ Trips + Sunday\ Trips}{7\ days}$	From ITE Trip Generation Manual data (unclear if state or national average)
2: Primary Trip Length	$= (H\text{-}W\ Length * H\text{-}W\ Share) + (H\text{-}S\ Length * H\text{-}S\ Share) + (H\text{-}O\ Length * H\text{-}O\ Share)$	H-W: Home to work H-S: Home to shopping H-O: Home to other Shares of each trip type are state averages, lengths of each trip type are county averages, reported by MPOs or generated using Caltrans Travel demand Models
3: Overall Trip Length	$= (Primary\ Trip\ Length * Primary\ Share) + (Primary\ Trip\ Length * 25\% * Diverted\ Share) + (0.1\ miles * Passby\ miles\ Share)$	Breaks down primary trip length into direct trips, diverted and pass by trips. The 25% of distance assigned is based on an assumption not sourced.
4: Annual Unmitigated VMT	$Average\ Daily\ Trips * Overall\ Trip\ Length * Total\ Units * 365\ days$	

Table A.2: Summary of VMT Reduction Equations from AHSC Quantitative Methodology

Name	Equation	Notes
5: VMT Reductions from Increased Density	$\left(\frac{Density - Required\ Density}{Required\ Density}\right) * 7\%$	The 7% elasticity is the most conservative estimate reported in Boarnet, Marlon and Handy, Susan (2010)
6: Land Use Index	$-\left(\frac{4 * (0.01 * ln0.01) + \left(\left(\frac{RS}{RS + MS}\right) * ln\left(\frac{RS}{RS + MS}\right)\right) + \left(\left(\frac{MS}{RS + MS}\right) * ln\left(\frac{MS}{RS + MS}\right)\right)}{ln6}\right)$	Index is applied to the share of the project area devoted to residential and public uses
7: VMT Reductions from Increased Land Use Diversity	$\left(\frac{Land\ Use\ Index - 0.15}{0.15}\right) * 9\%$	0.15 is assumed land use index for a single-use building The 9% elasticity is derived from the weighted average of 10 studies in Eweing, Cervero (2010)
8: VMT Reductions from Increased Destination Accessibility	$\left(\frac{housing\ development\ employment\ density - county\ employment\ density}{county\ employment\ density}\right) * 7\%$	The same elasticity as was used for residential density. CAPCOA uses the same calculation for residential and employment density.
9: VMT Reductions from Integration of Affordable Housing	$\left(\frac{Affordable\ Units}{Total\ Units}\right) * 4\%$	The 4% elasticity comes from Nelson/Nygaard (2005) report
10: VMT Reductions from Land Use Measures	$= 1 - (1 - Density\ VMT\ Reductions) * (1 - Diversity\ VMT\ Reductions) * (1 - Accessibility\ VMT\ Reductions) * (1 - Affordability\ VMT\ Reductions)$	Sums all Land Use reduction areas
11: VMT Reductions from Limited Parking Supply	$\left(\frac{Total\ Units * Parking\ Rate - Parking\ Space}{Total\ Units * Parking\ Rate}\right) * 50\%$	The 50% elasticity(?) comes from Nelson/Nygaard (2005) report

Name	Equation	Notes
12: VMT Reductions from Unbundled Parking Cost	$Unbundled\ Cost * \left(\frac{12\ months}{\$4{,}000}\right) * 0.4 * 85\%$	The 40% elasticity comes from Victoria Transport Policy Institute (2009) The 85% adjustment factor is based on CAPCOA figures (p.C3 of Appendix C): A reduction of 1 vehicle leads to an 0.85 reduction in vehicle trips
13: Total VMT Reductions from Parking Measures	$= 1 - (1 - Parking\ Supply\ VMT\ Reductions) * (1 - Unbundled\ Parking\ VMT\ Reductions)$	Sums all Parking reduction areas
14: VMT Reductions from Traffic Calming Measures	1%	Based on Cambridge Systematics & SMAQMD reports
15: VMT Reductions from Residential Transit Subsidy	$Elasticity * \frac{Recipients}{Total\ Units} * \frac{Duration}{30\ years}$	Variable elasticity is based on Nelson/Nygaard (2010)
16: Total VMT Reductions	$= Land\ Use\ VMT\ Reductions + Parking\ VMT\ Reductions + Traffic\ Calming\ VMT\ Reductions + Subsidy\ VMT\ Reductions$	
17: Annual Avoided VMT	$Annual\ Unmitigated\ VMT * Total\ VMT\ Reductions$	
18: Total Avoided VMT	$Annnual\ Avoided\ VMT * 30\ years$	Assumed lifetime of the project is 30 years

Table A.3: Descriptive statistics for all explanatory built environment variables constructed

Variable	Mean	Std Dev	Min	Q1	Q3	Max	# Missing
Car_DistToRail	42.24	56.53	0.13	2.82	74.02	299.36	230
Walk_DistToRail	44.19	59.49	0.00	2.84	75.22	301.66	50
PopDens_bg	6066.60	7526.69	0.12	1229.43	8087.22	162383.36	0.00
PopDens_half	5666.86	5945.87	0.12	1238.64	7980.18	75560.13	0.00
PopDens_one	5115.03	5007.86	0.12	1438.51	7124.86	47857.86	0.00

Variable	Mean	Std Dev	Min	Q1	Q3	Max	# Missing
PopDens_two	4375.00	4276.86	0.12	1344.63	6001.29	36572.19	0.00
EmpDens_bg	1932.21	8103.02	0.00	127.66	1550.49	491067.35	0.00
EmpDens_half	2126.05	7256.54	0.00	153.42	1939.71	291543.80	0.00
EmpDens_one	2238.96	6744.09	0.00	233.86	2287.90	220872.11	0.00
EmpDens_two	2119.40	4883.85	0.02	334.39	2352.39	89081.51	0.00
RetDens_bg	194.54	691.48	0.00	0.00	128.15	52331.43	0.00
RetDens_half	209.84	491.41	0.00	1.71	219.88	16745.93	0.00
RetDens_one	216.53	399.39	0.00	11.29	273.15	8958.46	0.00
RetDens_two	202.89	310.02	0.00	30.49	278.70	4829.17	0.00
EntDens_bg	292.45	1308.67	0.00	0.00	185.71	78771.73	0.00
EntDens_half	316.58	1096.98	0.00	2.84	263.12	34245.13	0.00
EntDens_one	318.40	913.08	0.00	17.34	314.53	21246.17	0.00
EntDens_two	283.27	660.76	0.00	38.64	302.47	12374.58	0.00
OffDens_bg	197.46	1714.17	0.00	0.07	67.77	92213.77	0.00
OffDens_half	232.48	1670.24	0.00	2.03	103.19	72015.29	0.00
OffDens_one	254.28	1559.30	0.00	6.20	140.44	60784.35	0.00
OffDens_two	241.14	982.57	0.00	12.83	166.50	21194.45	0.00
IndDens_bg	227.88	1148.23	0.00	7.71	129.98	76937.91	0.00
IndDens_half	256.63	1033.53	0.00	15.05	173.81	37120.27	0.00
IndDens_one	294.12	861.86	0.00	26.80	260.37	24814.58	0.00
IndDens_two	323.74	652.46	0.00	43.65	346.20	9652.91	0.00
SvcDens_bg	359.95	2119.35	0.00	7.14	179.31	160407.59	0.00
SvcDens_half	406.35	2380.17	0.00	12.02	238.06	123085.21	0.00
SvcDens_one	435.47	2277.65	0.00	22.17	314.30	94996.45	0.00
SvcDens_two	415.16	1490.93	0.00	31.28	353.94	34065.08	0.00
EduDens_bg	185.26	2664.27	0.00	0.00	99.61	213091.01	0.00
EduDens_half	190.29	871.29	0.00	0.00	168.26	54049.72	0.00
EduDens_one	199.33	492.82	0.00	15.37	207.25	9782.26	0.00
EduDens_two	188.34	337.57	0.00	32.33	222.05	4547.89	0.00
HlthDens_bg	380.97	2447.43	0.00	12.66	242.50	208733.04	0.00

Variable	Mean	Std Dev	Min	Q1	Q3	Max	# Missing
HlthDens_half	387.83	1194.31	0.00	15.81	312.38	23727.05	0.00
HlthDens_one	384.09	818.39	0.00	25.89	385.01	11845.55	0.00
HlthDens_two	348.59	593.18	0.00	41.42	428.76	7170.69	0.00
PubDens_bg	93.71	1763.89	0.00	0.00	0.00	161256.08	0.00
PubDens_half	126.06	1753.69	0.00	0.00	0.00	128160.16	0.00
PubDens_one	136.73	1128.49	0.00	0.00	17.95	38829.52	0.00
PubDens_two	116.27	528.95	0.00	0.00	53.00	8740.05	0.00
LowWage_bg	450.32	1270.86	0.00	35.02	445.47	47288.42	0.00
LowWage_half	482.63	1145.62	0.00	41.53	534.48	42362.91	0.00
LowWage_one	485.63	962.80	0.00	65.12	584.19	20642.11	0.00
LowWage_two	442.30	711.96	0.00	87.88	555.78	11097.99	0.00
MedWage_bg	594.58	1596.33	0.00	41.24	561.47	75047.61	0.00
MedWage_half	635.49	1441.24	0.00	52.94	699.35	48606.52	0.00
MedWage_one	650.65	1288.81	0.00	84.37	791.49	31842.37	0.00
MedWage_two	609.53	977.81	0.00	119.80	778.57	16094.73	0.00
HiWage_bg	887.30	5906.35	0.00	28.32	467.50	418882.32	0.00
HiWage_half	1007.93	5105.33	0.00	41.59	616.93	211476.86	0.00
HiWage_one	1102.67	4714.34	0.00	72.07	846.31	168387.63	0.00
HiWage_two	1067.57	3302.20	0.00	106.54	984.57	62061.71	0.00
StreetDens_GF18_bg	19.30	11.37	0.00	9.71	26.80	84.56	0.00
StreetDens_GF18_half	19.10	10.69	0.00	9.82	26.83	63.76	0.00
StreetDens_GF18_one	18.41	9.74	0.00	10.46	25.92	49.47	0.00
StreetDens_GF18_two	16.82	8.84	0.00	9.65	23.76	43.38	0.00
StreetDens_OX24_bg	12.16	7.20	0.01	6.03	17.51	48.89	203.00
StreetDens_OX24_half	11.95	6.38	0.01	6.60	16.84	35.15	75.00

Variable	Mean	Std Dev	Min	Q1	Q3	Max	# Missing
StreetDens_OX24_one	11.39	5.58	0.01	7.17	15.66	26.55	18.00
StreetDens_OX24_two_updated	10.23	4.84	0.09	6.64	13.94	21.37	15.00
IntersectionDens_OX24_bg	102.80	75.63	0.03	38.05	152.83	748.46	278.00
IntersectionDens_OX24_half	100.33	66.94	0.03	40.32	149.85	459.35	124.00
IntersectionDens_OX24_one	93.95	58.18	0.03	45.86	136.91	339.08	54.00
IntersectionDens_OX24_two	81.74	49.77	0.03	41.55	119.37	237.96	14.00
NetLoadDens_GF18_bg	332.56	407.46	0.00	148.74	393.67	15158.04	0.00
NetLoadDens_GF18_half	322.25	332.31	0.00	153.47	397.94	9825.82	0.00
NetLoadDens_GF18_one	314.35	304.15	0.00	169.34	385.09	6638.62	0.00
NetLoadDens_GF18_two	303.03	256.11	0.00	175.12	375.00	3683.39	0.00
NetLoadDens_OX24_bg	767.37	3212.39	0.67	213.17	638.12	288685.05	203.00
NetLoadDens_OX24_half	685.16	6632.16	0.67	220.11	660.56	695210.67	75.00
NetLoadDens_OX24_one	571.75	1541.41	0.67	244.12	651.16	58911.07	18.00
NetLoadDens_OX24_two_updated	516.58	491.64	0.67	256.31	641.59	7028.50	15.00
JobsWithin45_Car_2	76305.26	98141.01	12.00	11252	105027	684461	0.00
JobsWithin45_Transit_3	54124.48	125776.89	0.00	0.00	40170	1166204	0.00
HHBusStopCt_1mi_new	31.81	57.44	0.00	0.00	41.00	722.00	0.00

Variable	Mean	Std Dev	Min	Q1	Q3	Max	# Missing
HHBusStopCt_halfmi_new	8.81	16.32	0.00	0.00	12.00	232.00	0.00
HHBusStopCt_3eighthmi_new	5.10	9.67	0.00	0.00	7.00	120.00	0.00
HHBusStopCt_quartmi_new	2.30	4.65	0.00	0.00	3.00	64.00	0.00
HHBusStopDens_1mi_new	10.13	18.28	0.00	0.00	13.05	229.82	0.00
HHBusStopDens_halfmi_new	11.22	20.78	0.00	0.00	15.28	295.39	0.00
HHBusStopDens_3eighthmi_new	11.53	21.89	0.00	0.00	15.84	271.62	0.00
HHBusStopDens_quartmi_new	11.72	23.71	0.00	0.00	15.28	325.95	0.00
BGBusStopCt_bg_new	2.90	4.92	0.00	0.00	4.00	133.00	0.00
BGBusStopDens_2mi_new	9.59	17.38	0.00	0.00	12.40	190.95	0.00
BGBusStopDens_1mi_new	10.84	19.67	0.00	0.00	14.03	244.11	0.00
BGBusStopDens_halfmi_new	11.37	21.68	0.00	0.00	14.59	307.19	0.00
BGBusStopDens_bg_new	11.40	25.31	0.00	0.00	13.05	514.74	0.00

Appendix B: Data Processing

Section B.1: NHTS data discrepancies

In calculating our daily VMT per survey respondent, we calculated the VMT per person for each trip (dependent on the number of passengers) and then matched that VMT to the person ID of those survey respondents who were on the trip. However, in attempting to join the VMT values to survey respondents, we came across discrepancies in the listed person IDs from the trip and person datasets. It seemed that trips were not accurately joining to the correct person ID, and further exploration revealed that the number of people as well as the number of trips identified in the dataset did not match the values listed in the documentation. This issue occurred specifically within the California add-on to the 2017 NHTS, and we were able to remedy this issue by joining in data from the national version of the dataset.

Section B.2: Network distance to rail

One of the main built environment variables of interest in Task 2 is the distance to rail measure, which captures the proximity of each household to the closest rail stop. This variable can be calculated in two ways; either a straight line distance from the given household to the closest rail stop, or the distance along the existing street or pedestrian network. Given that the latter method is a more accurate representation of real-world conditions under which an individual would travel, we chose to calculate the variable this way. Though these infrastructural networks do not change significantly over time, we wanted to capture the conditions of the built environment at the time the NHTS data was collected, and therefore sought to use datasets constructed in 2017.

Table B.1: Pedestrian network distance to the nearest transit station (La Mesa, CA)

To calculate distance to rail along a network, we began by determining the appropriate datasets for our analysis. The rail station data came from previous research conducted by the authors. We chose to filter out any stations that were exclusively operated by Amtrak, as these stations are intended primarily for long distance travel and not often used for commute or other daily travel.

The main challenge in data procurement was identifying the proper network along which to do routing calculations from the identified household to the identified stops. We began by calculating driving distance with the 2017 road network TIGER/Line shapefile. However, we also wanted to represent walking distance as well, using a pedestrian network dataset from 2017. Identifying clean, comprehensive datasets for historical pedestrian networks proved difficult. We identified an online source (Geofabrik) that provided archived downloads of OpenStreetMap data, saved as a PBF (protocolbuffer binary format) file for the entire State. The walk network data was queried and extracted from the PBF file using the Python package pyrosm, with the network defined according to the package's preset filter for pedestrian networks.

Once the data was collected and processed, we used the Network Analyst extension supported by ArcMap 10.8 to conduct routing analysis. Using the OD cost matrix tool, we set the locations of households as our origins, the locations of the rail stops as our destinations, and the two respective networks as our network upon which to route. We then used the matrix to identify the closest destination to each respective origin and calculate the distance along the network. If an origin or destination was not located directly on the network, it was snapped to the closest point on the network, and the distance by which the point moved to snap was accounted for in the distance calculation.

When conducting this analysis for the pedestrian network, we found that many households did not connect at all to a rail station. This is due to the disconnected nature of pedestrian networks, and the fact that these models do not account for the fact that a pedestrian may be able to take a route, even if no predefined path exists. For instance, a number of households in one neighborhood were unable to be routed to their closest rail stop, given that the main road out of the neighborhood was not defined as walkable (though it is likely that these individuals could in fact walk out of their neighborhood anyways). Many other households were snapped to a small, disconnected portion of sidewalk directly outside their homes. Removing these disconnected portions of the pedestrian network proved to be challenging, as keeping only the largest connected portion would remove many significant pedestrian paths, and determining a threshold for which to remove or keep a portion of the pedestrian portion was difficult to identify.

In order to ensure that all households were routed to a transit station, each disconnected household was manually inspected and corrected. If households were snapped to a small disconnected portion of the network, their point was shifted slightly in order to snap to the closest connected portion instead. Most points only had to be shifted a few hundred feet at the maximum. All others that were significantly far from the connected network, or in a disconnected neighborhood as described before, were identified as being far enough from any transit stop to be of interest in our regression (greater than 2 miles from the closest rail stop). Therefore, these were left as is and marked as being outside any radius of interest for distance to rail.

Section B.3: Cleaning bus stop data

To develop the bus stop density variables for Task 2, we first needed to access geospatial data for all of the in-use bus stops in California. Bus stops change frequently, depending on demand and the resources of the transit agency responsible for those stops. Given that we were using the 2017 NHTS responses, current bus stop data may not reflect the built environment at the time of the survey and therefore we sought to use historical bus stop data.

While the California Open Data Portal does provide a statewide dataset for current transit stops, there exists no archive of datasets from previous years. We struggled to find any comprehensive and credible source whatsoever for this type of data.

Though we were unable to find any historical bus stop datasets online, our research team had geospatial data for bus stops in 2017 saved from a previous project. The data had been downloaded from each transit agency (for Bay Area bus stops), municipality, or region (for bus stops elsewhere) with bus service in operation in California at the time of interest. For a full list of all transit agencies and cities for which data was available, see Table B.3. Once the datasets were identified, we then worked to clean and combine all datasets into one comprehensive, statewide dataset.

As we began the task of cleaning and combining, we encountered a number of issues with the lack of standardization among the datasets. Each city or transit agency documented different attributes, whether that be stop name, bus routes served, frequency, etc. Some datasets included both in-use and not in-use stops, which had to be filtered out accordingly. In addition, many contained attributes that had incomplete or non-standardized entries. In our final comprehensive dataset, the only non-spatial information we were able to include for all bus stops was some form of a stop ID and the agency/agencies who operated the bus stop.

Most importantly for our bus stop density calculations, we saw significant variation in how stops were geocoded, both between datasets and within datasets themselves. Given that this dataset was being used to calculate the spatial density of bus stops within a given area, this was the point of most concern when considering how to clean the data. We observed the following differences in geocoding methods among datasets:

- **Bus stops opposite one another**: In the case where two bus stops are located opposite one another–serving the same route in opposite directions–different datasets geocoded these stops differently. Some represented two stops with one point, while others represented these with two. On occasion, some datasets would have one point for locations where there was truly only one stop, as well as one point for where bus stops were exactly opposite one another, making it very difficult to tell which points were only one bus stop while others were two. A sample of stops from each dataset were manually checked to confirm how stops were represented for that specific dataset.

- **Stops served by multiple routes:** In the case where one bus stop was served by multiple routes, different datasets also chose to geocode these differently. Some datasets geocoded one stop for each route, meaning that if one stop served multiple routes, this would be represented by multiple points either geocoded identically, or clustered together. Other datasets would represent this stop with a single point, and an attribute or attributes describing the routes which served that stop. We chose to represent the points by stop rather than number of routes; as we did not have reliable data for all of the datasets on which exact routes served each stop. In addition, routes change much more frequently than physical bus stops themselves, and so we could not be certain that the routes reported were accurate.
- **Stops at major transit stations:** In the case where a transit agency served a larger transit hub or rail station, we also saw variance in how this was geocoded. Some datasets geocoded a single bus stop for all routes that stopped at the transit station, while others geocoded multiple points. The difference in this issue from the previous is that these points were often more further spaced than when an agency had multiple points (representing one stop with multiple routes) on a single street corner. Therefore, each transit station had to be identified and stops were manually merged. If multiple datasets showed stops at the same transit station, then each dataset would have two points representing all stops at that station–one point per route direction.
- **Shared bus stops (Bay Area):** In the case of transit agencies within the Bay Area, some agencies shared bus stops in common. Sometimes they would be geocoded in exactly the same location, while other times they would be slightly offset. A sample of these stops were manually checked to confirm whether one stop was shared by multiple agencies, or if there were indeed two separate stops.
- **Other geocoding concerns:** There were other concerns with the accuracy or cleanness of some of the spatial data. Some geocoding was more spatially accurate than others.

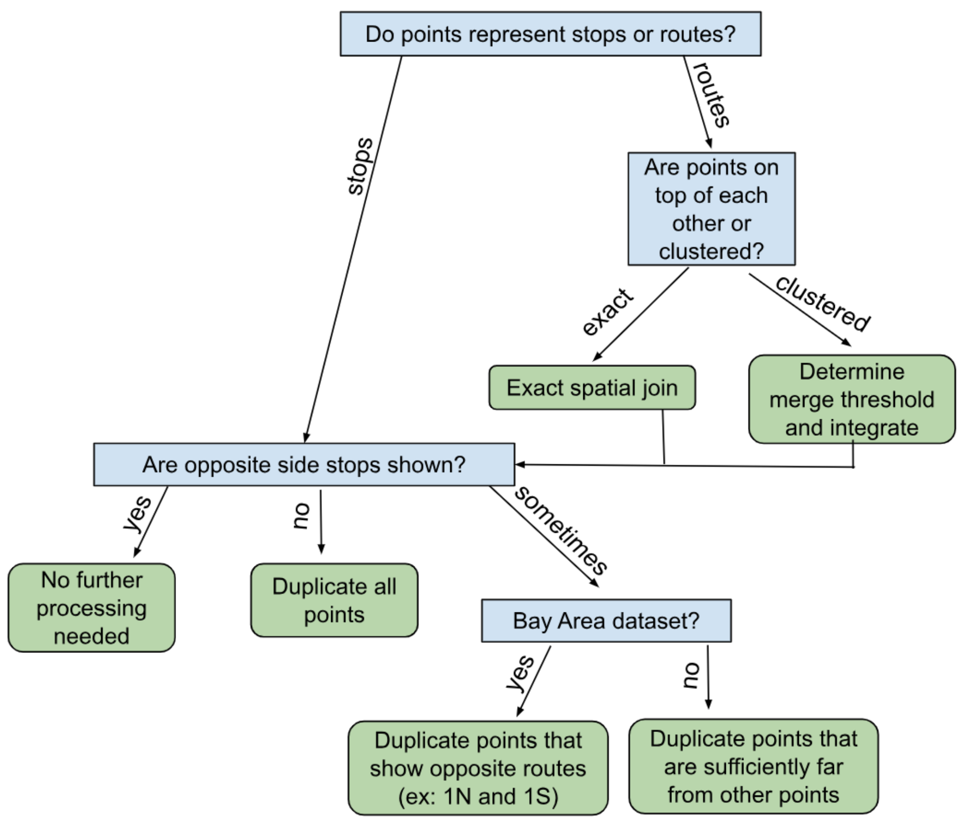
We took a number of steps to clean and standardize the data across the state. Due to the lack of reliable route data for all datasets, we chose to define each datapoint as a physical stop, rather than a route going through a stop. We also chose to define opposite-side bus stops as two unique bus stops. In addition, if only one bus stop existed but served both directions we chose to indicate this with two points, since this was a better indicator of directional mobility and made it easier to process the datasets. When two agencies shared a bus stop, we geocoded one stop per agency. Our final standardized dataset includes the bus stop name (if one was given), stop ID (according to how the original dataset chose to ID stops), and the name of the agency or city to which that stop belongs (as identified by the original dataset). See Table B.2 for how each geocoding concern was identified and addressed. See Figure B.1 for the order in which concerns were addressed.

Table B.2: Geocoding concerns and approaches

Geocoding Concern	Identification Methods	Approach
Bus stops opposite one another are geocoded as one stop	- Stop lists routes in opposite directions served (ex: 1N and 1S) - GoogleMaps Street View confirms two stops where one stop is geocoded - In the case where two agencies share a stop, one agency shows two stops and the other only shows one	- If it is determined that most/all points are actually opposite side stops, duplicate all points - If it is determined that some points are actually opposite side stops, duplicate conditionally (based on dist from other stops or route directions served)
Routes are geocoded rather than stops	- Multiple points geocoded in the same spot, or clustered together - Multiple data points with the same stop ID but varying route IDs	- If points are geocoded in the exact same place, do an exact spatial merge - If points are clustered, determine an integration threshold and apply it to all points
Multiple agencies sharing one stop (Bay Area)	- GoogleMaps Street View confirms whether agencies share a stop or two unique stops exist - StopID is the same for both datasets	- No action necessary; however shared stops among points did help us identify the above concerns
Other issues	- Incorrectly geocoded points - Points with no geometry - Points marked not in-use	- Any points with the identified concerns were removed from the dataset - This step was completed prior to any other necessary cleaning

In the case of Bay Area bus stops, route data was actually significantly more standardized across different agencies. Therefore, in order to determine opposite-side stop information, we relied on the routes listed for a given bus stop. If the same route as indicated, with opposite directions (ex: 1N and 1S) for the same bus stop, then this point was automatically duplicated, as indicated in Figure B.1. Before processing any points in the Bay Area dataset, we first dissolved all points based on exact spatial coordinates.

Figure B.1: Dataset processing flowchart



Once the concerns were identified and the best approach was determined for each dataset, it was processed according to the methods described above. For an overview of the concerns identified and processing methods applied to each individual dataset, see Table B.3. Small concerns unrelated to the above description are noted with an asterisk. For larger concerns unrelated to the above description, these are described in more detail below.

Table B.3: Bus stop count by dataset, before and after processing

City/Agency Name	Number of stops (before)	Geocoding Concerns	Processing method	Number of stops (after)
California municipalities/regions (excluding the Bay Area)				
Amador	86	-Clustered points representing routes for one stop -No opp side stops shown	-Integrated (40 ft) -Basic duplication	158
Butte	542	*3 exact duplicates shown	*Removed 3 duplicates	539
Del Norte	118	-Some opp side stops shown	- Conditional duplication (6500 ft)	125
Fresno	2166	-Clustered points representing one stop -No opposite side stops shown *3 stops with no geometry	- Integrated (100 ft) - Basic duplication *Removed 3 stops	1694
Kern	238	-Some opposite side stops shown *Only one stop at transit station shown	- Conditional duplication (9800 ft) *Transit station stop duplicated	248
Lassen	96	N/A	N/A	96
Los Angeles	20013	*Dataset has reliable ID information	*Dissolved by ID	14715
Los Angeles (BRT)	81	*Dataset has reliable ID information	*Dissolved by ID	74
Madera	52	N/A	N/A	52
Marin	551	*2017 dataset proved too unreliable	*Had to replace with current 2024 data	551

City/Agency Name	Number of stops (before)	Geocoding Concerns	Processing method	Number of stops (after)
Mendocino	212	-Some opposite side stops shown	- Conditional duplication (13100 ft)	225
Merced	360	N/A	N/A	360
Mono, Inyo, Eastern Sierra	152	-Some opposite side stops shown	- Conditional duplication (9800 ft)	169
Monterey	1440	*Included not-in-use stops	*Removed 113 not-in-use stops	1308
Nevada	223	N/A	N/A	223
Sacramento, Sutter, Yuba, Yolo, & Placer	5093	*Many issues; as described below	*Dissolved by stop name & Census-designated place	4784
San Benito	72	-Clustered points representing one stop -Opposite side stops not shown	-Integrated (70ft) -Basic duplication	82
San Diego	3340	-No opposite side stops shown	-Basic duplication	6680
San Joaquin	1042	N/A	N/A	1042
San Luis Obispo	504	N/A	N/A	504
Santa Barbara	2832	*Dataset has reliable NODEID information -No opposite side stops shown	*Dissolved by NODEID -Basic duplication	1696
Santa Cruz	1025	*2 stops without geometry	*Removed 2 stops	1023
Stanislaus	337	-Some opposite side stops shown	- Conditional duplication (13100 ft)	342

City/Agency Name	Number of stops (before)	Geocoding Concerns	Processing method	Number of stops (after)
Tehama	90	-Clustered points representing one stop -Opposite side stops not shown	-Integrated (150 ft) -Basic duplication	140
Tulare	283	*Dataset has reliable OBJECTID information -Opposite side stops not shown	*Dissolved by OBJECTID -Basic duplication	458
Bay Area Transit Agencies				
AC TransBay	1327	N/A	N/A	1327
AC Transit	5685	N/A	N/A	5685
ACE	8	-Opposite side stops not shown	-Basic duplication	16
AirBART	2	N/A	N/A	2
American Canyon Transit	19		-Basic duplication	19
Benicia Transit	53		-Integrated (80ft) -Basic duplication	90
Dumbarton Express	104	N/A	N/A	104
Emery Go-Round	32		-Conditional multiplication	42
Fairfield-Suisun Transit	309		-Basic duplication	618
Golden Gate Transit	887	N/A	N/A	887
Napa Downtown Trolley	15		-Basic duplication	30

City/Agency Name	Number of stops (before)	Geocoding Concerns	Processing method	Number of stops (after)
Napa VINE	345		-Integrated (80 ft)	249
Petaluma Transit	54		-Integrated (80 ft) -Basic duplication	88
Rio Vista Delta Breeze	29		-Conditional multiplication	49
samTrans	2230	N/A	N/A	2230
San Francisco MUNI	3827	N/A	N/A	3827
Santa Clara VTA	3661	N/A	N/A	3661
Santa Rosa CityBus	462		-Integrated (80 ft) -Basic duplication	732
Sonoma County Transit	732		-Integrated (80 ft) -Conditional multiplication	1188
St. Helena VINE	19		-Conditional multiplication	19
Stanford Marguerite Shuttle	136	N/A	N/A	136
The County Connection	1176		-Integrated (80 ft) -Conditional multiplication	1931
TriDelta Transit	490		-Integrated (70 ft) -Conditional multiplication	860
Union City Transit	128		-Integrated (80 ft) -Basic duplication	240
Vacaville City Coach	163		-Integrated (80 ft) -Basic duplication	254
Vallejo Transit	254		-Integrated (70 ft)	479

City/Agency Name	Number of stops (before)	Geocoding Concerns	Processing method	Number of stops (after)
			-Conditional multiplication	
WestCAT	208		-Integrated (70 ft)	176
WHEELS	371		-Integrated (70 ft) -Conditional multiplication	645
Yountville Shuttle	18	N/A	N/A	18

Some datasets required additional manual processing beyond the methods described above. The dataset for bus stops in Sacramento, Sutter, Yuba, Yolo, & Placer was very messy and it was difficult to differentiate spatially between duplicated points representing one stop and two adjacent points actually representing two stops. Stop names given by cross streets were included in the attributes and appeared to properly differentiate between actual physical stops, though there were some stops with the same street names given they were located in different jurisdictions within the region (eg, 5th Ave and Q St could denote two separate stops located in two different cities). Therefore, we dissolved by name and the city in which the stop was located.

Identifying issues within the dataset and correcting them to develop an accurate bus density metric was a significant, time intensive task, and though much thought was put into the exact processing methods it is likely that the resulting dataset is still not entirely reflective of real world bus stop locations. These issues and concerns could be improved on in the future through proper documentation of bus stops by transit agencies, and standardized reporting procedures for all agencies as set by the State. Challenges with this may include a lack of staff capacity in these transit agencies to properly document transit stops. Thankfully within the past few years, the State has begun a more concerted effort to collect accurate and detailed transit stop data using the General Transit Feed Specification (GTFS) format, and the issues described here should be ameliorated as the quality and comprehensiveness of this data continues to improve. However, accessing and utilizing historical data on bus stops still poses a significant problem for researchers.

Section B.4: Network identification & density calculation

To evaluate the density of the street network within a given area, an appropriate network dataset had to be identified. We began by using the OSMnx Python package to query and extract the street network via the OverPass API from OpenStreetMap. The network was defined according to the package's preset filter for street networks. Once the network was extracted the edges and nodes were saved as separate shapefiles, which were then used to calculate street and intersection densities, respectively.

The initial network dataset posed a few challenges for analysis. First, the package's filter excluded private streets, which proved to be a problem as certain block groups were then calculated to have a street or intersection density of 0, indicating that there were no streets in the block group. For a number of these block groups, adjacent block groups had much higher densities which indicated that there was an issue with how the network was defined. We manually inspected a number of these block groups and noted that many streets within private gated communities or on otherwise private property were not included, though still technically accessible via private automobile. Second, we hoped to use 2017 network data for similar reasons as noted in Appendix Section B.2: Network distance to rail and were unable to access this historical data via OSMnx. We were able to download historical OpenStreetMap data as a PBF file from the website Geofabrik and then used the Python package pyrosm to extract the street network. We ensured that private streets were included in our query, and after re-running our street density calculations found that all block groups had non-zero street density.

Further inspection of the node shapefile revealed that nodes were not defined as intersections for the Geofabrik, but rather the connections between linear portions of street segments. For instance, a curved portion of street would be represented as 20 linear edges, connected by 19 nodes. Therefore, the node file was not usable in our calculation of intersection density.

Rather than try and develop a new node file for intersection density calculated on the Geofabrik network, we continued to use our original intersection density calculation.

Appendix C: VMT Estimates Workbook

This is an Excel Workbook that is attached to the final report. The Excel workbook provides example calculations of VMT estimates by block group, throughout the state. Note that this is a sample Workbook and the calculations made here should not be applied as they exist currently. Model outputs and variable coefficients must be updated.

Appendix D: Python Scripts for Data Processing

The folder contains all the code and information on how we (1) Generated all relevant built environment metrics for each Census block group in California, including those at aggregated spatial levels and (2) Used this cleaned dataset to format the data for use in Appendix C. The script used to generate built environment metrics can be updated to include current day data, which can then be used to generate new VMT estimates.

To recalculate built environment metrics for each block group and find the subsequent new VMT estimates, scripts can be run in the order as seen below. Read below for more information on the data, data scripts, and outputs included here, as well as how these files can be adapted for future updates to the built environment metrics.

Script Order:

- **0A - OSMnx Block Group Network Densities.ipynb**
- **0B - GeoFabrik Block Group Street Density.ipynb**
- **1 - Aggregated Built Environment Characteristics (SLD).ipynb**
- **2 - Joining Aggregated Network Characteristics (OSMnx).ipynb**
- **3 - Joining Aggregated Network Characteristics (Geofabrik).ipynb**
- **4 - Joining Aggregated Bus Densities + Final File.ipynb**
- **5 - Built Environment Data Formatting for Appendix C.ipynb**

Data

This folder contains all spatial and tabular datasets used to generate built environment metrics and VMT estimates by Census block group. It includes base geographies, raw input data, distance matrices, and intermediate shapefiles. Many of the raw files reflect 2017 data and should be updated with the most current sources before rerunning the analysis. Refer to the notes within each description for details on appropriate data sources and update procedures.

Subfolder - geographies

This subfolder contains all necessary geographic data upon which to wrangle spatial data and join it to the correct geographies. Note that block group geographies should be updated to reflect new Census-designated boundaries. These change with every decennial Census.

tl_2017_us_CA.shp

This contains the polygon geography for the entire state. It is used to clip street network data downloaded from GeoFabrik to the state alone; as the download contains data for other western states as well. See the network_density subfolder in data_collection for more details.

tl_2017_06_bg.shp
This contains the polygon geographies for every block group in California and includes the land and water area of each block group. This data was sourced from TIGER/Line Shapefile data.

2017_bg_centroids.shp
This contains the latitude and longitude points for all GEOIDs in California. These were computed from tl_2017_06_bg.shp in ArcGIS Pro.

Subfolder – networks

This subfolder contains the files used to calculate street network density, both via OSMnx and Geofabrik.

2017CBG_streetnet.pkl
This is an intermediate output file that contains the network geometries for all block groups in California, using OSMnx. It can be used to calculate street and intersection densities by block group.

2017CBG_GDF_tot.shp
This is the data file that contains all the street and intersection densities by block group, based on the 2024 OpenStreetMap networks pulled using OSMnx; see the network_density subfolder in data_collection for more details. These can be updated using code in this subfolder.

us-west-180101.osm.pbf
This is the raw data downloaded from Geofabrik, in protocol binary format. This is historical data, archived from OpenStreetMap on Jan 1, 2018. It characterizes the street network around the time that the NHTS survey observations were collected, and was used in our regressions. Note that the intersection data from this source is not reliable, and therefore we do not include it here. See the network_density subfolder in data_collection for more details.

rds_byBG1.shp

This is the data file that contains street density by block group, based on the 2018 OpenStreetMap networks pulled using Geofabrik. Once data was extracted from the pbf described above, this data was pulled into ArcGIS Pro to break the statewide network down by block group. This is the resulting file from that process.

Raw data

Raw data included here should be updated to reflect the current built environment data. Each file description describes from what sources these metrics should be updated from.

CaliforniaSLD.shp
This is the raw data file for all data collected from the 2017 EPA Smart Location Database. Note that this data will likely not be updated soon, and as such metrics derived from this source will have to be collected elsewhere. Population and employment densities can be sources from the U.S. Census, American Community Survey 5-year estimates, and Longitudinal Employment Household Dynamics data. Jobs accessibility metrics can be computed using GTFS data and the Google TravelTime API.

BG_BusDens_V2.shp
This is the data file that contains the bus stop densities by Block Group, calculated in ArcGIS Pro. See **Appendix B.4** for more details on how the bus stop data was cleaned to generate densities. These can be updated with statewide GTFS stop data, which can be processed spatially in a similar format as described in **Appendix B.4**.

Block group distances

bg2bg_halfmi.json
This is a json file that contains all block group GEOIDs and an associated list of GEOIDs for block groups that are located within a half mile radius of the main block group of interest.

bg2bg_1mi.json
This is a json file that contains all block group GEOIDs and an associated list of GEOIDs for block groups that are located within a one mile radius of the main block group of interest.

bg2bg_2mi.json

This is a json file that contains all block group GEOIDs and an associated list of GEOIDs for block groups that are located within a two mile radius of the main block group of interest.

Intermediate outputs

These are intermediate files in the data collection process.

SLD_BECs_V2.shp
This is the intermediate output from “1 - Aggregated Built Environment Characteristics (SLD).ipynb” and contains all of the built environment characteristics that were pulled from the raw EPA SLD data (CaliforniaSLD.shp) and then aggregated to larger spatial scales.

BEC_AggBG_NEW.shp
This is the intermediate output from “2 - Aggregated Network Characteristics (OSMnx).ipynb” and contains all of the data from SLD_BECs_V2, as well as network densities from OSMnx, aggregated to multiple spatial scales.

BEC_AggBG_V2.shp
This is the intermediate output from “3 - Aggregated Network Characteristics (GeoFabrik).ipynb” and contains all of the data from BEC_AggBG_NEW, as well as network densities from GeoFabrik, aggregated to multiple spatial scales.

VMT estimate files

These files are used in the final VMT estimation script.

cbsa2fipsxw.csv
This is a crosswalk file that contains FIPS codes for all CBSAs in California.

personlevel_BEstats.csv
This file contains the mean and standard deviation values used to convert built environment variables into z-scores. These values are based on the means and standard deviations for the sample of observations from the NHTS, and not those of the block group observations themselves.

Data collection

This folder contains all necessary files to pull raw data and generate a cleaned and processed output, containing all relevant built environment characteristics for every Census block group in the state. Scripts here were utilized to clean datasets for Task 2 and 4 regressions and were generated prior to exact model specification. As such, the built environment dataset contains many more variables than were used in the final model, and not only a few of them are needed for VMT estimate generation, as described further in that section.

To generate our built environment dataset, we collected, cleaned, and processed data from a number of sources. Population density, employment density, and jobs accessibility measures were collected from the EPA Smart Location Database. Bus density was computed based on raw data from transit agencies and regional planning organizations, collected by the authors in 2017 for use on previous research. Street density was collected from OpenStreetMap via the Python package OSMnx for 2024 data, and via the website GeoFabrik for 2018 data. The code to recalculate these densities can be found in the network_density subfolder. To update built environment metrics, the raw data should first be updated and then the .ipynb files in this folder should be run in sequence. Output names should be updated to reflect versioning.

Subfolder - network_density

0A - OSMnx Block Group Network Densities.ipynb
This file uses the OSMnx package to generate graphs for every block group in California. Given it takes a very long time to generate graphs for all 23,212 block groups in the state, it saves these networks to a .pkl file as an intermediate step. It then uses these graphs to calculate street and intersection density. Note that when calculating metrics in the future, this script should be used to calculate current day network densities.

0B - GeoFabrik Block Group Street Density.ipynb
This file extracts the statewide street network in a usable format from the PBF download from GeoFabrik. This shapefile is then brought into ArcGIS Pro to snip the statewide to each block group, and calculate total street length as well as density. Note that this is the script we used to calculate street density for our purposes, but as noted above, script 0A should be used for future built environment characteristics.

1 - Aggregated Built Environment Characteristics (SLD).ipynb

This file reads in the raw EPA SLD data and outputs a cleaned dataset that includes aggregated built environment characteristics as a shapefile (SLD_BECs_V2.shp).

2 - Joining Aggregated Network Characteristics (OSMnx).ipynb

This file reads in the intermediate output from “1 - Aggregated Built Environment Characteristics (SLD).ipynb” and joins it to the network densities from OSMnx, aggregating these to multiple spatial scales.

3 - Joining Aggregated Network Characteristics (Geofabrik).ipynb

This file reads in the intermediate output from “2 - Joining Aggregated Network Characteristics (OSMnx).ipynb” and joins it to the network densities from GeoFabrik, aggregating these to multiple spatial scales.

4 - Joining Aggregated Bus Densities + Final File.ipynb

This file reads in the intermediate output from “3 - Joining Aggregated Network Characteristics (Geofabrik).ipynb” and joins it to the bus stop density, aggregating these to multiple spatial scales. Note that for future built environment calculations, bus stop densities would need to be updated manually before running this script. It should be ensured that the spatial formatting of these stops matches the process descibed in Appendix Section B.3: Cleaning stop data. This results in the final set of built environment characteristics needed to generate VMT estimates via Appendix C.

Output - BEC_AggBG_V4

This is the final output from “4 - Joining Aggregated Bus Densities + Final File.ipynb” and contains all built environment metrics at all spatial scales for every Census block group in California. This is saved both as a shapefile and a CSV.

Estimate generation

This folder contains all necessary files to format the final data contained in BEC_AggBG_V4 to replace the historical built environment data contained in the “BECs by BG” sheet of the Appendix C VMT Estimates Excel Workbook. Note that the scripts provided here are to update this sheet only. The model coefficients used to generate VMT estimates (contained in “BEC Coefficients”) cannot change without new travel survey data (note that the data we relied upon were from the 2017-18 NHTS). We generally assume that the relationship between the built environment and VMT is stable over short periods of time.

5 - Built Environment Data Formatting for Appendix C.ipynb

This file reads in the final output from the data collection step (BEC_AggBG_V4) and reformats column names to match those used in regression outputs, so that formulas in Appendix C work properly. It also fixes certain variables that were slightly miscalculated, and calculates new variables (network load density and transit:car accessibility ratio) that are used to generate VMT estimates by block group. In addition, it converts the variables considered in the model output to Z-scores, as this is what was used in the model. It also adds the name of the home CBSA to each block group, to calculate CBSA-based VMT variation.

Output – BEC_AggBG_Clean.csv

This file contains the final properly formatted built environment characteristics for all block groups throughout the state. The values in this csv can be directly copy-pasted into the “BECs by BG” sheet of Appendix C to automatically recalculate VMT variation estimates for all block groups.