Spatial microanalysis and equity assessment of joint relationships among destination choice, activity duration, and mode choice

July 2022

A report from the Pacific Southwest Region University Transportation Center

Konstadinos G. Goulias, UCSB
Rongxiang Su, UCSB
Hui Shi, UCSB
Jingyi Xiao, UCSB

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
PSR-21-19	N/A	N/A

4. Title and Subtitle Spatial microanalysis and equity assessment of joint relationships among destination choice, activity duration, and mode choice	**5. Report Date** July 22, 2022 **6. Performing Organization Code** N/A
7. Author(s) **Konstadinos G. Goulias, 0000-0002-7436-9447** **Rongxiang Su, 0000-0003-1966-7289** **Hui Shi, 0000-0002-6947-9668** **Jingyi Xiao, 0000-0001-8328-6363**	**8. Performing Organization Report No.** TBD
9. Performing Organization Name and Address METRANS Transportation Center University of Southern California University Park Campus, RGL 216 Los Angeles, CA 90089-0626	**10. Work Unit No.** N/A **11. Contract or Grant No.** USDOT Grant 69A3551747109
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590	**13. Type of Report and Period Covered** Final report (August 15, 2021 – August 14, 2022) **14. Sponsoring Agency Code** USDOT OST-R

15. Supplementary Notes
https://geotrans.geog.ucsb.edu

16. Abstract
This project develops a new integrated framework for experienced walking accessibility and assessment of disparities using motif and sequence analysis based on the 2012 California Household Travel Survey and the 2017 National Household Travel Survey for California. This is then complemented with other external data at the business establishment microlevel and at different geographic levels to account for residence in places of high social vulnerability. We develop many different model formulations to study Vehicle Miles of Travel (VMT) and show the power of quantile regression as analytical tool challenging the wisdom of VMT as a policy variable. We then turn to exploring the spatial distribution of time use and travel fragmentation demonstrating the power of spatial analysis in identifying geographies characterized by people with highly fragmented schedules and study the relationship with accessibility to opportunities and land use characteristics. This study ends with analysis of walking accessibility to retail and education opportunities that finds key findings include lower accessibility to retail and education opportunities for people living in places that are classified as populated by vulnerable residents (e.g., minority and lower income tracts). This analysis also indicates substantially higher and heterogeneous experienced accessibility among people that visit multiple distinct locations but with decreasing returns to the investment of visiting multiple locations. As expected living in higher density in terms of population and activity opportunities is also offering higher experienced walking accessibility even when we account for asymmetry in the distribution of accessibility indicators.

17. Key Words	18. Distribution Statement
Land use, accessibility, mobility, spatial clustering, activity-travel fragmentation, sequence analysis, travel survey, vehicle miles traveled, quantile regression, latent profile analysis, built environment, accessibility	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	94	N/A

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

Pacific Southwest Region UTC
University Transportation Center

Contents

About the Pacific Southwest Region University Transportation Center

The Pacific Southwest Region University Transportation Center (UTC) is the Region 9 University Transportation Center funded under the US Department of Transportation's University Transportation Centers Program. Established in 2016, the Pacific Southwest Region UTC (PSR) is led by the University of Southern California and includes seven partners: Long Beach State University; University of California, Davis; University of California, Irvine; University of California, Los Angeles; University of Hawaii; Northern Arizona University; Pima Community College.

The Pacific Southwest Region UTC conducts an integrated, multidisciplinary program of research, education and technology transfer aimed at *improving the mobility of people and goods throughout the region.* Our program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high growth areas.

U.S. Department of Transportation (USDOT) Disclaimer

Disclosure

Konstadinos G. Goulias (PI), Rongxiang Su, Hui Shi, and Jingyi Xiao, conducted this research titled, "Spatial microanalysis and equity assessment of joint relationships among destination choice, activity duration, and mode choice" at GeoTrans Laboratory, Geography Department, Division of Mathematical, Life, and Physical Sciences, University of California Santa Barbara. The research took place from August 15, 2021 to August 22, 2022 and was funded by a grant from the US DOT in the amount of $100,000. The research was conducted as part of the Pacific Southwest Region University Transportation Center research program.

Acknowledgements

This research was supported by the Department of Geography at the University of California Santa Barbara and the United States Department of Transportation Pacific Southwest Research Region 9 University Transportation Center. The authors are grateful to the anonymous reviewers for their valuable comments and suggestions on an earlier version of the manuscript and drafts of peer reviewed journal papers.

Abstract

This project develops a new integrated framework for experienced walking accessibility and assessment of disparities using motif and sequence analysis based on the 2012 California Household Travel Survey and the 2017 National Household Travel Survey for California. This is then complemented with other external data at the business establishment microlevel and at different geographic levels to account for residence in places of high social vulnerability. We develop many different model formulations to study Vehicle Miles of Travel (VMT) and show the power of quantile regression as analytical tool challenging the wisdom of VMT as a policy variable. We then turn to exploring the spatial distribution of time use and travel fragmentation demonstrating the power of spatial analysis in identifying geographies characterized by people with highly fragmented schedules and study the relationship with accessibility to opportunities and land use characteristics. This study ends with analysis of walking accessibility to retail and education opportunities that finds key findings include lower accessibility to retail and education opportunities for people living in places that are classified as populated by vulnerable residents (e.g., minority and lower income tracts). This analysis also indicates substantially higher and heterogeneous experienced accessibility among people that visit multiple distinct locations but with decreasing returns to the investment of visiting multiple locations. As expected living in higher density in terms of population and activity opportunities is also offering higher experienced walking accessibility even when we account for asymmetry in the distribution of accessibility indicators.

Spatial microanalysis and equity assessment of joint relationships among destination choice, activity duration, and mode choice

Executive Summary

In this report we first explain the reasoning and provide motivation for detailed spatiotemporal analysis of amenities provided to the California population to identify geographical pockets for which current policy may have unintended consequences. The three most important needs identified are: 1) understand the possible difference in sensitivity of travel behavior to the land use policy variables such as density of opportunities; 2) identify places in which fragmentation of time allocation makes it necessary to use the private automobile; 3) understand what motivates people to visit multiple distinct destinations in a day and the relationship of this behavior to availability of opportunities. In terms of analysis we first explore the classification of every US Census block in California using Latent Profile Analysis and a detailed inventory of business establishments. We derive in this way four distinct types that depend on the business establishment density. Then, we use the Social Vulnerability Indices (SVI) developed by CDC to explore the correlation between place of residence land use characteristics and SVI. Coastal areas have the lowest SVI and in urbanized environments also higher density of opportunities. There are many rural areas that also have high vulnerability. This motivates the microanalytic model specification developed in the rest of the chapters in this report.

One core analytical chapter develops and presents a new type of regression model with vehicle miles traveled (VMT) as dependent variable. The motivation for doing this is that VMT is a primary policy variable because it is strongly correlated with greenhouse gas emissions and the density and diversity of urban environments. Policies that increase density and diversity of land use are believed to decrease VMT presumably by replacing it with non-motorized travel. This causal inference has not been conclusive and may create further disparities by gentrification and an increase in long-distance travel for lower-income commuters. In this study, we explore the heterogeneity in the relationships between personal level VMT and the built environment, accessibility to opportunities and open spaces, and people's socio-demographic traits across various levels of travel demand. The present research first develops opportunity-based measurement of residential land-use indicators at the US census block level for the whole California. This is followed by the use of the Latent Profile Analysis as the land-use indicators identifying distinct land use patterns experienced by the respondents in the 2012-2013 California Household Travel Survey. Thus, we estimate quantile regression models to understand the heterogeneity in the relationships between VMT and residential built environment characteristics and people's socio-demographic traits. The results indicate different sensitivity to land use at various travel intensities implying different response to land use policies and challenges the wisdom of using VMT as the main policy variable.

Then, we turn to the data from the 2017 National Household Travel Survey in California from 26,078 survey participants, using sequence analysis to estimate a fragmentation indicator of people's daily schedules. This is followed by spatial clustering to find groups of observations with similarly high or low fragmentation using the longitude and latitude of their residential locations. Applying hierarchical sequence clustering within each spatial cluster we identify distinct patterns of time allocation. Using the Local Indicator of Spatial Association (LISA) we find a large portion (approximately 30%) of the sample with significant spatial clustering of fragmentation. We also find systematic and significant differences in membership to these clusters based on land use, county of residence, household and personal characteristics, and travel modes used. Sequence analysis pattern recognition within LISA spatial clusters shows systematically repeating time allocation patterns that include typical work and school schedules as well as staying at home patterns. However, each spatial LISA cluster is composed of different time allocation clusters. All this analysis taken together points out substantial and measurable heterogeneity in spatial clustering of fragmentation and the need for customized policy actions in different geographies.

The final analysis in this report explores the walking accessibility to opportunities by enumerating the distinct destinations visited in a day and the correlation between the number of destinations and accessibility. People that visit multiple locations also experience exponentially increasing with the number of locations walking accessibility. The 20 minute walking city is in essence composed of multiple destinations that are surrounded by many activity opportunities. We explore heterogeneity in experienced accessibility using multivariate regression for retail and education experienced accessibilities as a function of person and household characteristics, residence in one of the LPA types identified here, and SVI. Key findings include lower accessibility to retail and education opportunities for people living in places that are classified as populated by vulnerable residents (e.g., minority and lower income tracts). This analysis also indicates substantially higher and heterogeneous experienced accessibility among people that visit multiple distinct locations but with decreasing returns to the investment of visiting multiple locations. As expected living in higher density in terms of population and activity opportunities is also offering higher experienced walking accessibility even when we account for asymmetry in the distribution of accessibility indicators.

1. Introduction

In recent analyses of daily travel we find more than 80% of the traveling participants in the California component of the National Household Travel Survey (NHTS) visit 2 or more distinct locations outside their home. They do this for a wide variety of activity purposes and a wide variety of activity durations. This heterogeneity is even more pronounced when we analyze data from teleworkers (i.e., people that dedicate time to work from home). Visiting multiple distinct locations in a day is accompanied with heavy use of the private car contributing to the State's inventory of Vehicles Miles of Travel (VMT). This implies high greenhouse gas (GHG) and other air pollutant emissions and may contribute to wasted energy. Finding ways to offer variety of destinations for people while at the same time decrease the distances among them and/or serve by fast non-polluting services is the single most important purpose of the research in this project. The complexity of this problem is due to the many choices people face in their daily patterns. These include and are not limited to duration of an activity at a place, different locations visited to complete these activities, mode(s) used to reach locations, distance traveled to the locations, and the assembly of these activity locations in tours (i.e., a sequence of places and trips that start and end at major stations of activity such as home or places of work and school). People consolidate activities and trips in tours (called tour formation) to decrease travel time but they cannot always do this. Understanding tour formation behavior in space and time has advanced considerably since the early attempts to model this behavior in the 1980s and it is the foundation of the activity-based travel demand forecasting methods used by regions for their regional transportation plans and sustainable community strategies in California. However, the state of the art in this behavioral facet did not advance enough to produce clear taxonomies of behavior for different segments of the population and most important did not advance enough to clearly identify population segments that are not able to reach activity opportunities (also called amenities) by modes available to them. This lack of understanding on how people combine different options together is reflected in over-simplified statements of the type 15-minute cities. These recent popular ideas about the way cities should transform for better quality of life include advocacy for transforming opportunities and amenities for the residents of these cities (Bloomberg Businessweek, 2020). One proposition is to create a buffer around residences within which essential services are provided in relatively short travel time. These buffers are in essence accessibility indicators that are defined in terms of travel time defining cities as the "15-minute city" (Moreno et al. 2021), "20-minute city" (Capasso et al., 2020), or the "30 minute city" (Levinson, 2019). The description of these travel time buffer cities includes ideas of complete streets (smartgrowthamerica.org) with the addition of a mix of modes and principles of city design borrowed from practicing architects such as Gehl (2011). All these ideas align well with the California Senate Bill 375 motivated Sustainable Community Strategies (SCS) in the California Regional Transportation Plans (also known now as RTP/SCS). Assessing the number of opportunities and amenities within travel time buffers can be done at fine scales such as the US Census block and using counts of business establishments minimizing spatial error called ecological fallacy and we did this in GeoTrans more than a decade ago (Chen et al., 2011, Lei et al 2012). To produce complete assessment of cities just mapping these travel time-based accessibility indicators is not sufficient and behavior needs to be added (Joh et al., 2008). Moreover adding travel stages details about access to different modes (Boarnet et al., 2017) and correlating amenities, facility characteristics, and activity opportunities can inform specific interventions (Lin et al. 2014). These sophisticated methods, however, describe the environment in spatial and temporal detail and they pin-point to specific trips of individuals (e.g., commuting or leisure) but they do not take into account a complete daily schedule in space and time and are not designed in a way that interactions among persons become integral in the city offered quality of service for activities. We will do this in the project proposed here and include human interaction. Figure 1.1 is indicative of one type of interaction showing two patterns of activities and trips by two persons that live in the same place.

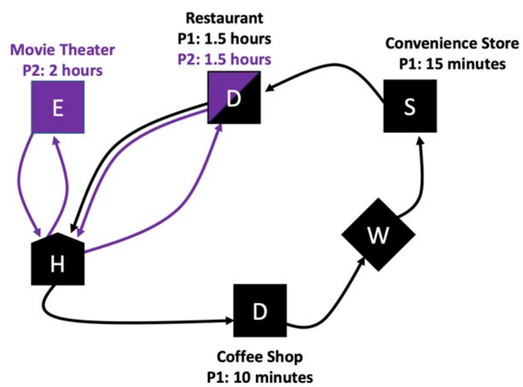

Figure 1.1 Daily schedule of two persons (artwork by Adam Davis)

Person 1 (the black trace) leaves in the morning from home driving alone the household car, stops at a coffee shop, then goes to work and stays for 7 hours. After work this person goes to a convenience store and then to a restaurant to meet Person 2 who came by bus (P2). Both persons return home in the household car. P2 during this day also went to a movie theater using the household car after dining with P1. Fundamental to understanding the decisions underlying the patterns we see in Figure 1.1 are models of spatial choice (e.g., business establishment locations that are the coffee shop, convenience store, restaurant, and movie theater). Choices that people make in space, however, are correlated due to their spatial proximity, spatial organization of activity locations, and urban form. Senate Bill 375 and the coordination between land use and transportation policies aim at increasing density and diversity of activity opportunities to change not only location choices of people but also their modes of traveling switching to active models such as walking and using the bicycle. This motivates the key aspects we emphasize in this project that are destination choice, activity duration, and mode choice. We do this analysis in an unconventional and more holistic way to capture their spatial and temporal order and correlation of the joint daily scheduling patterns of people.

In developing spatio-temporal taxonomies of daily patterns, it is not feasible to use trip chains as in Figure 1.1 because the combination of location types and interactions we observe in the data are a vast amount. We can extract from the data simplified patterns that use only the essential characteristics of their spatial organization. These are called motifs and they focus on the variety of destinations visited in a day to form networks of connectivity among these destinations. These networks keep the information of one- or two-way movements between destinations in a daily pattern. The difference between motif and trip chain is that a mobility motif does not distinguish locations (nodes) by activity types collected in a travel diary but by the geographical coordinates (sometimes manifested as an encrypted ID due to privacy concerns). For example, a trip chain of “Home-Work-Home” and a trip chain of “Home-School-Home” are the same if converted to motif representations (i.e., they will be the same motif with two nodes and a pair of bidirectional links between the nodes). In contrast, a trip tour that does not explicitly include destination choice uses analyst-defined activity categories and does not distinguish among activity locations in which people conduct the same activity (e.g., shopping activity at a grocery store or at a mall). The emerging tool of mobility motif is capable of capturing heterogeneous human daily mobility patterns with regards to the diverse interconnections among distinct individually visited locations and trips undertaken. An example of motifs we extracted from the same data we will use in this project is Figure 1.2.

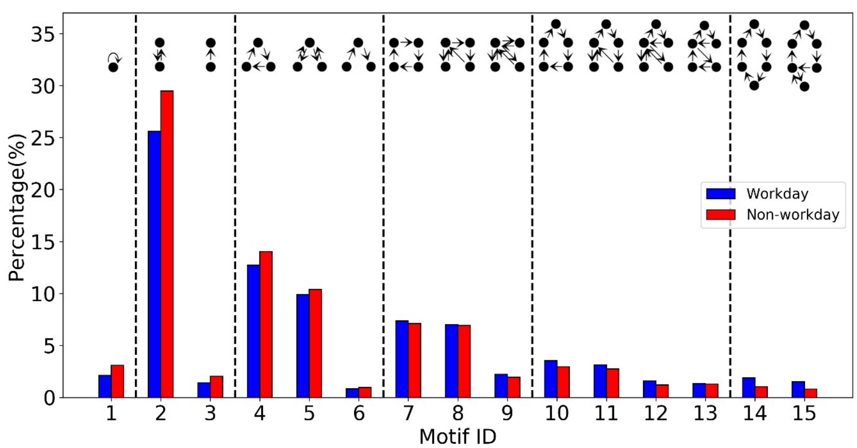

Figure 1.2 An example of motif extraction from observed data

This figure shows motifs selection is different between non-workdays (weekend days and official federal and state holidays) and typical non-holidays weekdays (i.e., workdays). Each dot is a distinct location and the arrows signify one way movement. The interaction shown in Figure 1.1 when studied in terms of motifs will be a two-person household with Person 1 choosing motif number 10 in Figure 1.2 and Person 2 choosing motif number 5. This representation not only accounts for possible variety seeking in destination choice but also captures essential elements of trip chaining. In one of our studies that used the idea of motifs and the 2017 California component of the National Household Travel Survey (California-NHTS) discovered that 16 unique motifs can capture daily mobility patterns of 83.05% of the total population in the data (Su et al., 2020). This makes motif pattern recognition an efficient summary of spatial activity-travel patterns. However, this is not enough for our analysis. The motifs pattern recognition needs enriching with added information about activity durations, types of activities, and travel distance and time. Figure 1.3 shows a combination of motif and minute-by-minute sequence analysis of activities and trips daily patterns that have been further classified using hierarchical clustering. Figure 1.3 is indicative of heterogeneity of human behavior and cyclical manner of activity engagement that changes by time of day, day of week, month of year but also in the life course of individuals. This is also the outcome of the interplay between behavior and available opportunities underlying the complex dynamics that are included in the data we have from diaries. In GeoTrans we developed sophisticated techniques to study scheduling of activities at fine detail, we can distinguish among different daily time allocation patterns we find in the population. Figure 1.3 is an example of the taxonomy we developed using NHTS diary data for a different project to study telecommuting (Su et al., 2021). In fact, Figure 1.3 has four distinct time allocation patterns of traditional commuters for the motifs that contains many diverse locations in a day (i.e., numbers 10 to 15). The four images in Figure 1.3 are stacked bar graphs that show at each minute of a 24 hour period the percent of people that travel or are at a place buying meals or services, working at a workplace or home, shopping, visiting friends and family and so forth. This method is based on sequence analysis and classification of the daily patterns of individuals using hierarchical clustering. Figure 1.3 contains one example from people that are commuters and shows four distinct patterns in a day from a wide variety of patterns we discovered (Su et al., 2021). Note the daily rhythms of arrival at work, lunch break from a small portion. Note also that the pattern of mostly staying at home is also the pattern with a higher fraction of people spending time buying goods. Similar patterns are obtained when we look within motif number 3 (in Figure 1.2) but with more people buying goods and meals during lunch time and after work. Depending on the time of day patterns in Figure 1.3 approximately 55% of the persons drive alone and another 20% drive another person. In addition, large proportion (more than 70% and different among the four patterns in Figure 1.3) of these persons are full time employed. These are four groups of people we expect contributes substantially to the California VMT and in turn to greenhouse gas emissions. They are however very different across the four groups and they also have differences within groups that we will pin-point in this project.

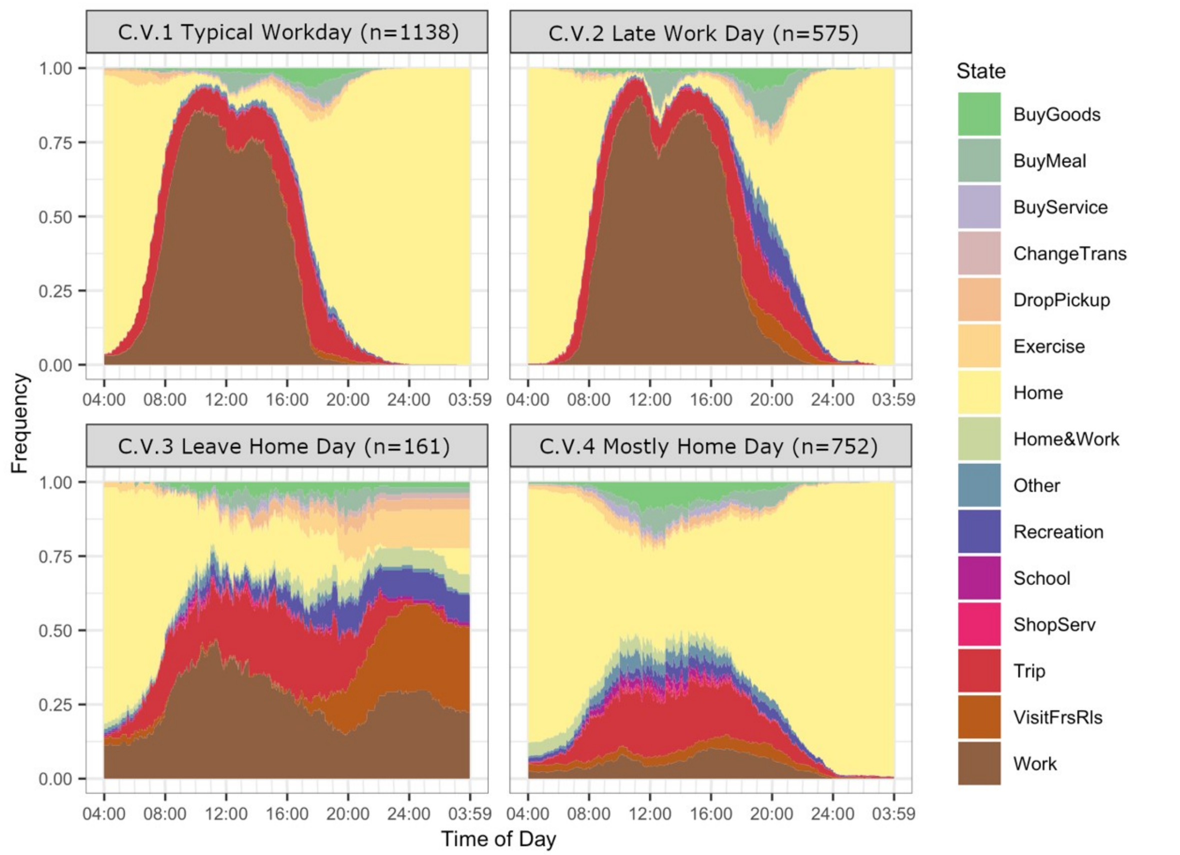

Figure 1.3 Minute-by-minute stacked bar graph of motifs that have four nodes

A major determinant of the genesis of these four different patterns and their high VMT is the built environment. One way to account for the built environment influence on behavior, is to develop models of opportunities and constraints that shape individual behavior and conceptualize action within the boundaries of possibilities represented at a scale commensurate with the scale of Figure 1.3 (these are accessibility indicators that also account for the opening and closing time of businesses – Chen et al., 2011). In this project we develop spatial and temporal relationships among destinations and activity type and duration at the locations in the motifs and we will find out, for example, if one of the four patterns in Figure 1.3 is associated with higher density and diversity of opportunities and if this also implies higher percentage of trips walking and riding a bicycle instead of driving alone. Most measures of accessibility are centered around places where people spend most of their time in a day such as home and work locations around which availability and reach of opportunities are measured. In this project we will compute a wide spectrum of accessibility indicators at different travel times (20, 30, 40 minutes) around each of the nodes in the motifs using as base data of opportunities the detailed business establishment information. Then, we test the correlation of similar patterns to Figure 1.3 with a variety of combinations of accessibility indicators to find the best explanatory power. In essence the task here is to define different types of accessibility indicators at fine spatial and temporal scale and then develop indices that will in turn be tested in a structural equation model that includes the choice of motifs and daily time allocation patterns as internal variables to the structural relationships. We experimented with similar methods using data from Pennsylvania and verified that accessibility should be considered for the entire daily pattern of locations visited. This complicates the analysis but a proof of its feasibility is our 2017 Pyke Johnson Award paper published in the Transportation Research Record (Lee et al., 2017). That paper shows there are different groups of people in their daily interaction patterns and they are influenced by different contexts (e.g., life cycle stage) and different environments they visit in a day. That study was done for State College, PA, which is a small town and includes a large University (PennState) with a mix of rural and suburban settings. We expect more variety in daily patterns using the entire California as case study in this project. This means we will obtain larger numbers of groups of patterns (clusters) and more diversity within each cluster. We also expect to find that people in a day visit places characterized by very different access to opportunities.

In this research project we explore equity in providing accessibility to communities and their individuals from different viewpoints to account for context in experienced accessibility in a way that respects the correlations between roles and responsibilities and concomitant enjoyment of access to opportunities to work, study, and play. In Chapter 2 we analyze and map access to opportunities by different social vulnerability indexed (SVI) US Census tracts. SVI is developed by the Center for Disease Control (CDC) to explore in four different themes geographic distribution of populations that are more or less vulnerable to threats. Since SVI is using percentile to rank Census tracts it makes it particularly suitable for comparisons and assessment of equity in accessibility provision. In Chapter 3 we address the problem of heterogeneous sensitivity of one key planning parameter, daily vehicle miles travelled (VMT). VMT is used in transportation planning due to its direct correlations with fuel consumption and CO_2 emissions. But as we show in our analysis VMT should be divided into different VMT types (i.e., driving alone, giving rides to other people, being a passenger etc.) and its correlation with land use is different depending on the amount of VMT a person accumulates in a day. This implies policies based on VMT will have undesirable impacts on the vulnerable. In Chapter 4 we turn to the theme of fragmented schedules we studied in GeoTrans multiple times for different behavioral aspects (in essence a summary indicator of the time allocation schedules of Figure 1.3). This time we explore spatial correlation of fragmentation and its correlation with amenities offered by the place of residence as well as person characteristics. In Chapter 5 we address experienced accessibility from the viewpoint of multiple distinct destinations people visit in a day. Preliminary analysis shows that people that experience a high level of opportunities also visit different destinations in a day and the experience of opportunities rises exponentially with the number of distinct locations visited.

2. Mapping Opportunities and Social Vulnerability

2.1 Built environment opportunity representation

To measure the built environment, we use data from the National Establishments Time Series (NETS) database in 2013, which is a comprehensive record of all business establishments in the United States. The variables we used from NETS include the distinct industry category (NETS classified them into 14 categories), number of employees, and geographic coordinates of each business establishment. The 2013 NETS point-based data were then aggregated to the US census block level and converted to a density scale (i.e., average the number of employees in each industry over the area of each block) to characterize the observed opportunities within each census block for the whole California (691,721 blocks in total). Table 2.1 shows the descriptive statistics for the 14 opportunity indicators. These 14 density-based opportunity indicators are used further in a classification method based on latent profile analysis to extract distinct land-use patterns for the 691,721 US Census blocks to describe residential built environment characteristics.

Table 2.1 Descriptive statistics for the 14 density-based opportunity indicators for 691,721 census blocks (unit: number of employees per square kilometer)

Variables	Description	Mean	Median	S.D.	Min	Max
EMPMAN	Manufacturing	179.25	0	6712.61	0	3670109.21
EMPPRO	Professional, scientific management, administrative, and waste management services	402.61	0	6497.45	0	1717171.01
EMPFINA	Finance, insurance, real estate, and rental and leasing	202.61	0	4335.72	0	1402006.27
EMPTRA	Transportation, and warehousing and utilities	64.63	0	2503.31	0	884989.35
EMPPUB	Public administration	120.10	0	5657.30	0	2229639.37
EMPART	Arts, entertainment, recreation, accommodation, and food services	191.34	0	4862.35	0	2661770.29
EMPHEA	Healthcare	194.79	0	6807.38	0	3109093.63
EMPWHO	Wholesale trade	105.66	0	2687.18	0	1120680.03
EMPINF	Information	91.98	0	3192.54	0	1212629.23
EMPOTH	Other services (except public administration)	141.00	0	3167.95	0	2046233.74
EMPCON	Construction	103.15	0	2045.23	0	848949.07
EMPRET	Retail trade	243.50	0	4566.24	0	1734064.12
EMPAGR	Agriculture, forestry, fishing and hunting, and mining	19.78	0	1763.04	0	1033585.97
EMPEDU	Education	75.40	0	3067.25	0	1547417.52

2.1.1 Latent profile analysis

Latent profile analysis classifies observations into latent categories based on similarities in their response variables. Unlike deterministic clustering approaches such as k-means cluster analysis, LPA estimates the probability of being in each latent profile conditional on individual's response variables (Nylund-Gibson and Choi, 2018; McBride et al., 2018). A series of models are estimated with one or more latent profiles during the process of model specification. By comparing these models using a range of fit statistics and taking into account the parsimony and ease of interpretation, we can select the best fitting model (i.e., optimal number of classes) (Bauer and Curran, 2003). These fit statistics include the Bayesian Information Criterion (BIC; Schwarz, 1978), Akaike Information Criterion (AIC; Akaike,1987), Adjusted BIC (Sclove, 1987), Entropy (an index ranging from 0 to 1, with higher value indicating greater classification accuracy), the Lo-Mendell-Rubin test (LMR; Lo et al., 2001) and the Bootstrap Likelihood Ratio Test (BLRT; Mclachlan and Peel, 2000). LPA is implemented in the R package tidyLPA (Rosenberg et al., 2018) and Mplus software (Muthen and Muthen, 1998).

The 14 density-based built environment indicators were transformed into log scale to mitigate the very positive skewed distribution (each variable was added 1 to avoid negative infinity when applying log transform). We estimated models ranging from one to five profiles with the goal of identifying a few discrete categories to represent the residential built environment characteristics. According to the fit indices as shown in Table 2.2, a five-profile solution is optimal (lowest AIC/BIC, highest Entropy, and p-value of BLRT). However, the four-profile solution is also acceptable. By comparing the standardized means of the 14 density-based built environment indicators of the four-profile model (Figure 2.1) and the five-profile model (Figure 2.2), we found that class 1 and class 2 in the five-profile model do not have clear separation (i.e., similar to each other), which indicates no need to use five classes. Furthermore, considering the interpretation and parsimony of the model, the four-profile solution is chosen for the rest of the analysis in this report.

Table 2.2 Model fit statistics for latent profile analysis of 1 to 5 class models

Fit statistics	1 class	2 class	3 class	4 class	5 class
Log-likelihood	-18846702	-17455513	-16765328	-16257031	-15947465
AIC	37693460.7	34911111.2	33530772.7	32514208.1	31895105.6
BIC	37693781.2	34911603.4	33531436.6	32515043.7	31896113
Entropy	n/a	0.976	0.98	0.989	0.99
prob min	n/a	0.969	0.963	0.974	0.973
prob max	n/a	0.996	0.999	0.999	0.998
n min	n/a	0.122	0.015	0.015	0.007
n max	n/a	0.878	0.854	0.623	0.622
p-value of BLRT	n/a	<0.001	<0.001	<0.001	<0.001

Note: AIC, BIC: lower values of AIC/BIC indicate better model fit. Entropy: measure the uncertainty of classification with higher value indicates lower uncertainty. prob min/prob max: minimum/maximum of the average probabilities for most likely class membership. n min/n max: the proportion of the sample assigned to the smallest/largest class. BLRT: comparing the improvement between neighboring class models, significant p-value of BLRT indicates significant improvement in model fit compared to the model with previous number of classes.

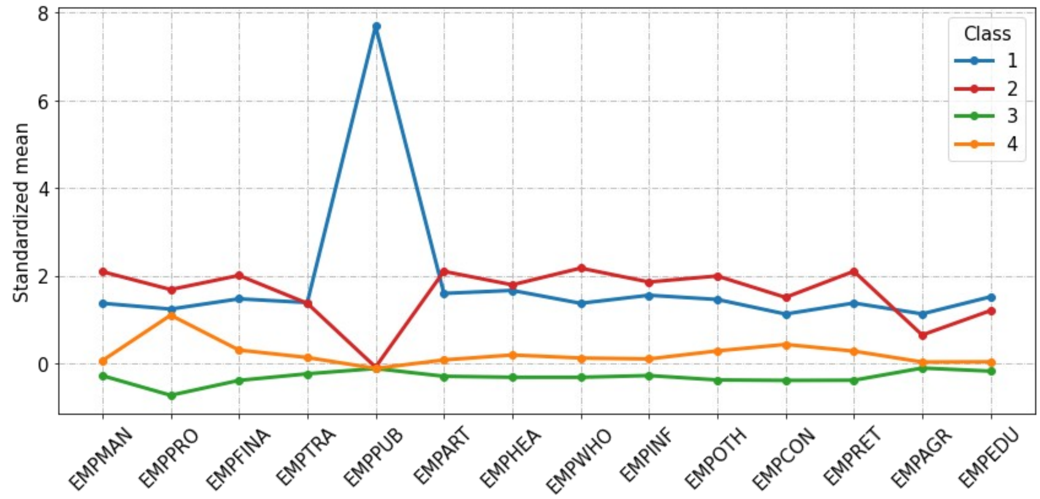

Figure 2.1 Standardized means of the 14 built environment indicators for the four-profile solution

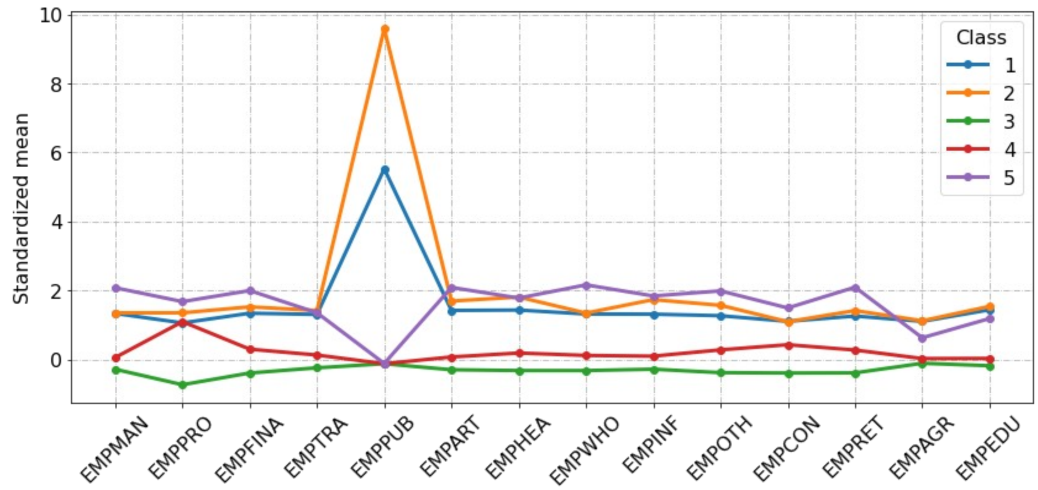

Figure 2.2 Standardized means of the 14 built environment indicators for the five-profile solution

Figure 2.1 that displays the selected solution of clustering also shows the standardized means of the 14 density-based built environment indicators for each latent profile. The first extracted group contains 10,375 census blocks (1.50%), and it shows a high level of EMPPUB (public administration) and moderately high levels on everything else. It is therefore characterized as the group of high density in EMPPUB. We called this group *LPA1_publicAdmin*. The second profile comprises 45,621 census blocks (6.6%) and shows very high densities in all industries except EMPPUB, EMPAGR (agriculture, forestry, fishing and hunting, and mining), and EMPEDU (education) and is labeled *LPA2_highDensity*. The third profile is the largest of all (n=430,658; 62.26%), and presents relatively lowest densities with respect to all the 14 categories of opportunities. We named this profile *LPA3_lowDensity*. The fourth profile consists of 205,067 census blocks (29.65%), and has moderately low levels of everything and is therefore labeled *LPA4_suburban*.

As an example of the spatial distribution of this block classification, the map in Figure 2.3 displays the geographic distribution of the four profiles in Los Angeles County which is the most populated county in California. The dominant *LPA3_lowDensity* shown in green is mostly distributed in less populated or/and less developed areas. Blocks classified into *LPA4_suburban* (in orange) are mostly surrounding or within the urban area. Blocks in *LPA2_highDensity* (in red) are mostly located in urban core and urban center with more opportunities. *LPA1_publicAdmin* (in blue) captures blocks where the airports or other public administration offices are located. In general, these four profiles manage to capture the diverse land-use patterns in urban and suburban areas in Los Angeles County, which proves the efficacy of using LPA to extract distinct land-use patterns.

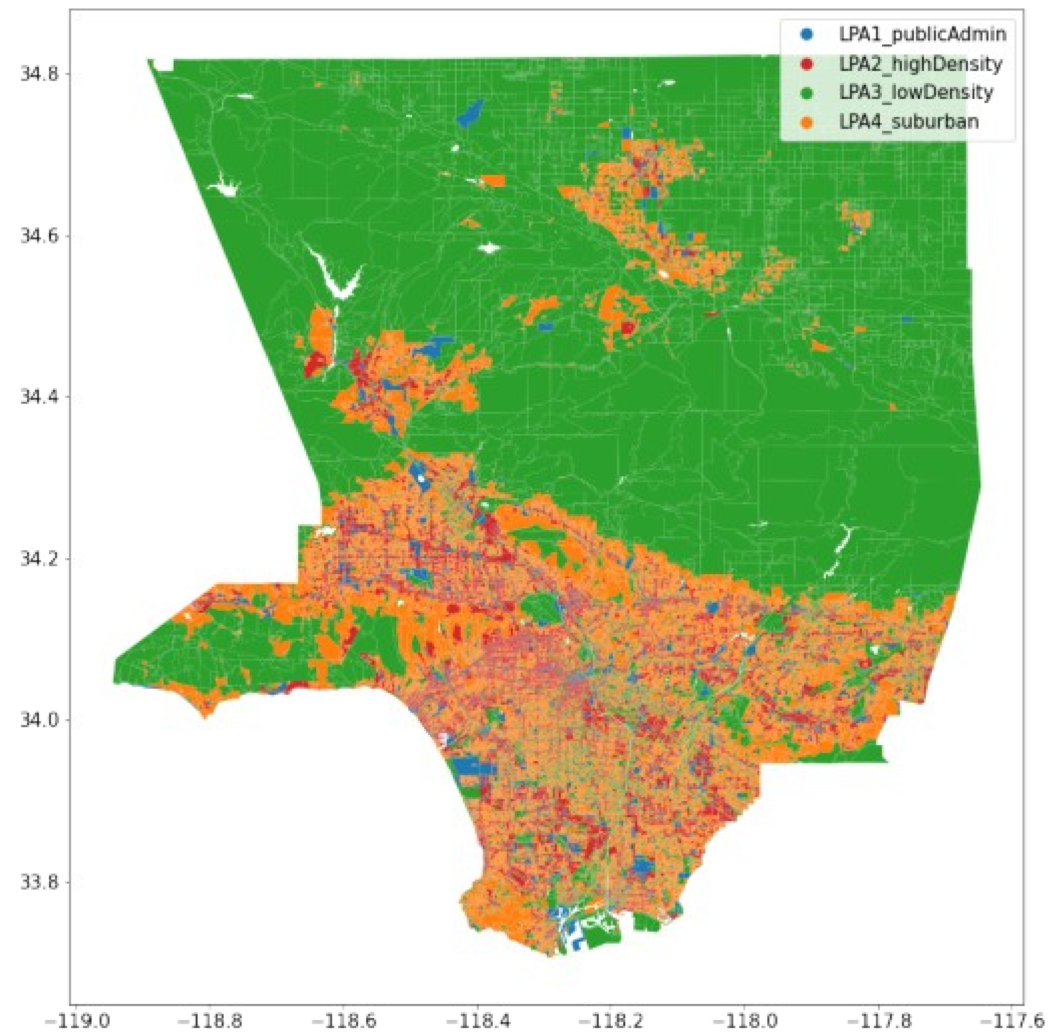

Figure 2.3 The four-profile classification for all census blocks in Los Angeles County, California

2.2 Social vulnerability and LPA classified Census block of residence

As mentioned earlier CDC developed social vulnerability indices to support further geographic analysis and the CDC web site mentions:

“Social vulnerability refers to the potential negative effects on communities caused by external stresses on human health. Such stresses include natural or human-caused disasters, or disease outbreaks. Reducing social vulnerability can decrease both human suffering and economic loss.” (https://www.atsdr.cdc.gov/placeandhealth/svi/index.html - accessed July 2022)

CDC derives different indices in four domains that form the basis of the SVI and they are: 1) socioeconomic status (comprising income, poverty, employment, and education variables); 2) household composition and disability(comprising age, single parenting, and disability variables); 3) minority status and language(comprising race, ethnicity, and English- language proficiency variables); and 4) housing and transportation(comprising housing structure, crowding, and vehicle access variables). Possible scores for each index and the overall index range from 0 (lowest vulnerability) to 1 (highest vulnerability). Figure 2.4 is a prepared county level mapping of SVI for each of the 4 themes/domains for LA County (https://svi.cdc.gov/Documents/CountyMaps/2018/California/California2018_Los%20Angeles.pdf – accessed July 2022). Notable is the level of spatial aggregation when compared to the LPA classification. Also, notable is the higher concentration of census tracts with high SVI in the northern portion of LA county and many tracts in the most urbanized portion. Notable, however, is the absence of high SVI along the coastal portion of the County. This is clearer in the plot of the combined SVI of Figure 2.5. The coastal region has a lower SVI (less vulnerable population) and large pockets of the urban environment and many areas far from the city centers are the places with the highest vulnerability.

We explore the correlation between SVI and LPA using descriptive statistics in the form of box plots. Figure 2.6 shows this relationship plotting the SVI index for each of the four LPA classified blocks. However, we should note the SVI is at the tract level and the LPA at the block level. The overall pattern shows somewhat symmetric distributions of SVI within each LPA group. Clearly emerging trends include higher socioeconomic and household composition and disability SVI in the low density LPA (mean, median, and the entire distribution) and lower SVI in the suburban LPA, presumably due to higher car ownership levels of the suburbs (one of the factors used in the housing & transportation SVI is percent of households with no car owned). Figure 2.6 also shows the overall SVI (combination of the four themes) and its correlation with the four LPA categories and again suburban residents appear to be less vulnerable and the residents in the LPA1 the most but all LPA have a wide spectrum of SVI residents. We will come back to these relationships in the following chapters in which we use as explanatory variables of behavior similar variables used to derive SVI and an expanded set of built environment and open space variables describing the residence of survey participants.

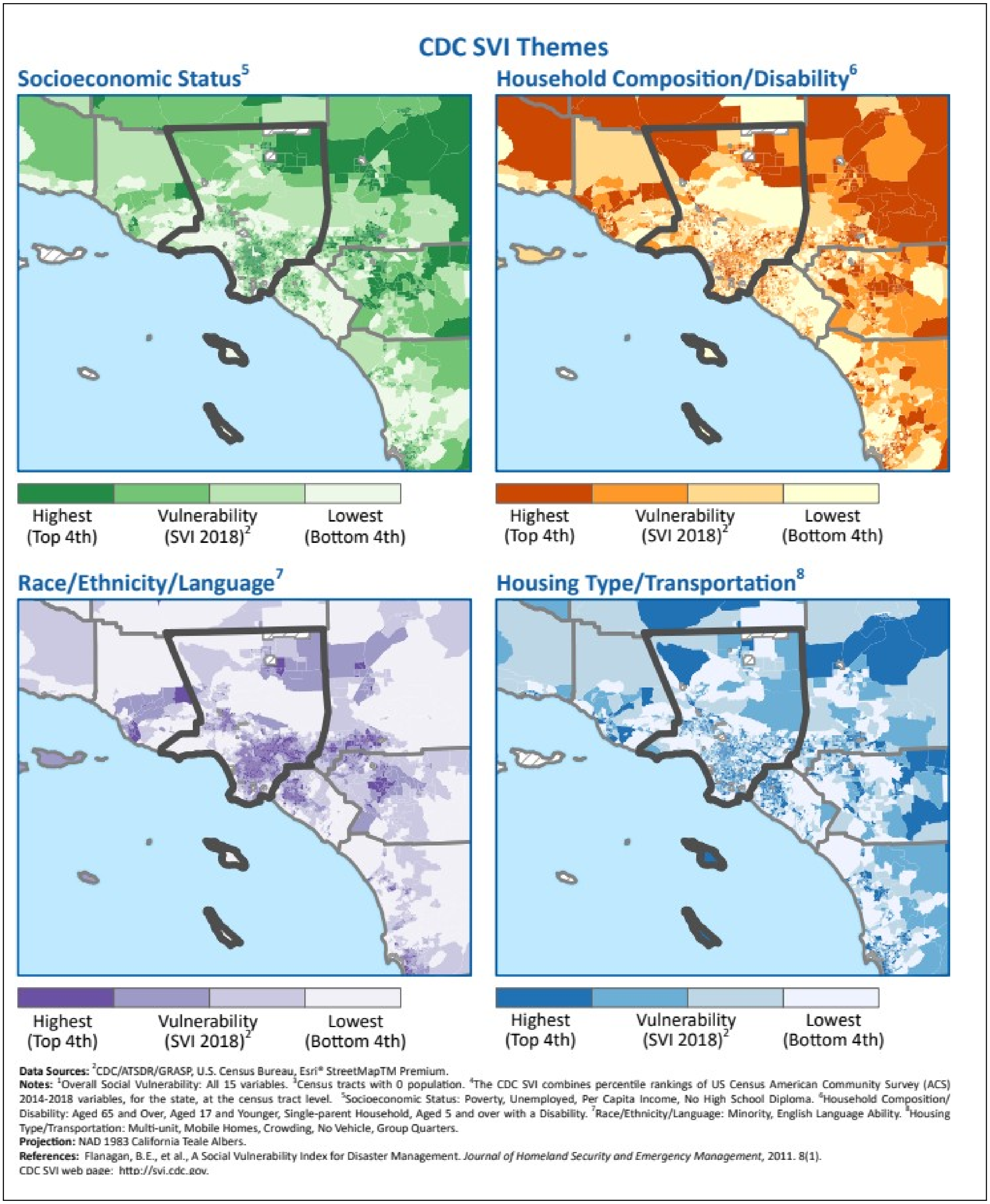

Figure 2.4 CDC SVI mapping of the four themes in LA County

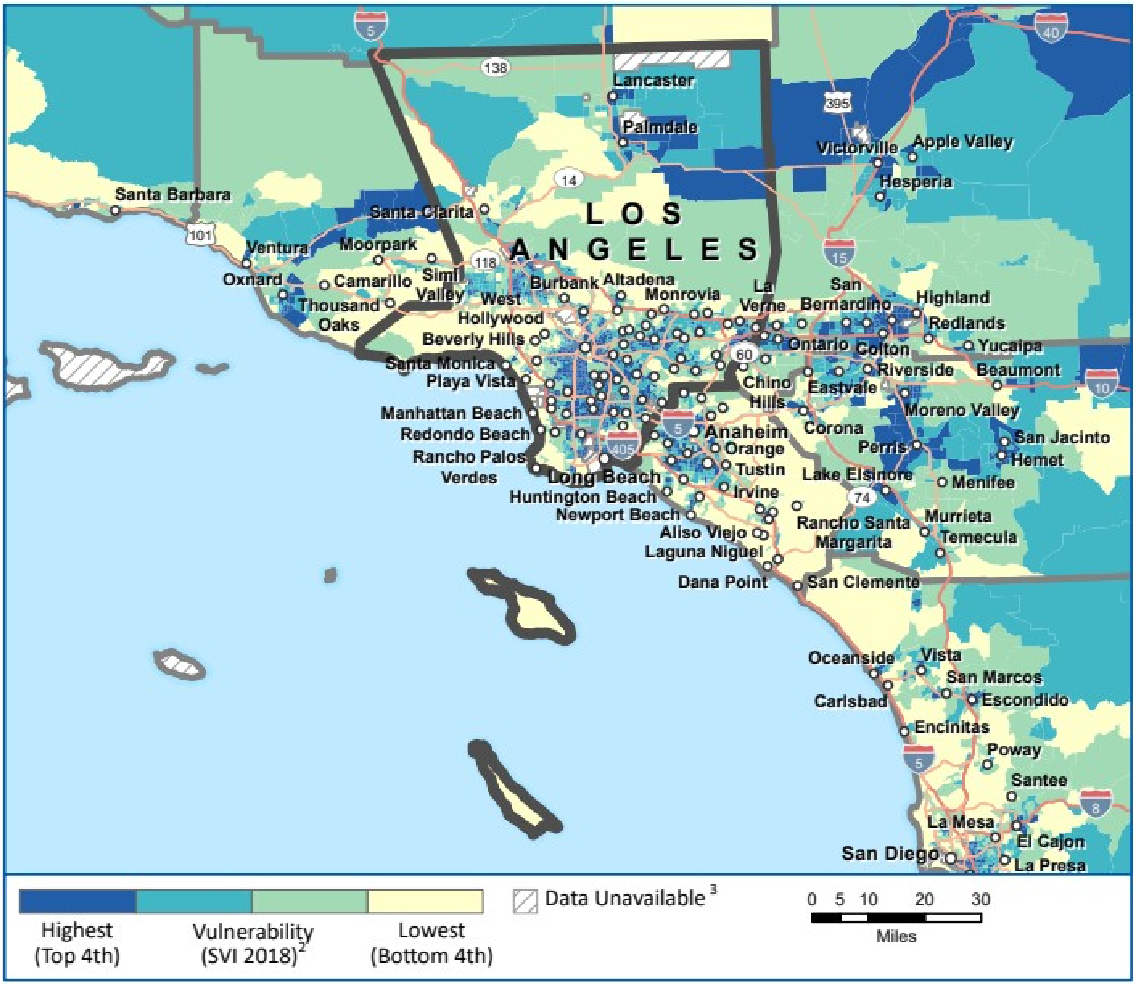

Figure 2.5 Overall Social Vulnerability (CDC developed Social Vulnerability 2018)

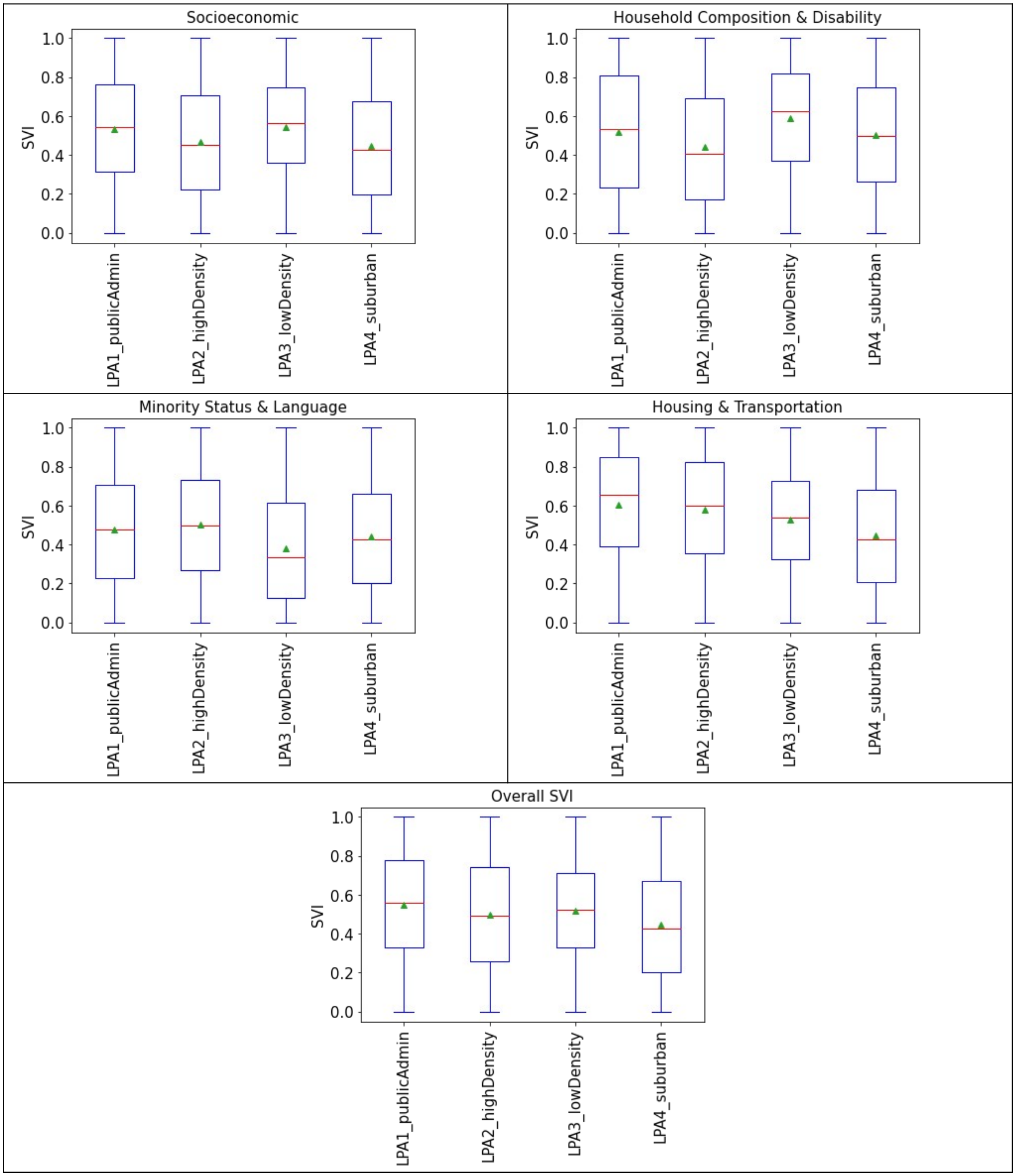

Figure 2.6 CDC SVI and LPA relationships

3. Quantile Regression Analysis of VMT

3.1 Introduction

Vehicle miles traveled (VMT) is a primary policy variable because it is strongly correlated with greenhouse gas (GHG) emissions and density and diversity of urban environments (Rentziou et al., 2012; Brownstone, 2008; Cervero and Kockelman, 1997). Policies that increase density and diversity of land use are believed to decrease VMT presumably by replacing it with personal miles walking and using bicycles (Rentziou et al., 2012; Ewing and Cervero, 2010, 2001). This causal inference has not been conclusive and may create further disparities by gentrification and increase in long-distance travel for lower-income commuters.

Existing studies on the relationship between VMT and the built environment mostly rely on single point estimates (Salon, 2015) and rarely consider the heterogeneity across segments of the population in terms of varying travel intensities. This might bias the conclusions as people at different travel intensities may present different sensitivity to land use. Developing models that can capture the heterogeneous impacts of built environment characteristics on VMT is essential for informing more targeted and efficient policies for populations of interest.

To fill this gap, we explore the heterogeneity in the relationships between personal level VMT and the built environment, accessibility to opportunities and open spaces, and people's socio-demographic traits across various levels of travel demand. The present research first develops opportunity-based measurement of residential land-use indicators at the US census block level for the whole California using data from Dunn&Bradstreet inventory of business establishments classified into 14 distinct industry categories. The analysis here also uses the LPA findings from Chapter 2 that was used to identify distinct land use patterns for the 691,721 US Census blocks delineating residential built environment characteristics. Using the same data, we create time-based buffered accessibility indicators for each block to compute access to businesses and their employment and access to open space. Then, quantile regression models (Koenker and Bassett Jr, 1978; Koenker and Hallock, 2001) are estimated using the data from the 2012-2013 California Household Travel Survey (CHTS) to understand the heterogeneity in the relationships between VMT and residential built environment characteristics and people's socio-demographic traits. Quantile regressions produce estimates of the correlation between an explanatory variable and a dependent variable at different levels or quantiles of the dependent variable and help us identify different sensitivity to land use across various travel intensities.

The contribution of this research is a focus on the heterogeneity across various travel intensities in the relationship between different types of VMT and residential built environment characteristics. This research challenges the use of VMT as the right policy variable because our findings indicate that the correlation of VMT with land use is substantially different among segments of the population at different levels of travel intensities.

3.2 Literature review

California legislation identifies VMT as a key performance parameter to control in decreasing GHG emissions because there is a direct relationship from miles driving a car to fuel consumption and to CO2 emissions (Rentziou et al., 2012). Legislation supports and promotes VMT decrease using land use policies (e.g., correlating VMT to residential density as in Boarnet and Handy, 2014). The California Air Resources Board (CARB) in its recommendation for policy implementation of legislative initiatives and its scoping plan uses VMT without distinguishing VMT accumulated driving alone, serving passengers, or being a passenger (https://ww2.arb.ca.gov/resources/documents/carb-2017-scoping-plan-identified-vmt-reductions-and-relationship-state-climate). This may lead to unintended consequences of policy implementation harming the vulnerable. For instance, people who cannot drive getting car rides to doctor appointments and captives to long-distance commutes due to jobs-housing lack of balance and costs (Mitra and Saphores, 2019). In fact, most of the existing literature only models the aggregated VMT instead of breaking it down into VMT driving alone, serving passengers, or being a passenger (see the reviews in Ewing and Cervero, 2010, and 2001).

VMT is also used as the key impact indicator in the relatively new impact analysis guidelines for the California Environmental Quality Act (CEQA - https://opr.ca.gov/ceqa/ - California governor's office of planning and research, 2018). Implementation of these guidelines in local jurisdictions does not distinguish among different types of VMT people accumulate in a day and does not identify possible differences in the behavioral sensitivity to the built environment changes among different people (see the example in The City of Santa Rosa Transportation and Public Works Department, 2020).

The literature on the relationship between travel and the built environment offers evidence of substantial variation found across different settings and social strata (see the series of research reports by the National Center for Sustainable Transportation, https://ncst.ucdavis.edu/tags/vehicle-miles-traveled). Ewing and Cervero (2001) summarized the most common travel outcome variables in the literature on travel and built environment and they are trip frequency, trip length, mode choice, and VMT as a composite measure of travel demand. The built environment is very often measured by the "five Ds" in the travel behavior literature including density, diversity, design (the original "three Ds" by Cervero and Kockelman, 1997), destination accessibility and distance to transit (Ewing and Cervero, 2010, 2001; Su et al., 2021b). *Density* is measured as the variable of interest per unit of area, such as population density, housing unit density, job density, etc. Previous studies demonstrated that higher residential density is negatively correlated with VMT (Boarnet and Handy, 2014; Ewing and Cervero, 2010; Ewing, 1997). For example, National Research Council research revealed that doubling residential density is associated with VMT reductions ranging from 5 to 12 percent (an elasticity of -0.05 to -0.12, see National Research Council, 2009). *Diversity* reflects the degree of mixed land uses. Entropy is frequently used to measure the diversity of land use. Higher entropy of land use is associated with lower car usage and higher non-motorized travel (Ewing et al., 2014; Ewing and Cervero, 2010; Cervero and Duncan, 2006; Kockelman, 1997). *Design* measures the street network characteristics in a neighborhood (Ewing and Cervero, 2001, 2010). Examples include average block size and intersection/street density. The variables of neighborhood design are associated with reductions in VMT because smaller block sizes and higher intersection or street densities can reduce the network distance to destinations and make other non-motorized modes more convenient (Nelson, 2017; Ewing and Cervero, 2010; Bento et al., 2005). *Destination accessibility* measures the ease of reaching destinations (Ewing and Cervero, 2001). The commonly used variables are the distance to the central business district and the number of job opportunities or other attractions within a given travel time (Ewing and Cervero, 2010). Higher accessibility to destinations may reduce VMT or encourage the use of other modes instead of car (Nelson, 2017; Ewing and Cervero, 2010). *Distance to transit* reflects the average shortest network distance from home locations or workplaces to the nearest transit stop (Ewing and Cervero, 2010). People living closer to transit may reduce VMT and presumably substitute transit trips for vehicle trips (Salon et al., 2012; Ewing and Cervero, 2010).

Existing literature mostly relies on single point estimates (Salon, 2015) and rarely considers the heterogeneity across segments of the population in terms of their varying travel intensities. This might be biased as people at different travel intensities may present different sensitivity to the built environment factors. To address this issue, we develop quantile regression models to capture the heterogeneous impacts of built environment characteristics on VMT across various travel demands. In addition, VMT is categorized as VMT driving alone, serving passengers, or being a passenger in order to distinguish the impacts of the built environment on different types of VMT.

3.3 Data used in this chapter

3.3.1 Accessibility indicators

Using the business establishment data described in Chapter 2, we create time-based buffered accessibility indicators for each block of in California to compute access to businesses and their employment given varying travel times by driving between block centroids. Travel time between each pair of blocks was computed based on the network distance between block centroids and the maximum speed limit of each road segment using the road network data from Open Street Map (https://www.openstreetmap.org). The routing algorithm was implemented mainly using the python packages OSMnx (Boeing, 2017) and NetworkX (Hagberg et al., 2008). In addition, open space accessibility was generated in this project using the open-access data of the California Protected Areas Database (https://www.calands.org/). Depending on its distance to the California coast, open space can be classified into three categories: ocean-view (no more than 100 meters from the coast), near-ocean (no more than 1000 meters from the coast), and far-ocean (more than 1000 meters from the coast) open spaces. Open space accessibility (at census block level) is calculated as the total area (in acres) of open space reachable by driving within certain time, similar to the employment accessibility.

3.3.2 Household travel survey

We use the data of the 2012-2013 California Household Travel Survey 2012-13 (NuStats, 2013), which was collected from February 1st, 2012 to January 31st, 2013. Every member in each participating household reported their individual and household-level socio-demographic information as well as a single-day travel diary from 3:00 AM on the assigned survey day to 2:59 AM on the following day. In the single-day travel diary, each participant reported their origin, destination, travel mode, trip distance, start time and end time of each trip. The CHTS data used in this study in total has 105,204 respondents from 41,495 households. According to the CHTS codebook, people's travel mode choices are classified into seven categories including walk, bike, transit, car drive alone, car drive someone else, car as passengers, and one category labeled "other" comprising all other modes. Table 3.1 summarizes the descriptive statistics for the variables that are used in further analysis. Total VMT for each person over the survey day is calculated by summing up the distances of all trips in private vehicles. We also break down VMT by VMT drive alone, VMT drive someone else, and VMT as passengers. As Table 3.1 shows, the average total VMT in a day in this sample is 19.28 miles. The mean VMT drive alone is 8.28 miles, the mean VMT drive someone else is 4.61 miles and the mean VMT as passengers is 6.40 miles. Among these people, 31,576 (30.01%) of the sample did not report any trip by vehicle in a day. The survey respondents only travel 0.93 miles taking transit and 0.34 miles by active modes (i.e., walk and bike) on average in a day.

Table 3.1 Descriptive statistics for the 2013 CHTS

(a) Continuous variables	Mean	Median	S.D.	Min	Max
Household size	3.31	3	1.55	1	8
Number of vehicles	2.07	2	1.02	0	8
VMT total	19.28	6.60	38.51	0	1285.40
VMT drive alone	8.28	0	22.55	0	1285.40
VMT drive someone else	4.61	0	21.06	0	844.66
VMT as passengers	6.40	0	25.30	0	844.66
PMT[a] as transit	0.93	0	8.67	0	779.10
PMT as active modes	0.34	0	2.87	0	309.66

(b) Categorical variables	Subgroup	Percentage
Sex	Female	51.29%
Age group	≤ 17	11.86%
	18-24	6.80%
	25-34	8.10%
	35-50	21.43%
	51-65	30.73%
	≥ 66	21.08%
Education attainment	Below bachelor's degree	50.33%
	Some college or associate's degree	13.85%
	Bachelor's degree or above	33.84%
Employment status	Yes	47.32%
Household income	Less than $24,999	14.06%
	$25,000 to $49,999	18.04%
	$50,000 to $99,999	29.95%
	$100,000 to $199,999	23.46%
	$200,000 or more	6.42%
Household structure	Have children ages ≤ 3	10.73%
	Have children ages 4-15	37.36%
	Have children ages 16-18	15.11%

Note: [a]PMT=Personal Miles Traveled. The percentage of the alternative option for some binary variables (e.g., sex, employment status) are omitted. The total percentage of some variables is not 100% because the corresponding survey question does not apply or participants refused to answer (below 1%).

3.3.2 Built environment and open space

The built environment is a structural determinant affecting people's travel behavior such as VMT (Handy et al., 2005). In Chapter 2 recall we used business establishment data to compute summary accessibility indicators and open space opportunities to compute summaries of access to open space for each US census block. These accessibility indicators were then spatially joined with the household locations of the survey respondents and used as explanatory variables in the regression models to explain the variation of person-level daily VMT. Given the high correlations among accessibility indicators across different travel times, we excluded those highly correlated variables and only included in the regression models accessibility indicators within 20 minutes of travel time. We chose 20 minutes because the average travel time of a trip in the survey is 21.32 minutes with a standard deviation of 32.48 minutes. Table 3.2 presents the descriptive statistics for these accessibility indicators for the 105,204 CHTS respondents. *Ocean view_20min*, *Far ocean_20min*, and *Near ocean_20min* denote the total accessible area of each of these three types of open spaces in 20 min by driving respectively. *Population density_20min* is the total population that is accessible within 20 min driving time over the corresponding total area of accessible blocks. We also spatially join the place of residence of each household with the LPA classification, developed and presented in Chapter 2, of the block within which the household resides.

Table 3.2 Descriptive statistics for the time-based buffered accessibility indicators for the 105,204 CHTS respondents

Variables	Mean	Median	S.D.	Min	Max
Ocean view 20min	10.843	0.000	25.093	0	245.591
Far ocean 20min	112.065	82.388	112.922	0	3539.250
Near ocean 20min	3.971	0.000	7.868	0	68.079
Population density 20min	1.107	0.736	1.105	0	27.137

Note: Ocean view 20min, Far ocean 20min, and Near ocean 20min are in acres (approximately 4,047 m^2) and divided by 10^6 to scale the estimates of this variable to the same level as other variables; Population density 20min is the population per m^2 and multiplied by 1000 to scale the estimates of this variable to the same level as other variables.

3.4 Quantile regression

Many studies on travel demand rely on ordinary least squares (OLS) regression to estimate the conditional means of the model parameters. It is limited to capturing the heterogeneous effects between explanatory variables and travel demand for people who travel less or more than the average (Su, 2012). Quantile regression (Koenker and Bassett Jr, 1978, Koenker and Hallock, 2001) is a more robust approach for understanding the source of heterogeneity in travel behavior as it models the relationship between explanatory variables and various quantiles of the VMT. In addition, OLS for hypothesis testing assumes the error terms are normally distributed while quantile regression does not make any assumptions on the distribution of the error terms (Koenker and Hallock, 2001). Therefore, quantile regression is a more appropriate approach for the present study on the heterogeneous impacts of built environment variables on VMT. Quantile regression seeks the coefficients b that minimize the objective function below.

$$\min_{\beta \in R^k} \sum_{i \in \langle i: y_i \geq x_i \beta \rangle}^{n} \theta \left| y_i - x_i' \beta \right| + \sum_{i \in \langle i: y_i < x_i \beta \rangle}^{n} (1-\theta) \left| y_i - x_i' \beta \right| \quad \text{(Equation 3.1)}$$

In this study, y_i is the vector of person-level daily VMT, x is the vector of all regressors including people's sociodemographic characteristics, time-based buffered accessibility indicators, and a categorical variable representing different latent profiles of residential built environment characteristics, β is the vector of parameters to be estimated, θ is between 0 and 1 indicating a θ^{th} quantile of the distribution of VMT. The parameter β for a given q can be estimated efficiently using linear programing methods (Koenker and Hallock, 2001). We use the R package quantreg (Koenker et al., 2022) to estimate the quantile regression.

3.5 Experiments

3.5.1 Correlation between LPA and VMT of the built environment characteristics

As an exploratory analysis, we look at the box plots of different types of VMT for the portion of respondents with a daily VMT greater than zero. Figure 3.1 shows that on average, residents in *LPA3_lowDensity* blocks have the highest total VMT (32.17 miles) as well as VMT drive someone else (7.94 miles) and VMT as passengers (11.06 miles). Presumably, people who live in low-density areas rely mainly on personal vehicles for their daily travel and they are very likely to travel with somebody instead of driving alone. People living in high-density areas (*LPA2_highDensity*) generate 26.38 VMT on average in a day which is 5.79 miles less than people in *LPA3_lowDensity*. In *LPA4_suburban*, we find even lower daily VMT (25.91 miles) compared to all other groups. *LPA1_publicAdmin* has relatively high daily VMT (30.87 miles) and also the highest VMT drive alone (15.81 miles) in comparison with other groups. Notable is the substantial difference between the average VMT and the median VMT for all groups as well as the shape of the distribution further supporting our selection of quantile regression as the appropriate data analysis technique to understand heterogeneity. In subsequent analysis, we further explore the propensity of making VMT and the heterogeneity in the relationships of VMT with the built environment characteristics, accessibility to opportunities and open spaces, and individual and household level socio-demographic traits.

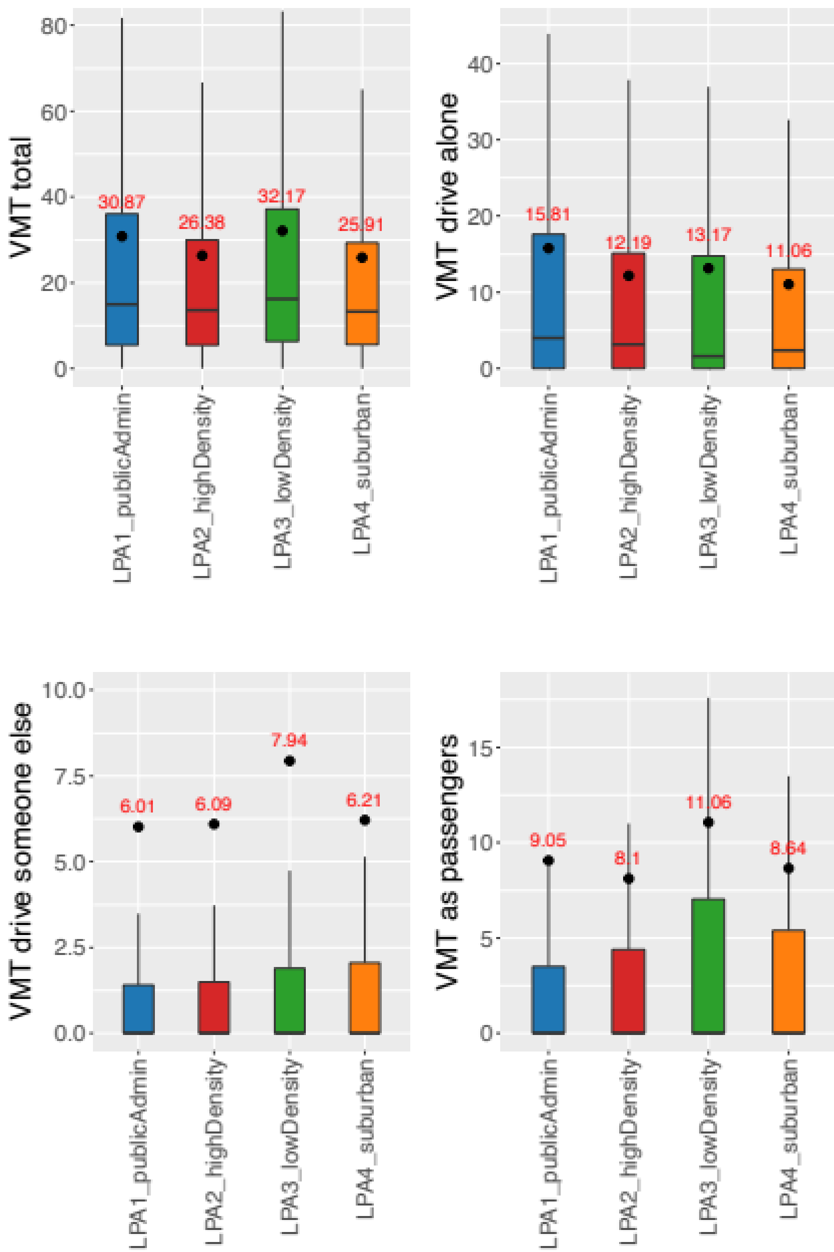

Figure 3.1 Box plots of four types of VMT by four latent profiles for the portion of respondents with daily VMT greater than zero. The black dots and the annotated numbers in red represent the mean.

3.5.2 Comparison of people with zero VMT

Among the 105,204 CHTS respondents, 31,576 (30.01%) reported zero VMT. We estimate a binary logistic regression model to identify significant factors that are associated with respondents who have non-zero VMT. To avoid multicollinearity, Variance Inflation Factor (VIF) was checked for each explanatory variable. Based on VIF, variables that are highly correlated with other variables were excluded in the final model. The final specification of the model was obtained by a systematic process of eliminating insignificant variables. Table 3.3 summarizes the outcome of the estimated binary logistic regression model. All of the explanatory variables are significant except the two dummy variables indicating if respondents have children aged under 4 or aged 16-18. According to the model, women are more likely to have VMT greater than zero compared with their male counterparts. Respondents aged 35 to 50 are most likely to have VMT greater than zero and this is followed by respondents aged 51 to 65, aged under 18, aged above 65, aged 25 to 34, and aged 18 to 24. The higher the educational attainment, the more likely people are to have non-zero VMT. Higher household income is also positively associated with the probability of having VMT greater than zero. Employed respondents tend to have VMT greater than zero compared to non-employees. The more vehicles a household has, the more likely an individual in the household to have VMT greater than zero. In terms of built environment variables, we found higher accessibility to ocean-view open space contributes to a higher probability of creating VMT while higher accessibility to near-ocean open space reduces the probability. People living in more densely populated areas are less likely to have VMT greater than zero. The estimates of the three latent profile dummy variables show that people living in *LPA4_suburban* are most likely to have VMT greater than zero and this is followed by *LPA1_publicAdmin*, *LPA2_highDensity*, compared to *LPA3_lowDensity*.

Table 3.3 Binary logit regression results using as dependent variable a binary variable indicating if respondents have VMT greater than zero on the survey day

Variables	Estimates (standard errors)
Female (base: not female)	0.077*** (0.014)
Age group (base: above 65)	
age under18	0.077*** (0.028)
age 18 to 24	-0.152*** (0.031)
age 25 to 34	-0.059* (0.030)
age 35 to 50	0.100*** (0.024)
age 51 to 65	0.096*** (0.020)
Education (base: below bachelor)	
some college	0.173*** (0.022)
above bachelor	0.412*** (0.019)
Household annual income (base: less than $25k)	
$25k to $50k	0.174*** (0.021)
$50k to $100k	0.387*** (0.020)
$100k to $200k	0.543*** (0.023)
more than $200k	0.600*** (0.036)
Employed (base: not employed)	0.762*** (0.017)
Household children structure (base: no children)	
has children aged under 4	NS
has children aged 4 to 15	0.412*** (0.018)
has children aged 16 to 18	NS
# of vehicles	0.207*** (0.008)
Ocean view 20min	0.002*** (0.0003)
Near ocean 20min	-0.003*** (0.001)
Population density 20min	-0.028*** (0.007)
Residential built environment (base: LPA3 lowDensity)	
LPA1 publicAdmin	0.093** (0.044)
LPA2 highDensity	0.060** (0.026)
LPA4 suburban	0.154*** (0.017)
Constant	-0.659*** (0.027)
Model statistics	
Number of observations	105,204
Log-likelihood Unrestricted	-59,874
Log-likelihood Ratio	8,808*** (df=23)
McFadden's Pseudo R^2	0.069

Note: *p<0.1; **p<0.05; ***p<0.01

3.5.3 Quantile regressions on VMT

To understand the heterogeneity in the relationship of VMT with the built environment characteristics, in this section we discuss the results from the quantile regressions on VMT and compare them with the traditional OLS models that estimate the conditional mean of VMT. Table 3.4 and Table 3.5 display the outcomes of the mean models and the median quantile regression models using as dependent variables VMT total, VMT drive alone, VMT drive someone else, and VMT as passenger respectively. It is noteworthy that in each model we excluded people with zero VMT as this can be a large portion of the sample and will bias the coefficient estimates. Figure 3.2 and Figure 3.3 show several selected variables and their estimates using quantile regression from 0 to 1 quantile using VMT total as the dependent variable.

First, we report the differences between mean and median models in terms of the significance of the estimates. In general, mean and median models present high similarity in statistical inference/significance but a few notable differences. Between models (1) and (2) that use total daily VMT as the dependent variable, we found that the dummy variable `age 18 to 24' is not significant in the mean model but becomes significant at 0.05 level in the median model. The coefficient is 0.608 indicates that for people with daily VMT at the median level, people aged 18 to 24 create 0.608 more VMT compared to people aged above 65. The estimate of the dummy variable `has children aged under 4' is significant in the mean model but is insignificant in the median model. The coefficient is -1.184 indicating that on average, people with children aged under 4 create 1.184 less VMT in a day compared with people without children. The dummy variable `LPA1_publicAdmin' is not significant in the mean model but is significant in the median model. The coefficient is -1.436 indicating that at the median level of daily VMT, people living in LPA1_publicAdmin create 1.436 less VMT compared with people living in LPA3_lowDensity. Between models (3) and (4) using VMT drive alone as the dependent variable, the coefficient of the dummy variable `some college' is 0.568 and significant in the median model but not significant in the mean model. The dummy variable `has children aged under 4' is insignificant in the mean model but becomes significant in the median model. `Ocean view_20min' and `Population density_20min' are only significant in the mean model. Between models (5) and (6) that use VMT drive someone else as the dependent variable, `age 18 to 24' is only significant in the mean model. A few dummy variables including `age 35 to 50', `above bachelor', `$50k to $100k', `more than $200k', and `has children aged under 4' are only significant in the median model but just at 0.1 level. Between models (7) and (8) that use VMT as passenger as the dependent variable, dummy variables including `age 35 to 50', `$50k to $100k', `has children aged under 4', and `LPA1_publicAdmin' are significant factors impacting VMT as passenger only in the median model.

Table 3.4 Estimation results of OLS versus quantile regressions on VMT total and VMT drive alone

	Dependent variable:			
	VMT total		VMT drive alone	
	(1)Mean	(2)Median	(3)Mean	(4)Median
Socio-demographic variables				
Female (base: not female)	−2.898***(0.319)	−1.217***(0.148)	−5.495***(0.312)	−2.493***(0.163)
Age group (base: above 65)				
age under18	−3.367***(0.676)	−1.761***(0.233)	−4.011***(1.552)	−0.917** (0.372)
age 18 to 24	NS	0.608** (0.292)	NS	NS
age 25 to 34	1.759** (0.712)	2.176*** (0.326)	2.069*** (0.705)	1.808*** (0.376)
age 35 to 50	2.938*** (0.553)	2.879*** (0.265)	2.329*** (0.560)	1.599*** (0.256)
age 51 to 65	2.075*** (0.488)	1.869*** (0.230)	1.676*** (0.477)	1.253*** (0.202)
Education (base: below bachelor)				
some college	2.690*** (0.511)	1.988*** (0.253)	NS	0.568** (0.235)
above bachelor	3.143*** (0.405)	2.495*** (0.207)	0.675* (0.380)	0.717*** (0.180)
Household annual income (base: less than $25k)				
$25k to $50k	NS	NS	NS	NS
$50k to $100k	1.688*** (0.479)	1.248*** (0.212)	1.283*** (0.482)	0.947*** (0.215)
$100k to $200k	3.344*** (0.511)	2.295*** (0.235)	1.928*** (0.514)	1.450*** (0.237)
more than $200k	3.731*** (0.723)	3.178*** (0.328)	2.637*** (0.711)	3.174*** (0.390)
Employed (base: not employed)	4.249*** (0.386)	4.737*** (0.190)	5.972*** (0.380)	4.725*** (0.168)
# of vehicles	1.624*** (0.179)	1.134*** (0.087)	0.893*** (0.174)	0.528*** (0.093)
Household children structure (base: no children)				
has children aged under 4	−1.184** (0.563)	NS	NS	−1.281***(0.341)
has children aged 4 to 15	−1.360***(0.392)	−0.681***(0.179)	−2.505***(0.410)	−1.744***(0.196)
has children aged 16 to 18	NS	NS	−0.952* (0.504)	−0.588** (0.258)
Built environment variables				
Ocean view 20min	−0.032***(0.007)	−0.013***(0.003)	−0.021***(0.006)	NS
Near ocean 20min	−0.076***(0.022)	−0.049***(0.010)	−0.055***(0.021)	−0.032***(0.008)
Population density 20min	−2.812***(0.158)	−0.508***(0.048)	−1.504***(0.157)	NS
Residential built environment (base: LPA3 lowDensity)				
LPA1 publicAdmin	NS	−1.436***(0.534)	NS	NS
LPA2 highDensity	−2.758***(0.593)	−2.467***(0.291)	−2.413***(0.577)	−1.792***(0.303)
LPA4 suburban	−5.063***(0.405)	−2.952***(0.217)	−4.006***(0.404)	−2.146***(0.226)
Constant	26.852***(0.695)	9.993*** (0.330)	20.146***(0.730)	8.759*** (0.333)
Observations	73,628	73,628	41,390	41,390

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3.4 Estimation results of OLS versus quantile regressions on VMT drive someone else and VMT as passenger

	Dependent variable:			
	VMT total		VMT drive alone	
	(1)Mean	(2)Median	(3)Mean	(4)Median
Socio-demographic variables				
Female (base: not female)	−5.324∗∗∗ (0.579)	−1.321∗∗∗(0.203)	1.650∗∗∗ (0.522)	0.577∗∗∗ (0.172)
Age group (base: above 65)				
age under18	−10.430∗∗∗(2.827)	−2.967∗∗∗(0.500)	−1.633∗∗(0.819)	−0.598∗∗(0.247)
age 18 to 24	−3.439∗∗ (1.604)	NS	1.975∗ (1.145)	0.922∗∗ (0.400)
age 25 to 34	NS	NS	NS	NS
age 35 to 50	NS	0.607∗ (0.354)	NS	0.760∗∗∗ (0.277)
age 51 to 65	1.646∗ (0.906)	1.464∗∗∗ (0.336)	NS	NS
Education (base: below bachelor)				
some college	NS	NS	5.500∗∗∗ (1.022)	1.466∗∗∗ (0.477)
above bachelor	NS	0.407∗ (0.242)	2.784∗∗∗ (0.800)	0.973∗∗∗ (0.325)
Household annual income (base: less than $25k)				
$25k to $50k	NS	NS	NS	NS
$50k to $100k	NS	0.479∗ (0.286)	NS	1.012∗∗∗ (0.257)
$100k to $200k	1.586∗ (0.945)	0.615∗∗ (0.303)	1.816∗∗ (0.813)	1.135∗∗∗ (0.267)
more than $200k	NS	0.784∗ (0.435)	2.771∗∗ (1.155)	1.310∗∗∗ (0.384)
Employed (base: not employed)	NS	NS	3.415∗∗∗ (0.724)	1.236∗∗∗ (0.297)
# of vehicles	1.100∗∗∗ (0.334)	0.513∗∗∗ (0.125)	2.018∗∗∗ (0.300)	0.927∗∗∗ (0.110)
Household children structure (base: no children)				
has children aged under 4	NS	0.599∗ (0.345)	0.415∗ (0.249)	NS
has children aged 4 to 15	−4.994∗∗∗ (0.686)	−2.231∗∗∗(0.230)	−3.047∗∗∗(0.638)	−1.264∗∗∗(0.219)
has children aged 16 to 18	−2.724∗∗∗ (0.822)	−0.898∗∗∗(0.258)	−1.192∗ (0.703)	−0.885∗∗∗(0.215)
Built environment variables				
Ocean view 20min	−0.028∗∗ (0.012)	−0.014∗∗∗(0.003)	−0.038∗∗∗(0.011)	−0.014∗∗∗(0.003)
Near ocean 20min	−0.090∗∗ (0.040)	−0.034∗∗∗(0.009)	−0.062∗ (0.037)	−0.045∗∗∗(0.011)
Population density 20min	−3.087∗∗∗ (0.290)	−0.477∗∗∗(0.077)	−3.083∗∗∗(0.253)	−0.442∗∗∗(0.041)
Residential built environment (base: LPA3 lowDensity)				
LPA1 publicAdmin	−4.653∗∗ (1.830)	−3.046∗∗∗(0.787)	NS	−2.926∗∗∗(0.762)
LPA2 highDensity	−2.128∗∗ (1.077)	−2.429∗∗∗(0.401)	−3.114∗∗∗(0.968)	−3.550∗∗∗(0.344)
LPA4 suburban	−4.868∗∗∗ (0.733)	−2.656∗∗∗(0.329)	−4.837∗∗∗(0.655)	−3.303∗∗∗(0.264)
Constant	31.849∗∗∗ (1.310)	11.508∗∗∗(0.510)	25.047∗∗∗(1.116)	10.008∗∗∗(0.397)
Observations	21,430	21,430	28,724	28,724

Note: ∗p<0.1; ∗∗p<0.05; ∗∗∗p<0.01

Pacific Southwest Region UTC
University Transportation Center

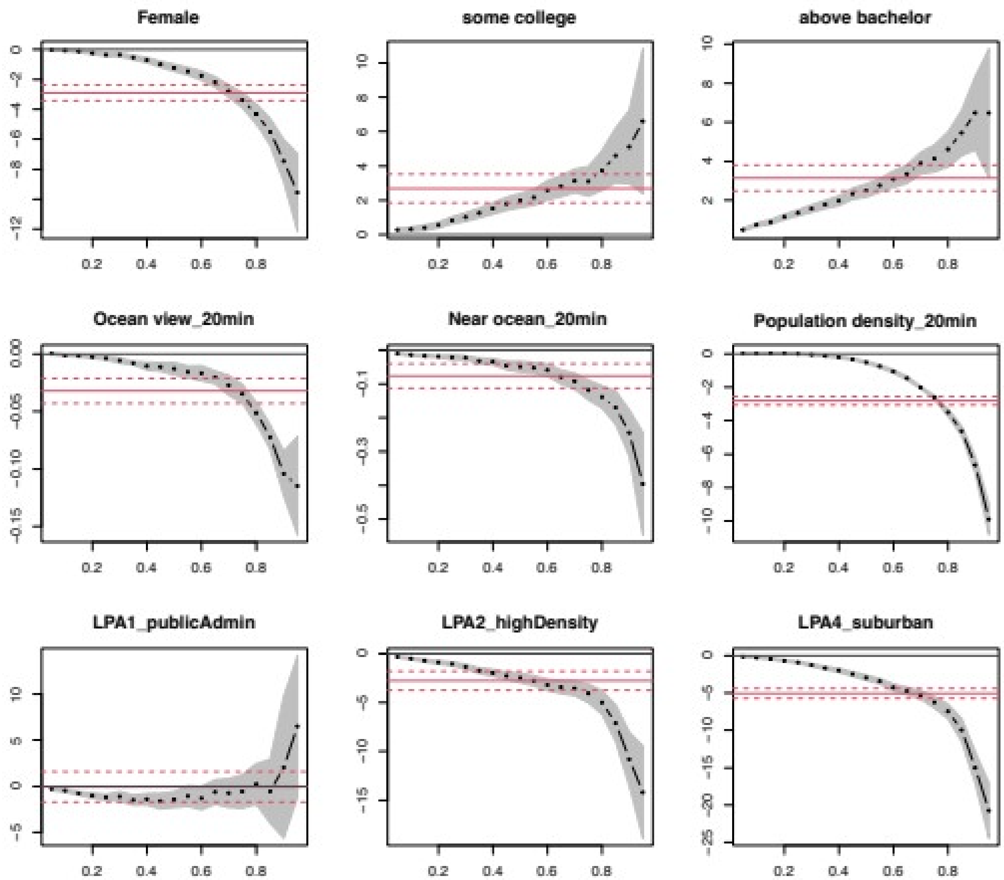

Figure 3.2: Partial outcomes of quantile regression using total VMT as the dependent variable. The x axis represents the quantile from 0 to 1. The y axis shows the coefficient estimates.

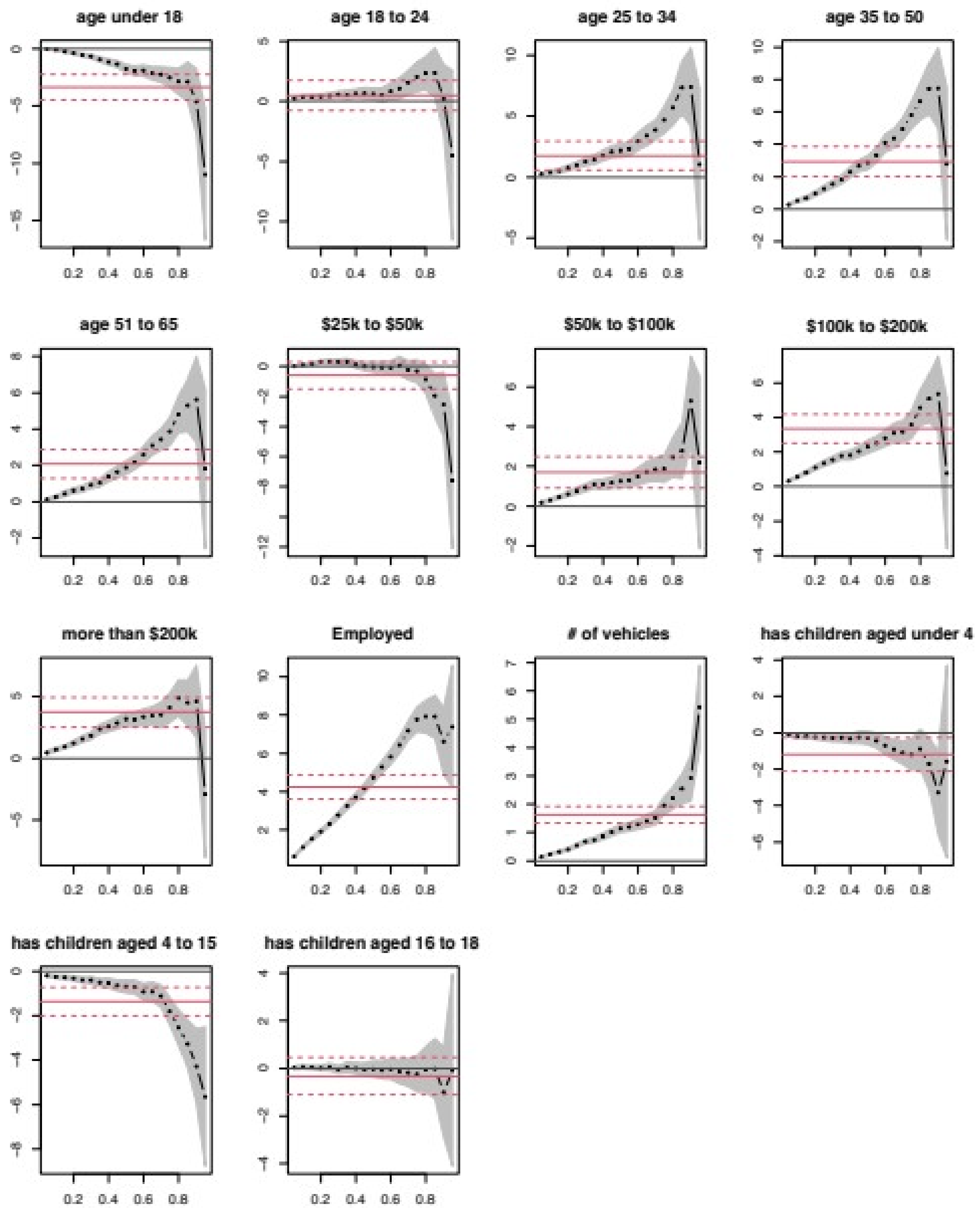

Figure 3.3 Partial outcomes of quantile regression using total VMT as the dependent variable. The x axis represents the quantile from 0 to 1. The y axis shows the coefficient estimates.

The signs of the coefficients that are both significant in the mean and median models are consistent. However, there are several main differences regarding the size of the coefficients between the mean and median models. On average, females create 2.898 less VMT in a day than their male counterparts. Among the survey participants with VMT drive alone greater than zero, we found females travel 5.495 miles less than males. In addition, females make 5.324 less VMT drive someone else and 1.650 more VMT as passenger on average. However, the median model (2) shows that females with daily VMT at the 0.5 quantile only travel 1.217 miles less than males. As shown in Figure 3.2, females with VMT at even higher quantiles such as above 0.8 travel far less than males (the coefficient is ranging from -4.315 to -9.551 miles). The median models (4), (6), and (8) also show large discrepancies compared to the corresponding mean model. Females with each type of VMT at median quantile make 2.493 less VMT drive alone, 1.321 less VMT drive someone else, and 0.577 more VMT as passenger compared to males. The estimates of the dummy variables of age groups indicate that respondents under 18 make less VMT regardless of the type of VMT. The mean models indicate very large differences between respondents under 18 and the reference group of age above 65, but the median models show smaller differences. This implies that simply relying on estimating the conditional mean might exaggerate the differences in VMT between different age groups. Both the mean and median models show that the age groups 25 to 34 make more VMT in a day significantly than the reference group.

Participants with higher educational attainment tend to make more VMT total, VMT drive alone, and VMT as passenger and they do not have a large difference in VMT drive someone else in comparison to the below bachelor's group. The coefficients of education dummy variables in the mean model (1) are greater than the ones in the median model (2). As Figure 3.2 shows, participants with higher educational attainment with VMT at higher quantiles make substantially more VMT in a day than the below bachelor's group. In terms of VMT drive alone, models (3) and (4) show that the above bachelor's group makes 0.675 more VMT drive alone on average and 0.717 more VMT drive alone at the 0.5 quantile. Model (7) shows that on average, people with some college education make 5.5 more VMT as passenger and people with above bachelor's education make 2.784 more VMT as passenger. However, the median model (8) only shows smaller differences compared to the below bachelor's group. It is 1.466 and 0.973 more VMT as passenger for some college's and above bachelor's groups respectively.

Higher-income groups and employed people are also likely to make more VMT total, VMT drive alone, and VMT as passenger in comparison to the lower-income group and non-employees respectively. In general, people in households with a higher number of vehicles make more VMT regardless of the type of VMT. In terms of household structure, both the mean and median models show that people who have children aged 4 to 15 make less VMT significantly regardless of the type of VMT than people who have no children. However, the coefficients between mean and median models are different. In particular, the mean model (5) suggests that people who have children aged 4 to 15 make 4.994 less VMT drive someone but the median model (6) only shows a 2.231 difference. The coefficient in the mean model (7) is −3.047 and in the median model (8) is −1.264. These results again tell us the large bias when relying on estimating exclusively the conditional mean and overlooking the heterogeneity of the relationships at different quantiles. In addition, people who have children aged 16 to 18 do not have a significant difference in daily VMT compared to people without children. But they tend to create less VMT drive alone, VMT drive someone else, and VMT as passenger. The corresponding coefficients of `has children aged 16 to 18' from mean and median models also present differences but not as much as the coefficients of `has children aged 4 to 15'.

Next, we discuss the impact of built environment variables on people's daily VMT. The results of mean models show that people with higher accessibility to ocean-view and near-ocean open spaces as well as higher accessibility to population density tend to make less VMT regardless of the type of VMT. But not all coefficients of these three built environment variables are significant in the median models (see the interpretation above). The sizes of the coefficients of these three built environment variables in the mean models are consistently larger than the ones in the median models. Particularly, the coefficients of `Population density_20min' in the mean models have substantially large sizes compared to the ones in median models. Figure 3.2 also shows that participants who have higher accessibility to ocean-view and near-ocean open spaces, as well as population density with VMT at higher quantiles, make substantially less VMT in a day. This implies that for those with higher travel demand, higher accessibility to ocean-view and near-ocean open spaces and population density can result in less VMT.

In terms of the four latent profiles of residential built environment characteristics, the mean model (1) suggests no significant difference in VMT total between people in LPA1_publicAdmin and people in LPA3_lowDensity. But the coefficient is -1.436 and significant in the median model (2), which indicates that at the median level of VMT total, people living in LPA1_publicAdmin create 1.436 less VMT than people living in LPA3_lowDensity. According to the models (1) and (2), people living in LPA2_highDensity create 2.758 less VMT in a day on average and 2.467 less VMT in a day at the 0.5 quantile, and people living in LPA4_suburban make 5.063 less VMT in a day on average while the median model only shows 2.952 miles difference. Similarly, the other mean and median models also suggest that people living in LPA2_highDensity and LPA4_suburban create less VMT drive alone, VMT drive someone else, and VMT as passenger compared to people living in LPA3_lowDensity. Models (3) and (4) indicate no significant difference in VMT drive alone between people living in LPA1_publicAdmin and people living in LPA3_lowDensity. Models (5) and (6) suggest that people living in LPA1_publicAdmin create less VMT drive someone else than people living in LPA3_lowDensity. According to models (7) and (8), the coefficient of LPA1_publicAdmin is only significant in the median model, which means at the 0.5 quantile of VMT as passenger, people living in LPA1_publicAdmin create 2.926 less VMT as passenger than people living in LPA3_lowDensity. Among all these models, we observe large differences in the size of coefficients between mean and median models. In particular, the size of the coefficients of LPA4_suburban in mean models is larger than the ones in median models. These results again imply the heterogeneity in the relationships between residential built environment characteristics and people's daily VMT which is neglected in the traditional approaches that model the conditional mean.

3.6 Conclusions and policy implications

The first portion of the research presented here leverages binary logit regression to identify significant factors that are associated with survey respondents who have non-zero VMT. The model results indicate that females, higher educational attainment, higher household annual income, employed individuals, and more vehicles in the household are factors that contribute to higher probability of creating VMT, which is consistent with existing literature. Additionally, the model also suggests that people living in low-density areas are most likely to make zero VMT compared to others. The second part of the analysis uses quantile regression to explore the heterogeneity in the relationship of VMT with residential built environment characteristics. The comparison of the mean and median models suggests that several variables such as certain age groups, some types of household children structure, neighborhood types, population density, access to open spaces that have ocean views, and household income can only be significant factors in either the mean model or the median model. This implies that relying on the mean model to assess policies increasing land use density and diversity to inhibit VMT will not lead to the expected outcomes. Instead, analysis of significant factors influencing VMT for people with different levels of travel intensities and the mix in neighborhoods of people with different intensities will lead to a more precise impact assessment either for CO2 emission or CEQA.

The findings in this research of substantial heterogeneity with respect to the relationships between the type of VMT and the built environment characteristics have additional implications. Recall that heterogeneity is also present in VMT drive alone, VMT drive someone else, and VMT as passenger. Land use policies that inhibit VMT drive someone else will have a deleterious effect when people try to help vulnerable family, relatives, and neighbors. In this case the policy actions have unintended consequences that have not been studied sufficiently to adopt a blanket VMT target (and related CO2 emission per person target) at the level of an entire region as it is practiced today in California.

This research challenges the traditional approaches based on estimating the conditional mean which may exaggerate the conclusions of the impacts of land use on VMT. For example, we found that on average people having higher accessibility to open spaces with ocean views and higher accessibility to population density create less VMT by driving alone. However, at the median level of VMT drive alone, these two built environment variables are no longer significant. In addition, the sizes of the coefficients of these three built environment variables in the mean models are consistently larger than the ones in the median models. It means that the impacts of these built environment variables on people at the median level of travel demand are lower than the average. In other words, for those with higher travel demand, higher accessibility to ocean-view open spaces and population density can result in substantially lower VMT. This means our inference about the impact of accessibility on some types of average VMT will be misleading in regional plans, sustainable community strategies, and CEQA.

Our overall findings suggest that examining heterogeneity in the relationships among travel behavior determinants and VMT by different types is necessary. This is especially important for informing regional policies that require more targeted solutions for different segments of population. Our study also raises doubts about the use of mean regression when the underlying distributions of the dependent variable are skewed and when relationships among variables depend on the intensity of the dependent variable. This points out the need to examine again the wisdom of using VMT as the key policy variable and possibly replace VMT with easy to understand but somewhat more complex descriptors of daily patterns that account for travel services people provide to family, friends, and co-workers (Goulias et al., 2020). In addition, changes in work arrangements amplified by the impact of the most recent pandemic may motivate people to visit multiple work locations in addition to using a residence as a place to perform work for pay tasks. This may increase VMT instead of decreasing VMT because people visit multiple distinct locations in a day that may not be served by any other mode except private cars. This has major implications on VMT by different types (Su et al., 2021a) and should also be accounted for in land use policies when VMT is the key policy indicator. This is a complex issue that also requires differentiating between VMT accumulated using fossil fuel internal combustion engine cars (undesirable) versus VMT accumulated using zero or very low emission cars (more desirable). This is left as future task of our research.

This study, however, has a few limitations. The built environment variables used in this research are focused on the built environment characteristics surrounding people's home locations. Land use at the destinations is only partially accounted by the buffers around residences used here. However, many trips in the middle of a day originate from places other than home (Su et al., 2020). Future investigations should develop new approaches to consolidate built environment characteristics in a more holistic and comprehensive way. This has been done in a way that accounts for self-selection bias (Mokhtarian and Cao, 2008) using structural equations models (SEM) for one region in California by de Abreu e Silva et al.(2012) accounting for work and job locations and car ownership. Repeating this type of SEM analysis properly modified to include quantile regression methods could account for the complex nature of decisions of residential choice, work and school location choice, car ownership and type and ultimately VMT by type can provide important policy impact assessments and inference about policy actions. This is left as a future task because quantile structural equations are still developed.

4. Spatial Distribution of Schedule Fragmentation

4.1 Introduction

Automation and information and communication technology motivate a move away from traditional forms of employer-employee relationships with fundamental changes about when, where, and how people work and play but also how they move from one place to another (Alexander et al., 2010, 2011; Su et al., 2021a). Past research has clearly demonstrated the impact of technologies on activity and travel flexibility and fragmentation of time allocation (Couclelis, 2000, 2004; Hubers et al., 2008; Lenz & Nobis, 2007; Merz et al., 2009). All this implies that we need to describe daily travel behavior patterns with indicators that can capture time allocation to activities and travel in fine detail and explicitly include heterogeneity in time allocation to activities and travel. Analytical approaches that account for time fragmentation exist but correlations of this fragmentation to land use patterns is rarely found in the land use – transportation research. For example, we still lack the ability to understand how travel behavior choices are spatially correlated in spite of a few very focused examples such as mode choice of children (Sidharthan et al., 2011). This undermines our ability to conclusively assess land use policy action effectiveness and it is a core objective of this research.

In travel behavior there are many different methods to study time allocation to activity and travel patterns. A few representative examples of extracting activity patterns from diary data include a plethora of methods that recognize the relationship of activity and travel episodes in a day, across different days, and persons interactions within social groups (Auld et al., 2009; Bhat & Koppelman, 1999; Eluru et al., 2010; Ettema et al., 1995a; Garikapati et al., 2016; Goulias, 2002; Schlich & Axhausen, 2003). From this pantheon of methods, sequence analysis emerged as a robust holistic approach to represent time allocation in activities and travel and provides the basic information to develop behavioral taxonomies (Auld et al., 2009; C.-H. Joh et al., 2002; Liu et al., 2015; C. Wilson, 1998a; W. C. Wilson, 1998a; Xianyu et al., 2017; Zhang & Thill, 2017).

A simple sequence can be a daily time allocation pattern in which a person stays at home (H) until 8:00 am, then travels to work which is considered as a trip (T) and stays at a workplace (W) for 8 hours and then travels back home (T) and stays there until the next day (H). This sequence can be described with H-T-W-T-H. A more complex sequence is a pattern like: Leave home (H) in the morning to escort children to different schools (EC1, EC2, EC3), go to workplace (W), eat meal with customers and colleagues (M), return to work (W), leave work to pick up children from their schools (EC1, EC2, EC3), go shopping (S), return home (H), work from home (H) in preparation for the next day. This sequence can be described with H-T-EC1-T-EC2-T-EC3-T-W-T-M-T-W-T-EC3-T-EC2-T-EC1-T-S-T-H.

The letters in the above patterns are called the “states” in a sequence and we usually also count the amount of time in each state. Then, comparison (called sequence alignment) between patterns of sequences using a numerical assessment of differences is performed to distinguish between similar and different sequences. After the assignment of a numerical difference between two sequences is done, we classify sequences into types by examining the time-of-day profiles in time allocation to activities and travel.

Ideally we would like to have complex sequences in high density environments in which people enjoy activity participation and travel using sustainable and healthy modes such as walking and bicycling. Unfortunately, both simple and complex patterns are often characterized by trips (driving alone or using cars for hire such as taxis and ride hailing services) made in motor vehicles with fossil fuel internal combustion engines producing pollutants. This conflicts with policy initiatives combating climate change and leads to the most important research questions of this study, which are also the research gaps we found in the literature and they are:

(1) Do fragmented schedules cluster in space based on the most recent 2017 California component of the National Household Travel Survey?
(2) Are there places in California with residents who have complex activity sequences but with a desirable mix of public transportation and active transportation (walk and bike) and where are they?
(3) Are activity sequence patterns that contain many taxi and ridehailing trips in a day in specific geographical areas producing congestion and air pollution?
(4) Do simple activity sequence patterns that contain long commuting trips cluster in low density and diversity environments?

To answer the aforementioned questions, we developed a method that identifies first spatial and then temporal behavioral clusters using travel diary data in this chapter (and the paper that has been published from material in this report (Shi et al., 2022)). In comparison with previous methods, the method in this study combines spatial clustering with sequence analysis, enabling spatiotemporal analysis of people's daily schedules. Additionally, we conducted this study using driving distance, which is more accurate representation than Euclidean distance, to identify neighborhoods and compute access to opportunity indicators that are in turn correlated with spatiotemporal behavior.

4.2 Literature Review

Time fragmentation is important in travel behavior analysis not only because of its threat to congestion and pollution but also due to concerns about social exclusion. Lucas (2012) defines this as time-based exclusion: “…other demands on time, such as combined work, household and child-care duties, reduces the time available for travel (often referred to as time-poverty in the literature)”. In a similar line of thought Couclelis (2000, 2004) hypothesizes that economic, societal and political developments increase the individual’s flexibility in scheduling daily activities. However, this added flexibility may work in such a way that detriments some individuals by imposing constraints and releases constraints for others who in turn enjoy the advantages of free time. In this study *fragmentation* of activities and travel is defined as the sequencing of many short and long activities and trips that happen in a person’s daily schedule.

When a schedule is made of many activities and travel episodes in sequence we see multiple switching between different activities in a day. Patterns like this may lead to very short periods of personal time or no time at all, which is also called *time poverty* (Turner & Grieco, 2000). A similar idea is *time deficit*, which is computed by finding the time left in a day for household work after subtracting periods allocated to physiological activities (sleep, personal care) and work for pay. If the time is negative or very short, a household faces time-deficit. To compensate for this deficit, households purchase services and when their income is low they face a combination of income and time poverty (Vickery, 1977). Harvey and Mukhopadhyay (2007) show that single working parents suffer the most from a combined income and time poverty. Batur et al. (2019), using American Time Use Survey data also show that this type of time deficit and time poverty is associated with lower subjective well-being. Moreover, fragmentation combined with stress may lead to other health problems (Halpern, 2005). Fragmented time allocation patterns favor private cars as modes because many activities have variable and flexible start and end times that are decided on the spur of the moment and can be serviced only by on-demand services such as Transport Network Companies (TNCs, e.g., Didi, Uber, Lyft) or the cars and bicycles in the household fleet. The urban form and density of living environments are also associated with fragmented schedules, because higher density places allow people to engage in many activities within a small geographic area (Hough et al., 2008; Yuan & Raubal, 2012). Suburban residents were also found exhibiting high fragmentation even when one controls for the household structure (Novák & Sýkora, 2007).

One way to quantify fragmentation is sequence analysis in which activity and travel are defined as “states” between which people transition in a day. In travel behavior, Wilson (1998a; 1998b) used biology-inspired sequence alignment methods (i.e., comparison of strings of patterns of activities in a day) to study the sequences of activities, Joh et al. (2001) explored different techniques to introduce space in sequence analysis and Zhang and Thill (2017) developed a pattern recognition method that combines sequence alignment with network analysis to better represent relationships and hierarchy in activities within a day. Moreover, sequences of activities and the daily transitioning from one activity to another as well as the amount of time spent in each activity is important for activity-based travel demand analysis because it allows to build predictive models (Auld et al., 2011; Ettema et al., 1995b; Přibyl & Goulias, 2005).

The analysis in this chapter belongs to a systematic search for defining fragmentation in out of home activities and travel, extraction of a taxonomy of daily schedules based on sequences, study of the determinants of differences among behavioral patterns, and exploration of the correlation of fragmentation with land use. In the first exploration using sequence analysis to quantify fragmentation, McBride et al. (2019) used two measures (*Entropy* and *Turbulence*) depicting variety and frequent switching between activities and travel. In that analysis substantial fragmentation was found in activity participation among persons of age groups in the bracket 25–65 years old, amplified by the presence of children in the household, and reaching peak fragmentation values in age groups 25-34, 35-44, and 45-54 years old. As expected, these persons had lower fragmentation in the weekend days with Sundays and Saturdays being distinct and substantial differences among the other days of the week. Moreover, financial poverty was found to be an inhibitor to fragmentation. The place of residence was also found to influence fragmentation with suburban dwellers having the most fragmented schedules. Gender differences were found only for one indicator of fragmentation. Repeating this analysis using a statewide database (California Household Travel Survey, (NUSTATS, 2013)) using an indicator of fragmentation called *Complexity*, McBride et al. (2020) found nine distinct daily scheduling patterns with significant differences in their fragmentation, very different activity and travel behaviors within each pattern, and systematic differences in daily patterns. The Complexity indicator used in that analysis is used in this chapter and defined in detail later in the analysis section. The determinants of differences among patterns were found to be gender, employment, children in the household by different age groups, car ownership, residential classification (i.e., urban, suburban, exurban, rural) and density of customer service business establishments surrounding the survey respondent residence. A comparison of fragmentation between men and women in the same household showed that employed women with children have higher levels of fragmentation than men. This confirms other research on time-based exclusion mentioned earlier (Goulias et al., 2020).

Using the same sequence analysis techniques enriched with a network science pattern recognition technique and the same database we use in this research (i.e., the National Household Travel Survey California Component of 2017), Su et al. (2020; 2021a) found different destination choice patterns containing a few typical daily sequences (e.g., commuters to a single place of work and students to a single education location) but also substantial heterogeneity in fragmentation. In a study of seniors using the same database Su et al. (2021b) found strong positive correlation between complex destination choice patterns and urban residential status. These findings provide background support for the analysis in this study. Past analyses offer strong evidence that fragmented schedules are more likely to be in urban and suburban environments and less fragmented schedules are more likely to be in rural environments. However, we do not have strong evidence of spatial clustering. For example, are people living in urban environments with fragmented schedules surrounded by similar persons in terms of their activity patterns? The positive correlation of high density and diversity of land use with the complexity indicators is not enough to answer this question. In addition to studying the significance of regression coefficients of land use indicators we also need to explore if fragmentation exhibits spatial correlation and spatial clustering.

Spatial clustering is the idea of grouping observations based on spatial dependence and spatial homogeneity (Jacquez et al., 2008). It also includes identifying observations that although close in terms of distance or travel time are very different. In this research we have points of complexity values corresponding to the daily activity pattern of every person and the longitude and latitude of the residence of each of these persons. We need to find how many of these persons live in close proximity to each other and if their travel behavior is contradicting our planning objectives outlined in the introduction. This is in essence a problem of spatial point pattern analysis, and we want to group these points based on proximity, contiguity, interaction, and value of complexity/fragmentation. Grubesic et al. (2014) classify spatial clustering methods into four types that are the nonhierarchical (in essence more elaborate versions of k-means and k-medoid partitioning methods), hierarchical (a progressive inclusion of points into groups or exclusion from groups based on a measure of closeness among points), scan methods (geographic windows used to identify neighborhoods with high occurrence of points), and spatial autocorrelation based methods (assessment of similarity between neighboring locations globally or locally to detect high correlation locations). There are also latent variable model based spatial clustering methods that provide the opportunity to classify locations based on a wide array of point attributes such as the Latent Class Clustering Analysis (Ravulaparthy et al., 2012). In this chapter, we propose a disaggregate approach for improving classification, which combines a method for local spatial autocorrelation with a hierarchical method for sequence classification. In comparison to previous methodologies, our method takes geographical and temporal factors into account simultaneously, which contributes to a more complete solution to the research questions provided in the introduction.

4.3 Data Used

The data used in this study comes from the California component of the 2017 National Household Travel Survey (California-NHTS). NHTS contains a survey questionnaire that collects data on social and demographic information about each participating household, every individual in the household, and a single day travel diary of each person (the geocoded data are available through a special process of confidentiality agreement with CALTRANS). The one-day travel diary is on a pre-assigned day spread from April 24th, 2016 to April 24th, 2017 to cover an entire year. The diary day for each household can be any weekday, weekend day, or holiday. In total, we have 55,819 persons in the California-NHTS. Since human mobility patterns differ strikingly between weekdays and weekend days (Xiao et al., 2020) and we aim to probe the relationship between space and fragmentation of the most common daily schedule of people. For this reason, only persons from households that were assigned non-holiday weekdays for their diary have been selected. Moreover, respondents that do not start their day or end their day at home, respondents that did not leave their home during the interview day, and samples with no neighbors within a four-kilometer driving distance based on Open Street Map data have also been excluded. This leaves us with 26,078 daily records (see Figure 4.1). For this analysis, we use person and household characteristics reported in Table 4.1. From the diary data we use the timing of trips, modes used, and activity types at origins and destinations. Trips (the one-way movement from an origin to a destination) are reported between 4:00 AM on the survey day until 3:59 AM on the following day.

To incorporate potential systematic trends between fragmentation spatial clusters and the different built environment settings, we use a built environment dataset containing both community design and accessibility measures of 23,190 U.S. census block groups in California (CalTrans, 2020). Based on these characteristics the residence of each participant was classified in urban vs rural environment as reported in Table 4.1. Apart from the variables given in Table 4.1, this study provides twelve dummy variables for counties where observations reside. The analysis also includes four synoptic variables to enable better understanding of people's travel behavior: (1) travel time ratio, which is defined as the total travel time (spending on changing travel modes, dropping off or picking up others, and trips) in a day divided by the total time outside the home (Dijst and Vidakovic, 2000), which is a gauge for people's trade-offs between travel and activity time. A high TTR can indicate that the travel costs associated with the activities are high; (2) Gini index, which quantifies the daily variation of mode choices (such as walking, biking, driving alone, driving others, ect.) (Xiao et al., 2020). A Gini index value of zero means only one mode was used and a higher Gini suggests that more modes are used. This captures the propensity of travel mode switching frequency. High switching with high fragmentation is preferable to no switching and use of car as a driver alone; (3) stay-at-home ratio, which is equal to the total amount of time spent at home divided by 1440 minutes (24 hours) to capture inactive persons who may also have low fragmentation; and (4) the driving distance between household residence and workplace (if a respondent does not go to work, this variable is assigned to zero). This variable captures commuting costs and also captures jobs-housing lack of balance due to high housing costs in attractive places of employment (Islam & Saphores, 2022; Mitra & Saphores, 2019). We divided the actual driving distance in kilometers by 1,000,000 to achieve a visible coefficient in the multinomial logit (MNL) regression model.

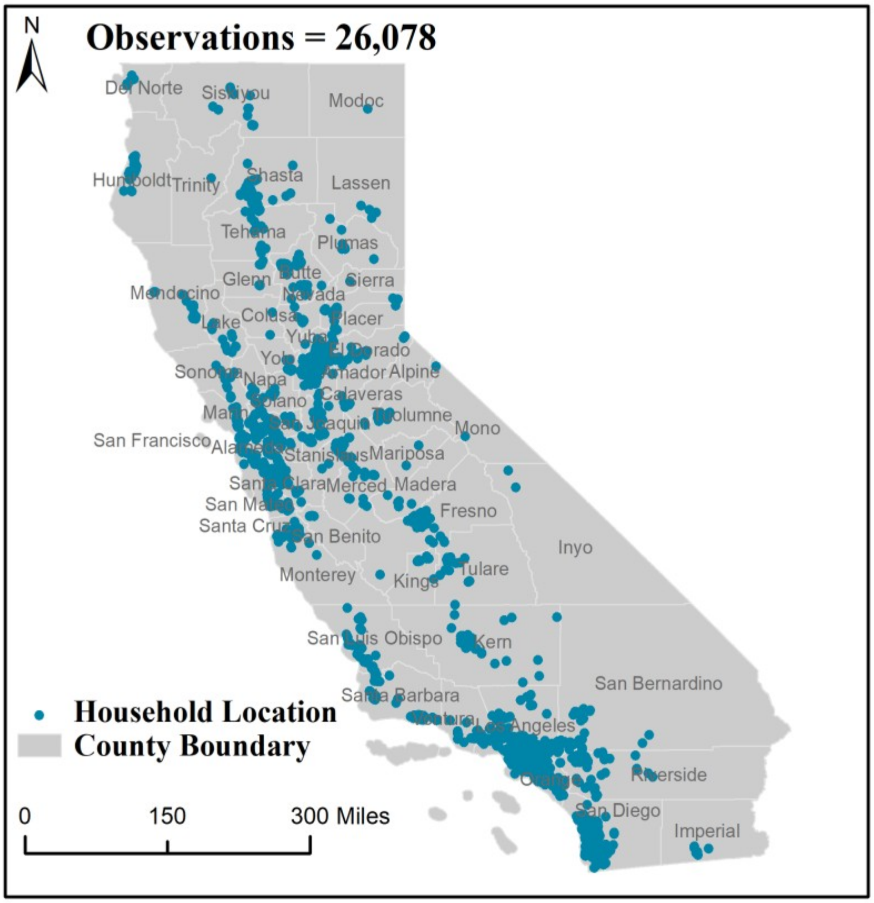

Figure 4.1. Spatial distribution of respondents' household locations.

Table 4.1. Selected variables to explain the fragmentation (n=26,078).

Variable	Description	Descriptive statistics
Age	The age of the respondent in years	Under 18: 13.72% 18 to 25: 5.76% 26 to 45: 23.21%
Sex	Respondent's binary gender	Women: 52.37%
Full-Time	If the respondent is a full-time worker	Full-time worker: 37.08%
Driver	Driver status	Yes: 81.75%
Walk	Count of walk trips	Min: 0.00 Mean: 5.49 Median: 4.00 Max: 99.00
Bike	Count of bike trips	Min: 0.00 Mean: 0.50 Median: 0.00 Max: 75.00
Public Transit	Count of public transit usage	Min: 0.00 Mean: 1.16 Median: 0.00 Max: 150.00
Ridesharing	Count of rideshare app usage	Min: 0.00 Mean: 0.4549 Median: 0.00 Max: 99.00
Travel day	Travel day - day of week	Monday: 17.13% Tuesday: 21.26% Wednesday: 20.78%
Household income	Respondent's household income	$50,000 or more: 68.12%
Household structure	Respondent's household structure	one adult, no children: 8.54% 2+ adults, no children: 22.30% one adult, youngest child 0-5: 0.54% one adult, youngest child 6-15: 2.53% one adult, youngest child 16-21: 0.97% 2+ adults, youngest child 0-5: 11.57% 2+ adults, youngest child 6-15: 16.29% 2+ adults, youngest child 16-21: 5.94% one adult, retired, no children: 7.05%
Residential location type	The built environment of household location[1]	Urban Core: 0.77% Urban District: 1.61% Urban Neighborhood: 5.30%

[1]Urban core is the area with a dense population and is accessible to employment by a traditional street network with strong local and regional multimodal connectivity. Urban district and urban neighborhood are placed lower on the location-efficiency spectrum.

4.4 Methodology

This chapter uses sequence analysis, spatial clustering, and hierarchical clustering that are described in this section. The flow chart representing the sequential order of our analyses is shown in Figure 4.2. Following the computation of *Complexity*, LISA analysis was performed to find groups of observations with similar high or low fragmentation in space, and within each LISA cluster, hierarchical clustering was conducted to detect distinct patterns of time allocation. Simultaneously, we estimated a MNL model in order to deduce the significant determinants of LISA spatial clusters.

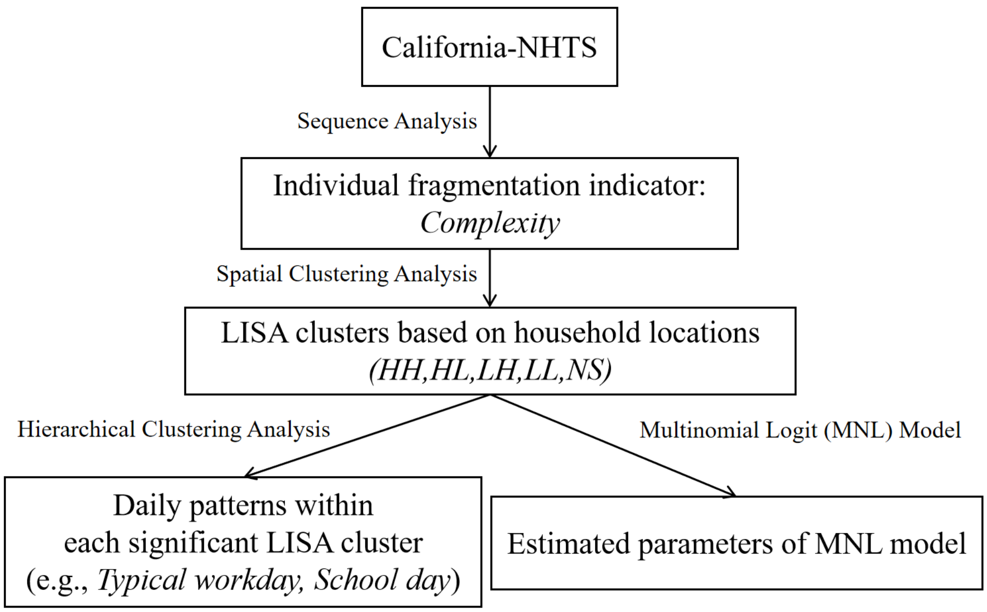

Figure 4.2. Flow chart of spatial-temporal analysis of people’s time fragmentation.

4.4.1 Fragmentation indicator

A sequence is a series of discrete time points and each time point is assigned to a “state.” A “state" can be Home, Work, School, Trip, or any other classification needed. A subject can move from one discrete “state” to another. The states in this study are defined as follows: Home with unspecified activity (*Home*); Work from home as in telecommuting and home stay combined (*Home&Work*); Work at a workplace or at other places (*Work*); Work-related meetings or trips (*WorkRelated*); Education at the school location (*School*); Attend child care (*ChildCare*); Attend adult care (*AdultCare*); Visit health care centers (*HealthCare*); Drop off or Pick up someone (*DropPickup*); Change type of transportation (*ChangeTrans*); Purchase goods such as groceries, clothes, appliances, gas (*BuyGoods*); Purchase services such as dry cleaners, banking, service a car, pet care (*BuyService*); Go out for a meal, snack, carry-out (*BuyMeal*); Run other general errands such as post office (*ShopServ*); Do volunteer activities without payment (*Volunteer*); Participate in religious or other community activities (*Community*); Engage in recreational activities such as visit parks, movies, bars (*Recreation*); Exercise (*Exercise*); Visit friends and/or relatives (*VisitFrsRls*); and all Other (*Other*). Travel between these places is also considered a "state" noted as *Trip* (and distinguishable from *DropPickup* and *ChangeTrans*). The daily sequence has a length of 1440 which is the total minutes in an assigned survey day. Every minute of the day contains one of the 21 distinct "states" for each person.

Fragmentation in this research is the value of an indicator called *Complexity* and it combines *Entropy* in Eq. 4.1 and transitions between distinct states within a sequence (*x*) defined in Eq. 4.2. The explanation here follows Gabadinho et al. (2011) and McBride et al. (2019) closely.

$$h(x) = h(\pi_1, \pi_2, ..., \pi_3) = -\sum_{i=1}^{s} \pi_i \log(\pi_i)$$

(Equation 4.1)

$$C(x) = \sqrt{\frac{(l_d(x)-1)}{(l(x)-1)} \frac{h(x)}{h_{\max}}}$$

(Equation 4.2)

where $h(x)$ shows the entropy function of x, which represents the sequence, s is the number of distinct states, and π_i is the proportion of occurrences of i th state in the considered sequence. To include the number of transitions between different states, *Complexity* ($C(x)$) combines entropy with the number of transitions $l_d(x)-1$ in a sequence x, normalized by the maximum theoretical entropy $h_{\max}$ (entropy is at its highest when people spend equal amounts of time on each state) and the length of the sequence $l(x)$ which equals 1440. In summary, $C(x)$ accounts for the diversity of visited states in a day expressed as the geometric mean between the proportion of transitions between states and a normalized version of the within-sequence entropy $h(x)$. In this way, a daily pattern state diversity is measured by entropy and adjusted for the proportion of transitions, which also takes into account the number of occurrences of each state. This is why Person 5 has a higher *Entropy* value than Person 6 but a lower *Complexity* value in Table 4.2. *Complexity* takes a value between 0 and 1, with zero corresponding to entropy zero and no transitions (e.g., staying at a single place for the entire day of observation). High complexity values correspond to higher fragmentation and lower complexity values correspond to low fragmentation as Table 4.2 indicates.

Table 4.2. Examples of Sequences and Fragmentation

Person	(State, Duration in minutes)	Entropy h(x)	Complexity C(x)
Person 1	(Home,150)-(Trip,40)-(Home,1250)	0.127	0.008
Person 2	(Home,630)-(Trip,33)-(Work,247)-(Trip,35)-(Home,495)	0.639	0.024
Person 3	(Home,420)-(Trip,98)-(Recreation,198)-(Trip,62)-(Home,662)	0.732	0.026
Person 4	(Home,250)-(Trip,30)-(Work,170)-(Trip,5)-(BuyMeal,40)-(Trip,10)-(Work,335)-(Trip,20)-(Home,580)	0.924	0.041
Person 5	(Home,220)-(Trip,5)-(Work,225)-(Trip,10)-(BuyMeal,40)-(Trip,10)-(Work,270)-(Trip,5)-(Home,115)-(Trip,15)-(Exercise,45)-(Trip,10)-(Home,470)	1.025	0.053
Person 6	(Home,165)-(Trip,5)-(DropPickup,1)-(Trip,4)-(Home,120)-(Trip,10)-(BuyGoods,10)-(Trip,10)-(BuyGoods,15)-(Trip,5)-(Home,40)-(Trip,10)-(Community,105)-(Trip,10)-(BuyMeal,30)-(Trip,20)-(Recreation,40)-(Trip,30)-(Home,65)-(Trip,5)-(DropPickup,5)-(Trip,5)-(Home,730)	0.861	0.065

4.4.2 Spatial clustering of fragmentation indicator

The spatial clustering analysis in this study is based on identifying spatial correlation among fragmentation indicators and discovering if there is a systematic spatial distribution of these indicators that is significantly different from a completely random spatial distribution, and then performing hot spot analysis of the spatial associations found in the data. Paez and Scott (2004) define succinctly this as “the tendency of variables to display some degree of systematic spatial variation. …(this) means that high variable values are found near other high values and low values appear in geographical proximity”. In this study a local version of Moran's I called Local Indicators of Spatial Association (LISA) (Anselin, 1995) is conducted to derive clusters with “neighborhood” defined by a driving distance threshold of four kilometers (any neighbors that share the same household locations with the targeted sample will be given a distance of ten meters to avoid zero division in the spatial weights matrix). This threshold is chosen based on a sensitivity analysis (see Appendix B). The LISA is below.

$$LISA_i = \frac{x_i - \bar{x}}{\sigma^2} \sum_{j=1}^{n} w_{ij} (x_j - \bar{x})(i \neq j)$$

(Equation 4.3)

where $LISA_i$ denotes the local Moran’s I for sample i, x_i and x_j are the values of x (complexity) in the corresponding spatial location i and j, respectively; $\bar{x}$ represents the average of the x value; σ^2 is the overall variance of all samples; n is the total number of spatial units; and i is the i[th] unit; w_{ij} denotes the row standardized weight of the neighbors. The function used is 1 over the driving distance between the respondent's household and the neighbors' households within a four-kilometer buffer. For each sample, the normalized weights of all neighbors add up to one. This indicates that each sample may have a unique set of neighbors. In this study, the average number of neighbors is 45, with a minimum of 1 and a maximum of 304. In comparison to the conventional methods of creating the spatial matrix (such as rook contiguity, queen contiguity, and k-nearest), calculating distance based on road networks is more reliable and reasonable, as the Euclidean distance does not take land use into account and thus cannot accurately approximate the real distance in certain circumstances (Shahabi et al., 2003).

LISA values range from −1 to 1. A value close to 1 suggests a positive autocorrelation, while a value close to −1 refers to a negative autocorrelation. If its value is close to 0, it indicates a random spatial distribution. There are four significant spatial association groups that one can derive in this way, they are a cluster of high values surrounded by high values (HH), a cluster of low values surrounded by low values (LL), a cluster of high values surrounded by low values (HL), and a cluster of low values surrounded by high values (LH)) plus a group of observations that are not significantly different than average fragmentation called the nonsignificant distribution (NS). In this analysis, comparison random scatter is created using 999 permutations and the results reported use the 90% confidence level.

4.4.3 Hierarchical clustering of sequences

The sequence analysis above produces a very large number of sequences and many of them have similarities in the number of activities, trips, and timing of both. To understand the types of sequences we need to create groups of similar patterns. Using this type of data, developing groups of daily patterns is done efficiently using hierarchical clustering (McBride et al., 2019; Su et al., 2020; Su et al., 2021a). In the pattern analysis literature (Kaufman & Rousseeuw, 2009), dissimilarity between two sequences is measured by the number of operations needed to make two sequences exactly the same and this is called a “distance”. Typically, the distance between two sequences is the minimum combination of substitution and indel (i.e., insertion and deletion). In addition to these two possible operations, there is one termed inversion. However, both indel and inversion alter the order in which activities occur. As Gabadinho et al. (2011) noted, using indel decreases the significance of time shifts in comparisons, whereas using substitutions emphasizes positional similarity. In our scenario, the sequential order of daily activities and time allocation is critical for pattern detection, as successive activities are likely to have an effect on one another (Joh et al., 2001). Additionally, indel and inversion have the drawback that their costs are usually determined by a manually chosen constant that lacks credible support, while substitution costs are determined by the transition rates between each pair of distinct states in all sequences. Thus, while measuring the dissimilarity of two sequences, we consider only substitutions, and the Hamming distance is used since it quantifies the least number of substitutions required to convert one sequence to another. Explicitly, here we use the same technique as in Su et al. (2020) and follow Gabadinho et al. (2011) and the method called agglomerative nesting clustering method (AGNES). In essence we have sequences of the same length (1440 minutes in a day) and each minute is assigned to one of the 21 states mentioned above. Then, we use an algorithm that assigns each sequence to its own group and adds to each group another sequence computing the differences among groups of sequences and how well each step performs in terms of distinguishing different sequences. This is done computing within-cluster sum of squares (WSS) and average silhouette coefficient (Silhouette) and these are used widely for many different types of clustering algorithms (Kaufman & Rousseeuw, 2009). The outcome of this step groups sequences that belong to distinct types of time allocation patterns in a day. To visualize the overall patterns of time allocation in each group we can use a stacked bar chart as illustrated below.

Figure 4.3 shows a stacked bar chart depicting the daily time allocation pattern of a group of person-days. In this study, there are 21 states of activity (see graph legend). The x axis displays the time of day in minutes, starting at 4:00 a.m. and ending at 3:59 a.m. in the following day. The y axis indicates the relative frequency of persons undertaking one of the 21 kinds of activities. The example demonstrates a typical commuting mobility pattern that the majority of individuals in this group begin their day at home, leave home for work early in the morning, stay at workplaces during the day, gradually return home after 4 p.m., and spend the night at home. Furthermore, people in this cluster spend a modest amount of time on trips, buy goods or meals, exercise, drop off and pick up other people, and visit friends or relatives.

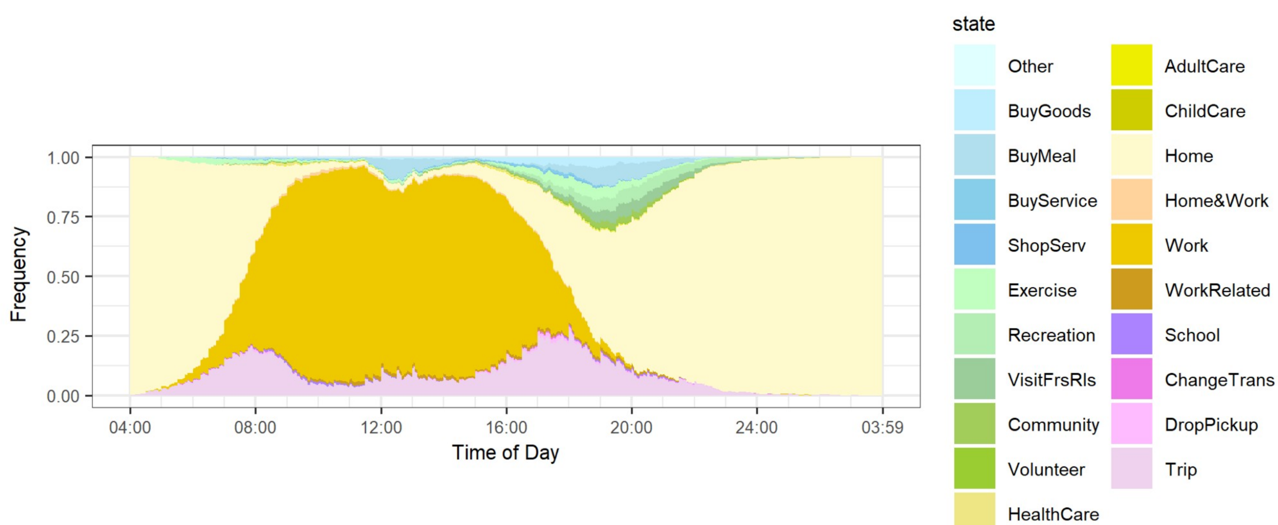

Figure 4.3. An example of daily pattern of min-by-min activity sequences

In summary and based on Figure 4.2, the data analysis steps are as follows. First, we create for each person in the sample sequences of 1440 letters representing activities or travel. Then, we compute the complexity indicators for each person (in essence a person-day) and then run a LISA analysis using the longitude and latitude values of the residence of each person and the values of the complexity. Each person is then classified in one of the 5 LISA clusters (spatial clustering categories that are HH, HL, LH, LL, and NS). After this, we find similar sequences within each LISA cluster applying hierarchical clustering within the LISA cluster to find distinct sequences like the Figure 4.3. In parallel, we estimate a multinomial Logit model to identify significant determinants of the LISA spatial clustering.

4.5 Results

4.5.1 Comparison of fragmentation of LISA generated clusters

Five LISA clusters categories have been generated from the travel diaries of the 26,078 respondents. The number of samples and descriptive statistics of the complexity values for each category are shown in Table 4.3 and the corresponding box plot is presented in Figure. 4.4

On the whole, the mean complexity of Californians is 0.037 and the standard deviation is 0.016, with a minimum of 0.002 and a maximum of 0.145. This indicates substantial heterogeneity of people's daily activity and travel fragmentation for many persons with patterns mostly spanning from sequences of Persons 3 to sequences of Person 4 of Table 4.2. From past analyses we know this is attributed to different ages, household income and location, occupation, and within household task allocation (McBride et al., 2020; Su et al., 2020; Su et al., 2021a). Regarding the descriptive statistics of the complexity values within each LISA cluster (Table 4.3), the highest mean complexity (0.055) is observed in the HH group due to the diversity in activities throughout the day, followed by the HL cluster. On the contrary, people in LL and LH clusters have a low average complexity as expected because of just a few activities in a day, with the mean complexity of 0.024 and 0.026, respectively. More than 70% of samples belong to the non-significant (NS) type, suggesting that the spatial distribution of most people's daily fragmentation is not significantly different than spatial randomness.

Table 4.3. Descriptive statistics of the complexity values for each LISA cluster.

Clusters	Observations (Percent)	Complexity				
		Min	Mean	Std.	Median	Max
HH	2,363 (9.06%)	0.037	0.055	0.013	0.053	0.124
HL	1,048 (4.02%)	0.037	0.050	0.011	0.047	0.116
NS	19,080 (73.17%)	0.002	0.037	0.015	0.035	0.145
LH	1,269 (4.87%)	0.003	0.026	0.007	0.027	0.037
LL	2,318 (8.89%)	0.003	0.024	0.008	0.025	0.037
Total	26,078 (100%)	0.002	0.037	0.016	0.035	0.145

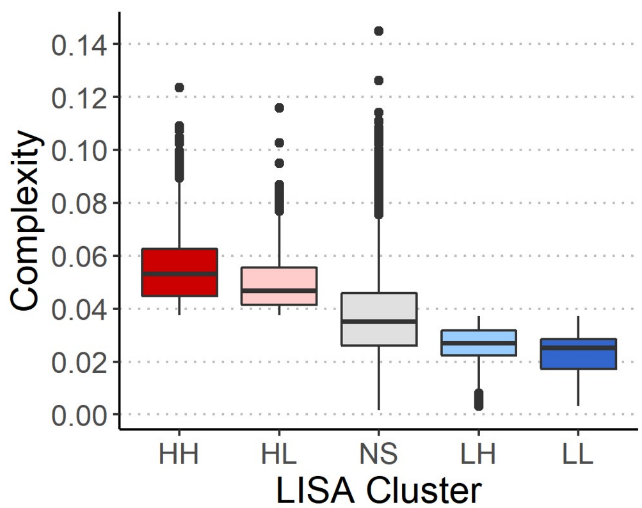

Fig. 4.4. Box plot for complexity indicators among different clusters generated by LISA.

4.5.2 Spatial distribution of different LISA clusters

To visualize the clusters and outliers, we mapped the percent of four significant clusters for each county in each county of California in Fig. 4.5. HH cluster (2,363 persons and 9.06% of sample): Respondents in this cluster have high complexity with spatial homogeneity, which indicates that people have similar complexity with their neighbors. Most of them are distributed in coastal or metropolitan areas, such as San Francisco Bay Area and Ventura but also Yuba and Yolo that are adjacent to Sacramento. These are higher density areas with dense infrastructure and recreation facilities. Observations in the HH clusters are also found in less developed regions such as Madera and Trinity. This is partly due to the fact that Yosemite National Park is located in Madera, which motivates residents to travel out for nature viewing. While Trinity is the fourth least populous county in California, the HH proportion is statistically insignificant due to the small sample size (29 persons in total).

HL cluster (1,048 persons and 4.02% of sample): respondents in this cluster have high complexity but are surrounded by samples with low complexity, with a strong spatial heterogeneity. The percentage of participants in this cluster of each county is generally lower (ranging from 1% to 5%) compared to that of HH type except Sierra, Tehama, Inyo, Merced, and San Benito. The reason why Sierra has a relatively high percentage is because of its small samples, with only 7 respondents. Additionally, Merced is home to some big corporations, and with relatively developed highways and public transportation infrastructures. These variables contribute to the diversity of local people's travel behaviors. As for the other counties listed above, even though they are not metropolitan areas, they have recreational areas, offering hiking, camping, boating, fishing, and so on.

LH cluster (1,269 persons and 4.87% in sample): samples in this category have low complexity but surrounded by high complexity. The spatial distribution of counties with large LH proportions is partly similar to that of those with high HH percentages. This is because these two clusters are near each other. Despite the fact that 50% of samples in Alpine county belong to the LH cluster, only two persons completed the survey on workdays. Thus, this proportion makes no sense. It is also worth noting that more than 15% of respondents in San Francisco county experience low fragmentation. This also happens in other counties of the San Francisco Bay Area.

LL cluster (2,318 persons and 8.89% in sample): people in this group have low complexity and their neighbors too. Counties with relatively large proportions of people in LL cluster mainly concentrate in rural areas or natural reserves, like Calaveras (31.86%), Siskiyou (18.48%), Tuolumne (17.90%), and Tehama (17.28%). Those areas are more likely to be located in the eastern and northern parts of California and are characterized by mountainous terrain or forests.

NS cluster (19,080 persons and 73.17% in sample): since the majority of people belong to this cluster, many counties have a high proportion of NS daily patterns, particularly Colusa, Lassen, Inyo, and Napa counties, where the percentages exceed 80%. However, their sample sizes are all less than 100.

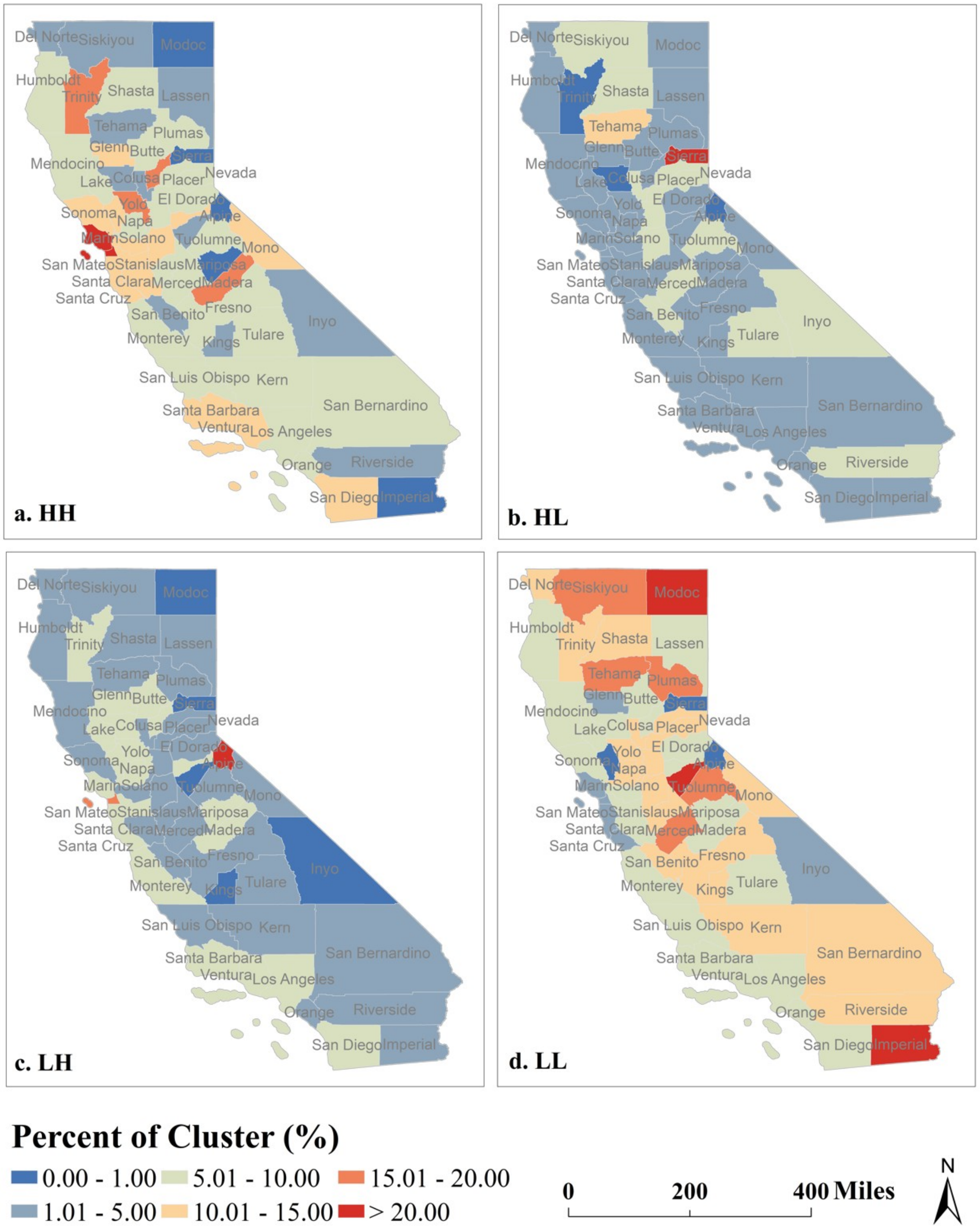

Figure 4.5. Percent of corresponding respondents for each of the LISA clusters.

Pacific Southwest Region UTC
University Transportation Center

4.5.3 Time allocation patterns within each LISA cluster

In this section we explore the within spatial cluster time allocation pattern diversity using hierarchical clustering of sequences to find distinct daily activity and travel patterns and classify each observation in one of these patterns. Figures 4.6 to 4.9 show people's daily patterns for the HH, HL, LH, and LL LISA clusters. To be specific, the LH cluster exhibits four distinct daily patterns, the HH and LL clusters show three distinct daily patterns, and the HL cluster displays only two distinct daily patterns. This shows considerable time allocation heterogeneity within each LISA cluster.

Although different LISA clusters may have dissimilar daily travel patterns, the typical workday appears in all LISA clusters. It is the typical commuting pattern in which people travel in the early morning to work and go back home in the evening. People belonging to high complexity clusters tend to take a lunch break and usually visit some other places (such as supermarkets, grocery stores, gyms) before getting home (Su et al., 2020; Xiao et al., 2020), especially for the samples falling into the HH cluster who are also parents or caregivers for the elderly (Su et al., 2020). However, people in LH cluster or LL cluster have relatively simple travel patterns, normally working until they are off duty and then return home directly.

School day is another typical pattern that is found in all significant LISA clusters except the HL cluster. This is mainly because children cannot have high complexity sequences when their closest neighbors (their parents) have low complexity due to escorting children to school practices. It is also worth noting that school going persons in the HH cluster allocate more time to other activities after school including but not being confined to going out to buy goods and meals or for recreation, doing exercise, and visiting friends.

Discretionary day cluster includes respondents spending a substantial proportion of time traveling and participating in leisure activities. They belong in the two high complexity clusters (of the 1,579 persons with a discretionary day pattern, 999 are in the HH cluster and 580 are in the HL cluster). This pattern is also associated with long-distance travel and is likely overlooked in travel demand forecasting (Su et al., 2020). Meanwhile, it is noticeable that a few people in this group also go to work for a short time during the day of the NHTS interview.

In addition to the above three pattern types, late work day pattern is also found in the LH cluster in 48 person schedules. Compared to people with a typical workday pattern, those respondents start working later and finish later as well, however, they participate much less in other activities. Home day is another typical pattern that has 2,216 (32.21% of the sample in significant LISA clusters) people who spend most of their time at home with a few trips to run errands and just going out for a short time (e.g., buying goods). This pattern is only observed in low complexity clusters.

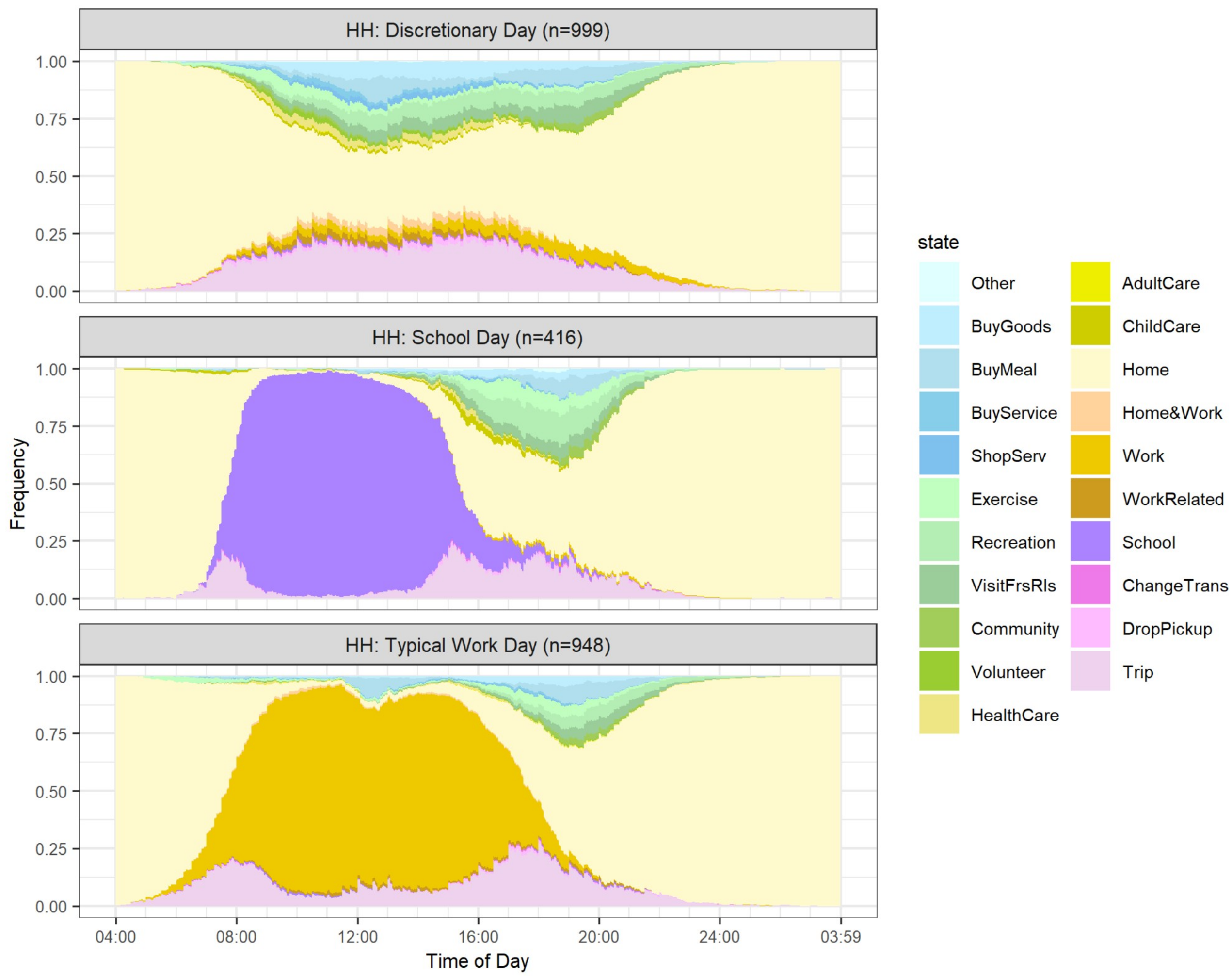

Figure 4.6. Daily time of day patterns of activity sequences for people in HH cluster.

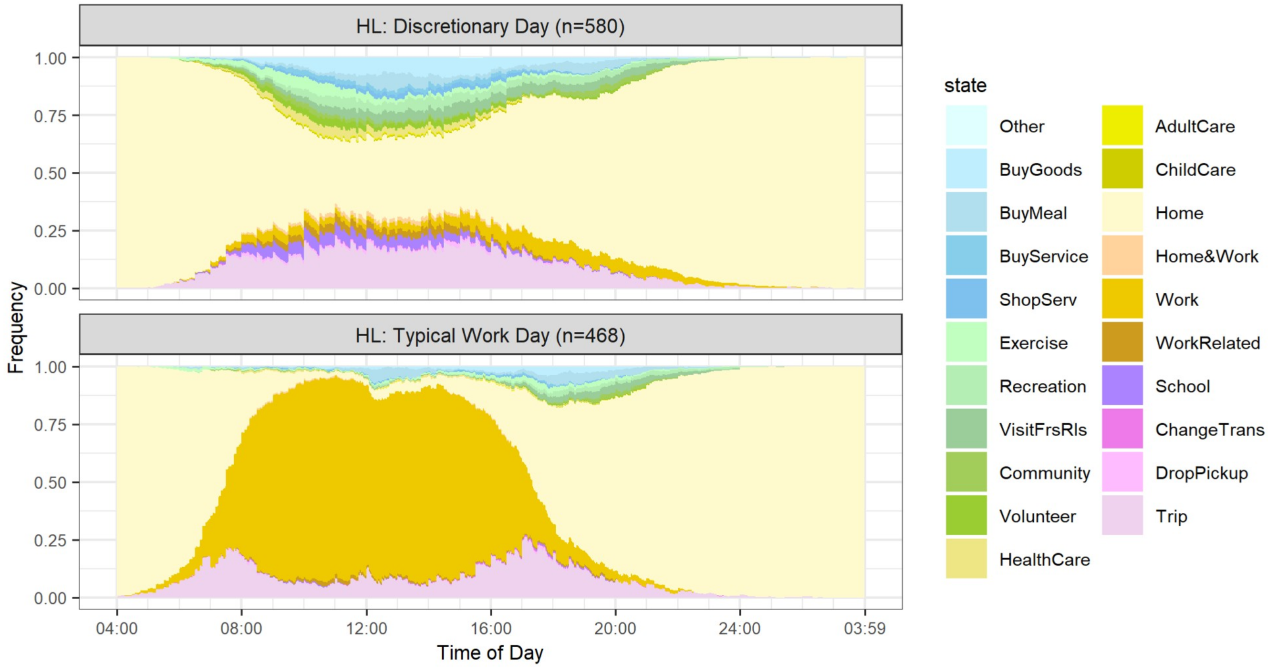

Figure 4.7. Daily time of day patterns of activity sequences for people in HL cluster.

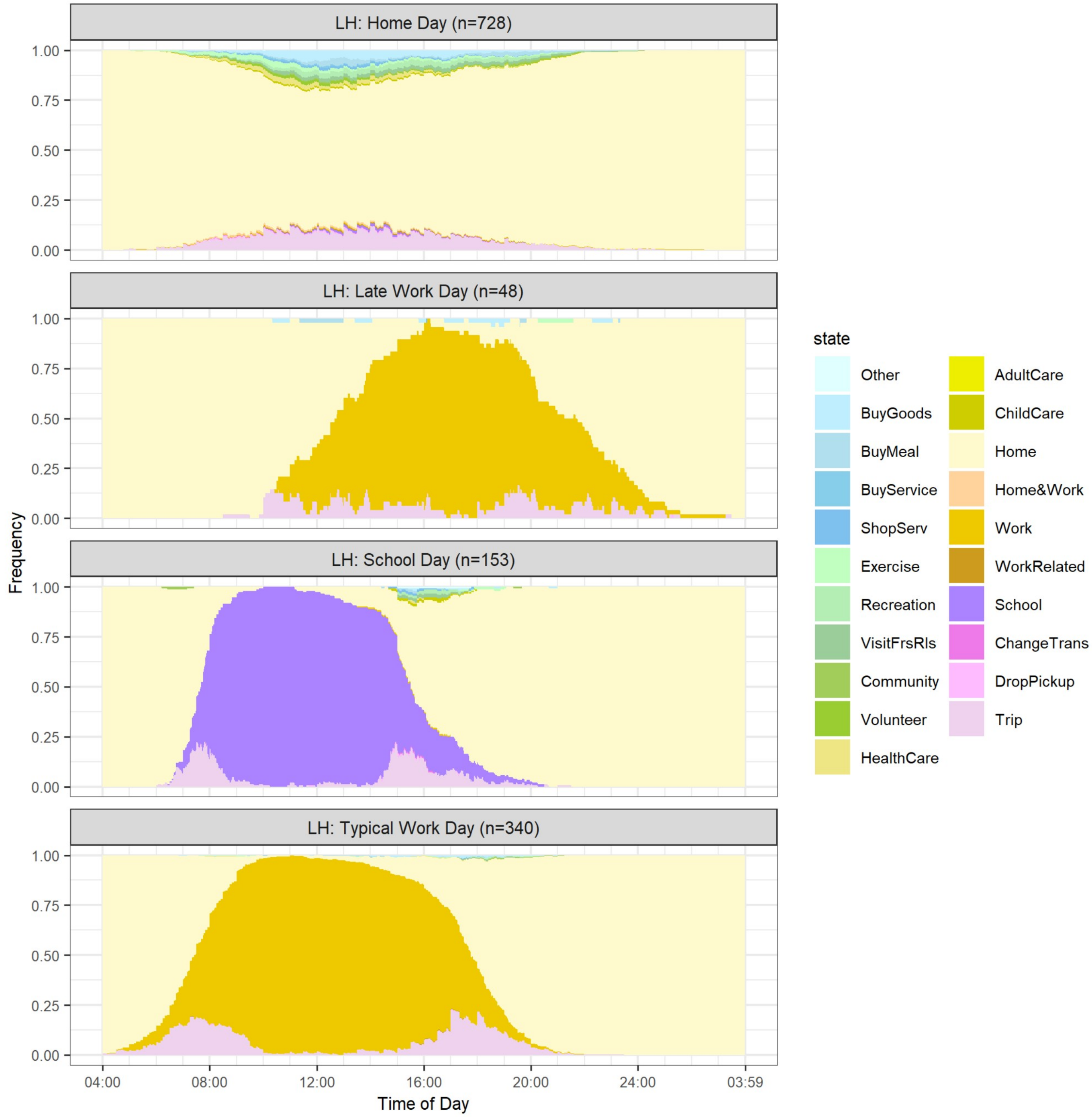

Figure 4.8. Daily time of day patterns of activity sequences for people in LH cluster.

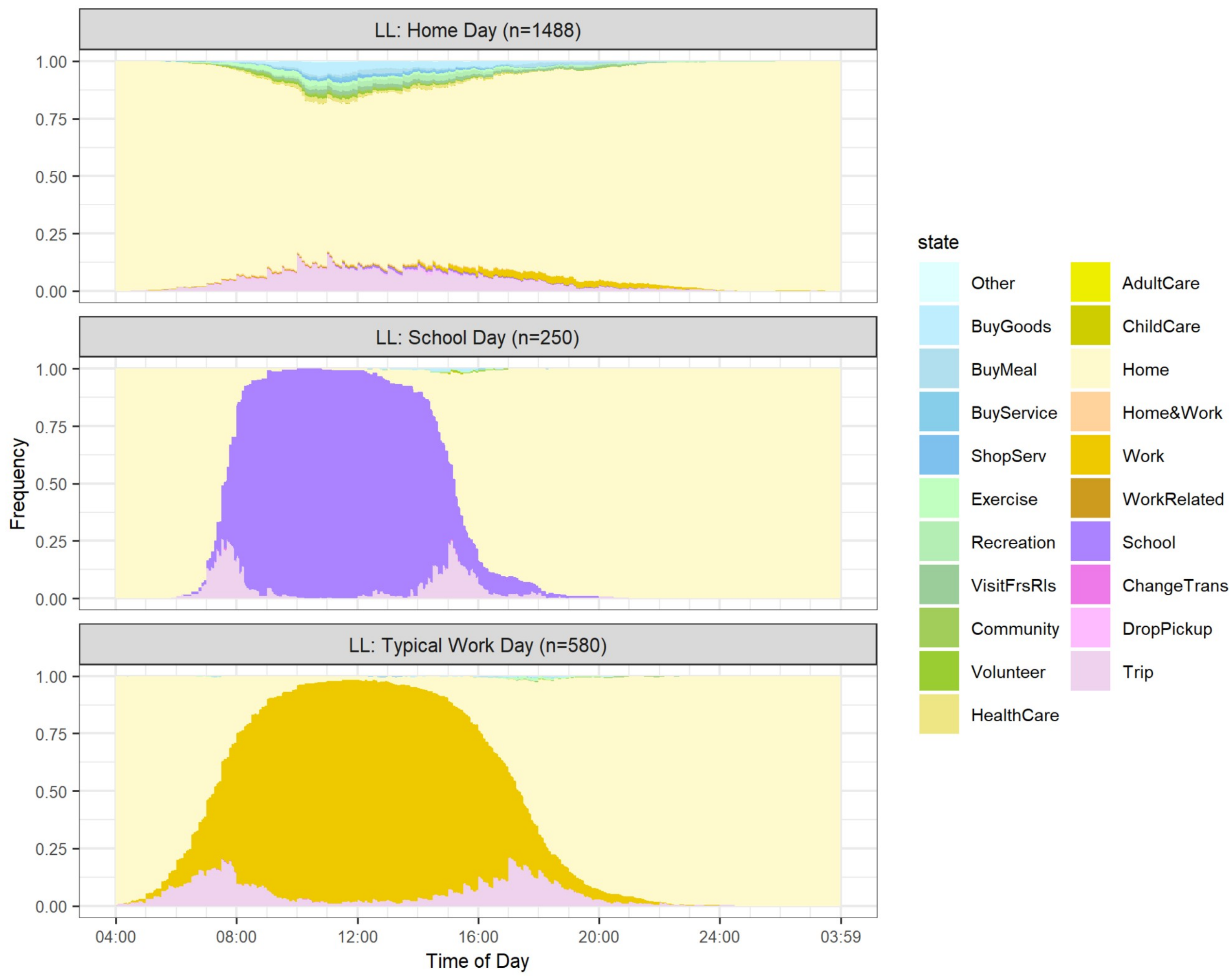

Figure 4.9. Daily time of day patterns of activity sequences for people in LL cluster.

To investigate the differences in time allocation across different patterns in further detail, we calculated the average duration of states for each pattern and present them in Table 4.4. Before that, we classified *BuyGoods*, *BuyMeal*, *BuyService*, and *ShopServ* under *ForErrands*, *Exercise*, *Recreation*, and *VisitFrsRls* under *ForFun*, *Community* and *Volunteer* under *ForService*, *HealthCare*, *AdultCare*, and *ChildCare* under *ForCare*, and *Home&Work*, *Work*, and *WorkRelated* under *ForWork*. Then states with fewer than 30 participants in each pattern were excluded to ensure the statistical significance of those mean values.

It is worth noting that people spend the majority of their time at home in all patterns, owing to the fact that certain activities, including sleeping, selfcare, and household work, are mandatory. Additionally, trips are recorded in all daily patterns, but individuals in high complexity clusters (HH and HL) spend a greater amount of time on the road. Individuals who have a discretionary day are more likely to go out for errands and recreation, whereas those who go to work late during the day or who attend school in low complexity groups are less likely to engage in shopping activities. Community or volunteer activities are more likely to be in clusters with high complexity, but also in home days with the low complexity. This is also true for caring related activities. Given that discretionary days and home days also include working and schooling, the average duration is significantly less than that of work and school days. Furthermore, changing travel modes and driving to drop/pick up persons are more frequent in high complexity clusters, particularly changing travel modes, which is exclusively found in HH clusters. This is mainly because the HH population has greater access to a higher variety of infrastructure facilities in developed places such as San Francisco.

Table 4.4 The average duration of states (in minutes).

Patterns	Home	ForErrands	ForFun	ForService	ForCare	ForWork	School	ChangeTrans	DropPickup
HH: Discretionary Day	955.24	57.39	137.26	135.00	116.02	269.03	153.07	47.80	39.51
HH: School Day	800.26	41.58	108.35	112.74	96.10		415.94	17.97	13.50
HH: Typical Work Day	741.11	39.58	91.56	111.35	83.45	418.10		18.01	18.82
HL: Discretionary Day	1005.62	52.94	138.12	147.23	103.85	255.05	318.82		30.96
HL: Typical Work Day	790.13	34.93	87.33			424.49			12.09
LH: Home Day	1260.94	43.66	122.97	163.73	115.89	195.17			37.87
LH: Late Work Day	903.31					463.75			
LH: School Day	937.25						439.39		
LH: Typical Work Day	820.20	19.85				536.69			
LL: Home Day	1281.88	42.44	117.34	144.76	72.48	319.61			35.09
LL: School Day	975.70						421.82		
LL: Typical Work Day	836.95	19.05				529.35			

Note: The background color in a gradient is according to the data in this table.

4.5.4 The relationships between LISA clusters and people's characteristics

To offer further insights on the five different LISA spatial clusters, a MNL regression model is used to associate LISA clusters with respondents' characteristics and their household attributes. NS type is used as the reference category. The model fits the data well demonstrated by a chi-square likelihood ratio of 5,194.9 (with P-value less than 0.01). Table 4.5 includes only the coefficients that are significantly different from zero at the 0.1 level.

In terms of individual variables, children are most likely to be in the HH and LL clusters but less likely in the HL cluster, which suggests that children fragmentation tends to be comparable to that of their neighbors who often happen to be their parents (Chandrasekharan & Goulias, 1999, Bhat et al., 2013). This is primarily because children are usually accompanied by their adult parents to school but also other places for exercise and leisure. People between the ages of 18 and 45 are less likely to have high complexity but more active neighbors. This is partially because many persons in this group spend a lot of time commuting and working and are thus facing less available time for other activities in their schedules than other family members. Additionally, the gender result indicates that women are less likely to exhibit low complexity and they are surrounded by people with relatively high complexity. It is in line with previous studies that women are generally the ones taking more household responsibilities even when they also need to work (Jiao et al., 2020; Su et al., 2020; Turner & Grieco, 2000). Full-time workers are more likely to be in the LL cluster and this is in combination with the impact of the positive and significant coefficient of the home-work distance show the significant inhibition of for higher fragmentation. When it comes to the frequency of using different travel modes, we find that: (1) the HH and LH people prefer to walk for travel; (2) people in the HL and LL clusters are less likely to ride bicycles; (3) public transportation can contribute to people's low complexity by requiring them to spend additional time on the road; and (4) ride sharing is favored more by people with HH type as its convenience and high comfort level somewhat encourage people to have complex mobility patterns. This last finding points out the potential risk associated with the availability of TNC services in contributing to congestion and air pollution.

The Gini index coefficient (significant, positive, and large) shows that people using multiple modes are also more likely to be in the HH and HL groups. This also implies that people in the HH and HL live and work in environments that motivate the use of combinations of modes (Vich et al., 2019). As expected, people who use more diverse travel modes and spend less time at home are prone to have high complexity. Moreover, we find that the considerable driving distance between households and workplaces discourages people from creating complex daily routines, and that people are less likely to have high fragmentation levels between Monday and Wednesday presumably due to time commitments required for commuting and working at a workplace.

Regarding household structure, single persons with no children are split between two different spatial clusters but consistently surrounded by high complexity neighbors (HH and LH). The presence of children at different age groups plays different roles to spatial cluster membership. This may indicate that residential location choice, fragmentation, and household structure are strongly correlated in a way that household structure used in a regression model like this is capture multiple variables. Related to this is also the impact of household annual income that shows households at higher wealth levels tend to be in the HH cluster and then the LH cluster but not the other two.

The living environment indicators show a clear trend for the HH cluster with higher probability to belong in this cluster for people residing in the center of a city and its attenuation as we move to the suburbs. In contrast for the HL cluster all urban areas (core, district, and neighborhood) show higher and similar propensity of membership. This shows that low fragmentation people with high fragmentation neighbors are more likely to live in the most urbanized areas. This aligns well with other research showing urban areas that provide many activity opportunities (e.g., entertainment venues) and better-developed transportation facilities motivate and enable residents to go out frequently (Hough et al., 2008). The findings also support the conclusions that people who have higher propensity to be more active outside their homes, are able to find jobs in the center of cities, are also able to afford the higher cost of living in these urban environments offering more opportunities supporting a high fragmentation lifestyle. This creates unintended gentrification and has implications for the way people travel in these environments (Chatman et al., 2019). The opposite trend is found for the LL and HL clusters. Furthermore, twelve dummy variables representing counties are introduced to test if local travel habits and availability of infrastructure are correlated with clustering of fragmentation (we use the 12 county indicators that were found to be significantly different than zero to study propensity to belong in each of these spatial clusters). The five counties of Tehama, Tuolumme, Calaveras, Imperial, and Siskiyou, that are among the less urbanized California counties and offer less opportunities for public transportation, have large positive coefficients associated with the LL cluster. People living in these counties who are not in urban environments are more likely to populate this cluster. In contrast, people living in San Francisco and Marin (this is the county located North of San Francisco reachable by the Golden Gate Bridge) are more likely to be in the HH group supporting the hypothesis that there are other factors beyond degree of urbanization that determine higher schedule fragmentation. In fact, Madera, another county with high and positive coefficient for the HH group, is not a highly urbanized county and includes a large swath of land dedicated to open space. This support further the county by county exploration in Section 4.5.2. There are significant geographic differences among the spatial clusters and activity-travel fragmentation that cannot be attributed to urbanization alone requiring further scrutiny.

Table 4.5. Estimated parameters of multinomial logit model.

Variables	HH		HL		LH		LL	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Individual attributes								
Age (base case: over 45)								
Under 18	0.622***	0.128	-1.753***	0.281	0.343**	0.162	0.378***	0.124
18 to 25	0.028	0.111	-0.444***	0.169	0.038	0.148	0.043	0.113
26 to 45	-0.004	0.065	-0.357***	0.096	0.128	0.090	-0.014	0.076
Women	0.126***	0.047	0.023	0.066	-0.107*	0.060	-0.091**	0.046
Full-time	-0.391***	0.070	-0.063	0.096	0.130	0.095	0.670***	0.074
Driver	-0.072	0.105	0.176	0.173	-0.206*	0.110	-0.008	0.089
Walk	0.010***	0.003	0.004	0.004	-0.007	0.004	-0.009**	0.004
Bike	-0.006	0.010	-0.075***	0.026	-0.014	0.017	-0.078***	0.020
Public Transit	-0.011**	0.005	-0.014	0.009	0.016***	0.005	0.012**	0.006
Ridesharing	0.019**	0.009	-0.032	0.024	0.036***	0.010	-0.102***	0.029
Gini index	1.907***	0.098	1.574***	0.138	-1.667***	0.161	-1.573***	0.125
Travel time ratio	0.559***	0.109	0.008	0.152	-0.372***	0.130	-0.401***	0.096
Stay-at-home ratio	-5.086***	0.198	-3.540***	0.270	3.460***	0.269	4.548***	0.212
Home-work distance	-1.962**	0.832	-0.079	0.726	0.740	0.593	0.233	0.577
Travel day (base case: Thursday and Friday)								
Monday	-0.225***	0.068	-0.023	0.093	-0.014	0.084	-0.016	0.065
Tuesday	-0.248***	0.063	-0.048	0.087	-0.080	0.079	-0.022	0.061
Wednesday	-0.125**	0.061	-0.066	0.087	-0.129	0.081	-0.068	0.062
Household attributes								
Household income $50,000 or more	0.278***	0.058	-0.287***	0.074	0.318***	0.071	-0.540***	0.049
Household structure (base case: 2 or more retirees without children)								
One adult, no children	0.688***	0.096	0.154	0.120	1.002***	0.115	-0.427***	0.109
2+ adults, no children	-0.001	0.085	-0.652***	0.109	0.206**	0.103	-0.561***	0.081
One adult, youngest child 0-5	0.501*	0.268	-0.368	0.541	0.748**	0.361	-0.442	0.343
One adult, youngest child 6-15	-0.290*	0.161	-0.539*	0.279	0.061	0.235	-0.269	0.171
One adult, youngest child 16-21	0.136	0.223	0.279	0.273	0.315	0.310	-1.244***	0.391
2+ adults, youngest child 0-5	-0.207*	0.108	-0.242	0.148	-0.153	0.143	0.076	0.097
2+ adults, youngest child 6-15	0.208**	0.093	-0.290**	0.126	0.002	0.133	-0.040	0.092
2+ adults, youngest child 16-21	-0.216*	0.125	-0.552***	0.169	-0.007	0.164	-0.056	0.107
One adult, retired, no children	0.365***	0.121	0.440***	0.120	0.809***	0.110	-0.298***	0.090

Table 4.5 (continued). Estimated parameters of multinomial logit model.

Variables	HH		HL		LH		LL	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Spatial attributes								
Residential location type (base case: Suburban Neighborhood and Non-urban area)								
Urban Core	0.578***	0.211	-0.321	0.584	0.643**	0.256	-2.125**	1.018
Urban District	0.254*	0.152	-2.275***	0.772	0.642***	0.176	-1.173***	0.377
Urban Neighborhood	0.211**	0.092	-0.409**	0.175	0.585***	0.111	-0.279**	0.131
San Francisco	0.760***	0.152	-0.585	0.493	1.012***	0.177	-0.190	0.338
Marin	0.797***	0.212	-0.422	0.508	0.822***	0.306	-0.601	0.457
Madera	0.984***	0.378	0.103	0.723	0.871*	0.477	-0.227	0.544
Tehama	-1.358***	0.463	1.218***	0.218	-1.050**	0.519	0.481***	0.183
San Benito	-0.393	0.613	0.834*	0.477	-0.923	1.013	0.272	0.423
Tuolumne	-0.140	0.310	0.688***	0.261	-0.527	0.429	0.549***	0.186
Mariposa	-2.989	2.417	-0.208	0.742	0.830*	0.469	-0.152	0.490
Glenn	0.225	0.335	0.031	0.463	0.681*	0.363	-1.451**	0.635
Amador	0.657**	0.334	0.109	0.477	0.625*	0.380	-0.279	0.376
Calaveras	-0.990	0.723	-0.046	0.589	-1.434	1.005	1.234***	0.219
Imperial	-2.745	2.629	0.047	1.027	0.025	0.729	0.850**	0.385
Siskiyou	-1.073**	0.459	0.258	0.339	-0.234	0.427	0.536***	0.204
Constant	0.334*	0.192	-0.392	0.277	-5.299***	0.255	-4.766***	0.199
Summary statistics								
Observations		26,078						
Log-Likelihood Restricted		-24451 (df=4)						
Log-Likelihood Unrestricted		-21853 (df=172)						
Log-Likelihood Ratio		5194.9 (p<0.01***)						
McFadden R^2		0.106						

Note: *p<0.1; **p<0.05; ***p<0.01.

4.5.6 Conclusions

In this study, we firstly defined a fragmentation indicator based on 21 distinct states to represent the complexity of people's daily schedules. Then a spatial autocorrelation analysis was carried out to identify if clusters exist in terms of people's fragmented schedules over space. Finally, a hierarchical clustering algorithm was used to understand within spatial cluster diversity and a MNL model was developed to identify determinants fragmentation clustering using social, demographic, and place of residence type variables.

The results of spatial clustering show that the majority (more than 70%) of the complexity values of the people in the LISA analysis are not different than a randomized spatial distribution, which indicates people with average fragmented schedules can be found anywhere in California. With respect to the remaining approximately 30% of the sample clustered in four groups, it is noteworthy that the HH respondents are more likely to live in coastal and well-developed regions (e.g., San Francisco, Marin County), while the LL samples are more likely to live in less developed areas (e.g., Calaveras, Imperial, Siskiyou). Furthermore, the LH points seem to be near the HH samples and HL persons live close to the LL people, which potentially explains why the coefficients of the HH and LH clusters (or LL and HL clusters) in the MNL models are similar with each other in terms of household income and built-environment characteristics around the household residence. The within spatial groups hierarchical cluster analysis of time allocation sequences demonstrate substantial heterogeneity in time allocation in all spatial clusters. The HH cluster has three distinct daily patterns (including typical work day, school day, and discretionary day), but the HL clusters have only two distinct daily patterns (including typical work day and discretionary day). As for the LH and LL clusters, they have similar activity patterns, but the LH features a late work day in addition to the other three common daily patterns (home day, school day, and typical work day). Since children are most often accompanied to school, it is almost impossible for them to have higher complexity than the adults that care for them. As a result, there is no school day pattern in the HL cluster. Because of the significant variability within each LISA cluster, integrating LISA analysis with activity-travel sequence analysis appears to be an informative technique for identifying heterogeneity in space and time while accounting for spatial similarity. Moreover, this combination of two analytical methods allows for a holistic investigation of people's spatial distributions of multiple travel behavior indicators (e.g., travel by purpose and activity participation patterns in a day).

In the introduction of this chapter we also set the goal to identify the regions where people are more likely to use specific modes. The estimates of MNL model suggest that people in HH groups tend to use multiple modes, not simply the vehicle, but they are less likely to use public transit. This has the unintended consequence of adding to congestion and air pollution due to the use of ride hailing services offered with cars that are not zero emission vehicles. In those regions, the government can restrict people from driving private cars in city centers by parking pricing policies, incentivize the purchase of zero emission vehicles by TNCs, and enhance public transportation and walking and biking infrastructure as it is already promoted by Sustainable Community Strategies but with guided development (Mawhorter et al., 2020). In low-complexity areas, additional public and commuter bus lines can be built to lessen social exclusion, as residents are more likely to use public transit in those places. Additionally, since urban areas commonly have a high travel time ratio, which may be a result of traffic congestion, and the significant negative coefficient of distance between home and workplaces indicates a shorter commute and local employees can walk, cycle, or take the bus to work. Another possible policy here is time-flexible work that is also family friendly leading to less congestion and reduced costs to employers (Halpern, 2005; Van et al., 2019). This may decrease fragmentation and increase health benefits. However, it is unknown if it will lead to positive environmental impacts unless it is coupled with land use policies that increase density and diversity of activity opportunities and provision of multiple mode options and especially non-motorized transport options. Related to this is also the time of day of school scheduling. Most schools in the US end the school day before the end of the parents work day (Brown et al., 2016; Halpern, 2005) and implicitly favor a small fraction of children that are performing better early in the day (Callan, 1998). In fact, time flexibility for school going children may be another possible solution in releasing time pressure and achieving other equity objectives.

A third aspect addressing time poverty due to high fragmentation will also require to expand the policy action repertory in transportation and include customized services under integrated land use – transportation policies. This includes daycare centers for workers that are also parents, motivate cities to locate schools centrally in residential neighborhoods (Kim et al., 2016), and organize walking school buses (Kearns et al., 2003; Nikitas et al., 2019).

In addition to explicitly spatial-centered policies a policy that is already happening at very large scales due to the COVID-19 pandemic is the enabled and targeted substitution of physical mobility with virtual mobility for the persons that need this the most (see example in Kenyon et al., 2002). Despite the fact that certain research demonstrate that telecommuting can help mitigate some environmental impact (Harpaz, 2002), caution on this substitution should be mentioned because tele-life may lead to even higher levels of out of home activity-travel fragmentation indicating that targeting specific segments requires a fine resolution analysis in space, time, and social segmentation to discern fine differences in all dimensions (Su et al., 2021a) and then develop customized services for different segments of the population that are location specific.

Although this study provided a comprehensive method for analyzing the distribution characteristics of people's daily schedules and the relationship between their travel behavior and living environments, the built environment data employed in this study does not have the spatial and content resolution to explain spatial clustering completely. Therefore, we plan to extend the analysis here using much finer classification of the built environment (e.g., enumerating business establishments by business type in distance-based and travel time-based buffers, develop fine resolution in space and time availability of opportunities, and classify different places in California based on residential development, business availability, and recreational site availability). Additionally, while the current analysis provides a comprehensive comparison among different spatial clusters, it does not examine the various effects of each independent variable on people's fragmentation in different areas. In the future, we can experiment with spatial econometric models to address this issue. Moreover, time poverty in this analysis was only implicitly addressed and in a subsequent analysis we will need to create a more detailed taxonomy of type of time poverty (e.g., time poverty emerging from long commutes vs time poverty emerging from high levels of fragmentation). Another limitation of the study here is due to absence of more detailed information in the data used about time-allocation at home and other locations where people spend large amounts of time in a day. All of the above are left as future tasks.

5. Destination Choice and the 20 minute Walking City

In Chapter 1 of this report we provided a short review of the recent popularity in the idea of a "20 minute city" receives. In transportation planning is important to consider pedestrian (walking) accessibility in the coordination of land use with transportation facility provision because this is one of the most important benefits of increasing land use density and diversity. Computing indicators of walking accessibility entails many considerations that are different from automobile accessibility because the networks need to be constructed in different ways. Liu et al (2021) discuss the issues in computing pedestrian accessibility in detail and provide a few solutions. In this research we use the methods and software mentioned in that paper. The ingredients to compute accessibility for each US Census block at different travel time buffers by different modes are the same described in Chapters 2 and 3 of this report. In Chapter 3 we attached accessibility indicators to the place of residence of survey participants and reported the significance of these indicators in explaining daily VMT by type. In this chapter we take that analysis one step further and consider the daily pattern experienced accessibility, which is the accessibility a person enjoys in a day as they visit different locations. To do this we use the ideas of motifs in Chapter 1 to count for each person distinct locations. Then we compute for each of these visited locations the amount of opportunities encountered within 20 minutes of walking time. If a person visits the same location repeatedly in a day we do not count the opportunities multiple times. In this way we extract the benefit of variety seeking in destinations (Borgers et al., 1989) and the within-a-day sequencing of destination choices (Davis et al., 2020). Then, using the same reasoning of experienced accessibility as the outcome of daily behavioral pattern and amenities provided by the built and natural environments we analyze their correlation with accessibility.

5.1 Destination choice and walk accessibility

Figure 5.1 shows the amounts of experienced 20-minute walking accessibility by type as a function of the number of distinct destinations visited in a day by the CHTS survey participants. This is done by summing up the number of accessible opportunities by type within 20 minute walk distance (see section 3.3.1 the description of deriving the accessibility indicators) from each distinct location each individual visited in a day. These accessibility indicators include accessible area of three types of open spaces (in acres), accessible numbers of employees in 14 industry categories, and accessible lane miles of five types of roads (retrieved from OSM). People visited more than 9 distinct locations are lumped together in Figure 5.1 because they only account for 0.8% of the total CHTS sample and for the sake of clearly showing the change trends of accessibility indicators along with increasing number of distinct locations visited. As shown in the figure, many of these accessibility indicators grow exponentially with the number of destinations with a few exceptions in the area open space (ocean view and far ocean), access to service and motorways/freeways, and very few employment types. We explore this relationship in depth by considering two examples of 20-minute walking experienced accessibility (i.e., retail and education accessibility) accounting for other factors that also provide evidence of disparities in experienced accessibility.

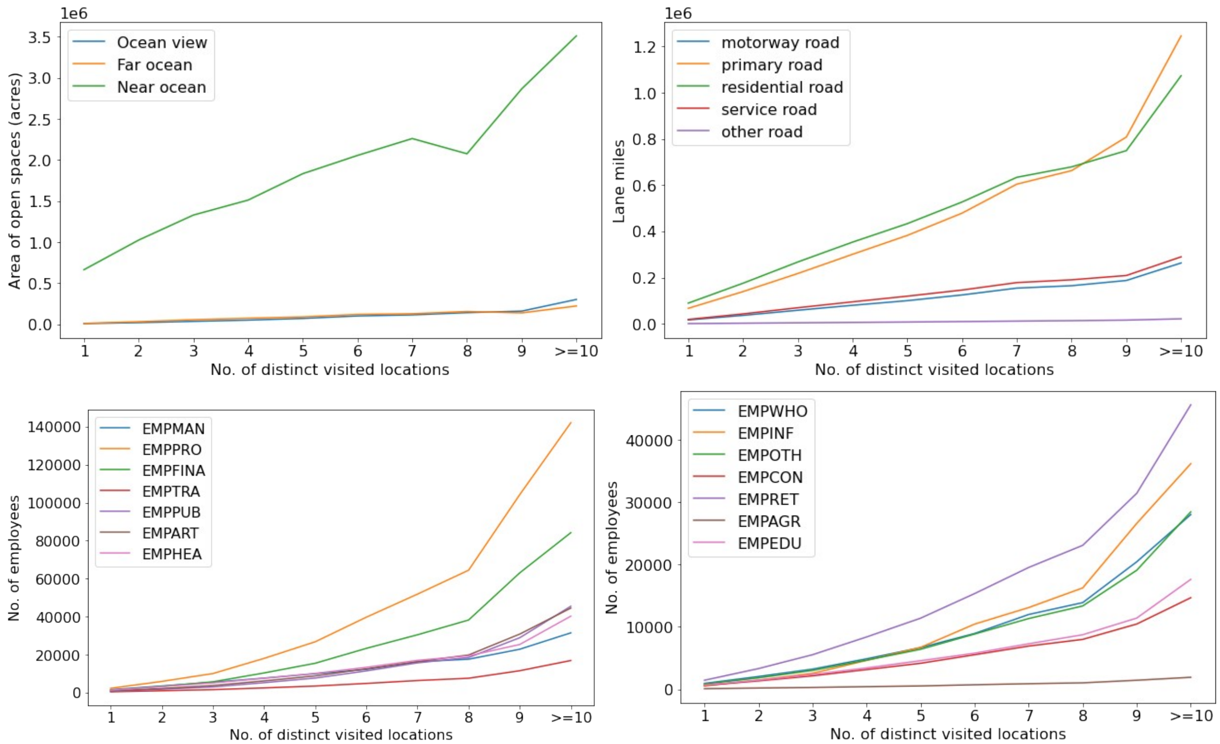

Figure 5.1 Relationships between number of distinct destinations visited in a day and experienced accessibility by type.

5.2 Regression models of experienced accessibility

The same approach, quantile regression as in Chapter 3 is applied to explore the heterogeneity in the relationship of 20-minute walking experienced accessibility with people's characteristics, travel behavior and the built environment. Following Chapter 3, 20-minute walking experienced accessibility to different types of opportunities is converted to density scale by averaging over the total area of accessible blocks within 20 minute walk. In this section, we discuss the results from the quantile regressions and compare them with the traditional OLS models that estimate the conditional mean of 20-minute walking experienced accessibility. To avoid multicollinearity, Variance Inflation Factor (VIF) was checked for each explanatory variable. Based on VIF, variables that are highly correlated with other variables were excluded in the final models. The final specification of each model was obtained by a systematic process of eliminating insignificant variables.

5.2.1 Retail accessibility regression

Table 5.1 displays the outcomes of the mean model and the median quantile regression model using as dependent variable 20-minute walking accessibility to retail. First, we report the differences between mean and median models in terms of the significance of the estimates. In general, mean and median models present high similarity in statistical inference/significance but a few notable differences emerge. The dummy variable 'Female' is significant in the mean model but becomes insignificant in the median model. The coefficient of 'Female' in the mean model is -8.154 indicating that on average, women have lower accessibility to retail. The estimate of the dummy variable 'age under 18' is not significant in the mean model but is significant in the median model. This means at median level of retail accessibility, people aged under 18 have lower accessibility to retail than people aged above 65. In addition, the dummy variables '\$50k to \$100k', '\$100k to \$200k', and 'has children aged 16 to 18' are only significant in the mean model, while the variable VMT as passengers is only significant in the median model.

The signs of the coefficients that are both significant in the mean and median models are consistent. However, there are several main differences regarding the size of the coefficients between the mean and median models. For instance, people aged 25 to 34 on average have 22.995 more retail accessibility in density scale than people above 65 as shown in the mean model. However, the median model shows that people aged 25 to 34 with retail accessibility at the 0.5 quantile only have 8.344 less retail accessibility than people above 65. Participants with higher educational attainment, higher annual household income, employed, and who live in densely populated areas tend to have higher retail accessibility compared to their counterparts in less densely populated areas. However, the mean model shows larger discrepancies compared to the median model. Participants who live with children and own more vehicles tend to have lower retail accessibility. However, the discrepancies are smaller at 0.5 quantile of retail accessibility compared to the outcome in mean model.

The SVI indicators suggest that people living in locations with higher vulnerability in terms of their socioeconomic status and minority status and language have lower retail accessibility. But people who live in areas of higher vulnerability in terms of housing and transportation tend to have higher retail accessibility. Presumably, these are people who have lower car ownership and live in dense areas where retail opportunities are more concentrated. Similarly, the mean model shows larger coefficients compared to median model. The three LPA dummy variables indicate that people in LPA1 publicAdmin and LPA2 highDensity enjoy higher retail accessibility and people in LPA4 suburban have lower retail accessibility compared to people in LPA3 lowDensity. This is reasonable as retail opportunities are less concentrated in low density and suburban areas (plus suburban areas tend to be predominantly residential). The three VMT variables suggest that people who drive alone or drive someone else more are more likely to have lower retail accessibility. The coefficients of the number of distinct visited locations are both positive in the mean and median models which implies that more distinct destinations people visit in a day, they are more likely to experience higher retail accessibility. However, the coefficients of the square of the number of distinct visited locations are negative for the mean and median models suggesting that the increment of retail accessibility by visiting additional destination declines as the number of distinct visited locations increases (i.e., decreasing marginal benefits with increasing distinct destinations visited in a day).

Table 5.1 OLS versus quantile regression on retail accessibility within 20min walk

Variables	Dependent variable: Retail accessibility	
	(1)Mean	(2)Median
Female (base: not female)	−8.154∗∗∗ (2.110)	−1.475 (0.903)
Age group (base: above 65)		
age under 18	2.836 (4.359)	−3.385∗∗ (1.639)
age 18 to 24	9.966∗∗ (4.881)	3.528∗ (2.086)
age 25 to 34	22.995∗∗∗ (4.624)	8.344∗∗∗ (2.065)
age 35 to 50	13.591∗∗∗ (3.609)	3.079∗ (1.592)
age 51 to 65	2.276 (3.110)	1.147 (1.275)
Education (base: below bachelor)		
some college	11.027∗∗∗ (3.336)	2.295∗ (1.243)
above bachelor	21.510∗∗∗ (2.773)	12.777∗∗∗ (1.266)
Household annual income (base: less than $25k)		
$25k to $50k	3.764 (3.334)	−1.438 (1.336)
$50k to $100k	14.931∗∗∗ (3.107)	−1.214 (1.267)
$100k to $200k	24.568∗∗∗ (3.484)	1.416 (1.503)
more than $200k	69.091∗∗∗ (5.142)	14.757∗∗∗ (2.759)
Employed (base: not employed)	41.564∗∗∗ (2.560)	11.799∗∗∗ (1.111)
Household children structure (base: no children)		
has children aged under 4	−9.786∗∗∗ (3.696)	−4.800∗∗∗ (1.370)
has children aged 4 to 15	−40.358∗∗∗ (2.630)	−23.804∗∗∗ (1.094)
has children aged 16 to 18	−7.482∗∗ (3.181)	−0.796 (1.352)
# of vehicles	−8.007∗∗∗ (1.129)	−2.593∗∗∗ (0.437)
Population density 20min	156.050∗∗∗ (0.602)	99.826∗∗∗ (0.551)
SVI Socioeconomic status	−89.728∗∗∗ (6.497)	−75.334∗∗∗ (2.760)
SVI Minority Status & Language	−250.544∗∗∗ (5.841)	−70.583∗∗∗ (2.226)
SVI Housing & Transportation	111.192∗∗∗ (4.832)	79.597∗∗∗ (2.181)
Residential built environment (base: LPA3 lowDensity)		
LPA1 publicAdmin	75.097∗∗∗ (7.290)	37.411∗∗∗ (4.556)
LPA2 highDensity	30.599∗∗∗ (3.989)	61.639∗∗∗ (2.519)
LPA4 suburban	−26.999∗∗∗ (2.745)	−3.116∗∗∗ (1.049)
# of distinct visited locations	55.286∗∗∗ (1.930)	47.834∗∗∗ (0.972)
Square of # of distinct visited locations	−2.318∗∗∗ (0.220)	−2.521∗∗∗ (0.128)
VMT drive alone	−0.177∗∗∗ (0.049)	−0.080∗∗∗ (0.021)
VMT drive someone else	−0.180∗∗∗ (0.048)	−0.136∗∗∗ (0.021)
VMT as passengers	0.052 (0.040)	−0.039∗ (0.020)
Constant	−73.769∗∗∗ (5.946)	−53.614∗∗∗ (2.407)
Observations	114,126	114,126

Note: ∗p<0.1; ∗∗p<0.05; ∗∗∗p<0.01

5.2.2 Education accessibility regression

Table 5.2 presents the outcomes of the mean and the median models using as dependent variable 20-minute walking accessibility to education opportunities. We first report the discrepancies in terms of the significance of variables between mean and median models. The dummy variables ‘age 35 to 50’ and ‘$50k to $100k’ are significant only in the mean model. On the other hand, the variables ‘age 51 to 65’, ‘has children aged 16 to 18’, ‘SVI Socioeconomic status’, ‘LPA2 highDensity’, ‘VMT drive alone’, and ‘VMT as passengers’ are only significant in the median model. Similar to the observation in Table 5.1, the signs of the coefficients that are both significant in the mean and median models are consistent. However, there are a few main differences in terms of the size of the coefficients between the mean and median models. We summarize below a few major findings.

Female and non-female survey respondents do not have significant difference in accessibility to education opportunities as shown in both the mean and median models. As the mean model shows, people aged 18 to 24 tend to have the highest accessibility to education followed by age groups 25 to 34, under 18, and 35 to 50. People aged above 65 and aged 51 to 65 have the lowest accessibility to education and there is no significant difference between them. However, the median model suggests smaller discrepancies between all age groups and the age group 18 to 24 still has the highest accessibility to education but followed by age groups under 18, 25 to 34, and 51 to 65. And people above 65 and aged 35 to 50 show no significant difference and have lowest accessibility to education opportunities. People with higher educational attainment, higher annual household income, employed, and who live in densely populated areas tend to have higher experienced accessibility to education compared to their counterparts. However, the mean model shows larger discrepancies compared to the median model. Both the mean and median models suggest that participants who have children aged under 4 and aged 4 to 15 tend to have lower accessibility to education presumably they do not visit locations that add to their experienced accessibility. However, the median model also shows that participants who have children aged 16 to 18 tend to have higher accessibility to education compared to those who have no children and this is an indication of driving their children to schools. The coefficients of the number of vehicles in both models are negative and significant indicating that people who own more vehicles are likely to have lower accessibility to education. In terms of the three SVI indicators, both mean and median models suggest that people living in places with higher index of vulnerability in terms of their minority status and language have lower accessibility to education and people who are more vulnerable in terms of housing and transportation tend to have higher accessibility to education (similar to retail). The variable of SVI Socioeconomic status is negative and only significant in the median model indicating that at 0.5 quantile level of education accessibility, places of higher vulnerability in terms of their socioeconomic status have lower accessibility to education. The three LPA dummy variables indicate that people in LPA1 publicAdmin enjoy higher accessibility to education (e.g., university campuses) and people in LPA4 suburban have lower accessibility to education compared to people in LPA3 lowDensity. The mean model suggests no significant difference between people in LPA2 highDensity and LPA3 lowDensity but the median model suggests people in LPA2 highDensity enjoy higher accessibility to education than people in LPA3 lowDensity. Both models agree that the more people drive someone else, the less their experienced accessibility to education becomes. The median model also suggests that people who drive alone or taking rides as passengers more are more likely to have lower accessibility to education. Similarly, we observe the same findings regarding the coefficients of the number of distinct visited locations and the square of the number of distinct visited locations: more distinct destinations people visit in a day, more likely to enjoy higher accessibility to education opportunities they are. The increment of the accessibility by visiting additional destination declines as the number of distinct visited locations increases (i.e., decrease in marginal returns to experienced accessibility to education).

Table 5.2 OLS versus quantile regression on education accessibility within 20min walk

Variables	Dependent variable: Education accessibility	
	(1)Mean	(2)Median
Female (base: not female)	−1.507 (1.223)	−0.086 (0.204)
Age group (base: above 65)		
age under18	13.845∗∗∗ (2.527)	4.621∗∗∗ (0.367)
age 18 to 24	42.979∗∗∗ (2.830)	4.888∗∗∗ (0.496)
age 25 to 34	20.302∗∗∗ (2.681)	2.748∗∗∗ (0.486)
age 35 to 50	6.082∗∗∗ (2.092)	0.389 (0.357)
age 51 to 65	0.430 (1.803)	0.656∗∗ (0.290)
Education (base: below bachelor)		
some college	5.933∗∗∗ (1.934)	−0.218 (0.262)
above bachelor	26.373∗∗∗ (1.608)	4.117∗∗∗ (0.333)
Household annual income (base: less than $25k)		
$25k to $50k	−0.302 (1.933)	−0.008 (0.277)
$50k to $100k	3.578∗∗ (1.801)	−0.123 (0.265)
$100k to $200k	10.472∗∗∗ (2.020)	0.812∗∗ (0.318)
more than $200k	28.510∗∗∗ (2.981)	5.574∗∗∗ (0.724)
Employed (base: not employed)	11.299∗∗∗ (1.484)	3.263∗∗∗ (0.261)
Household children structure (base: no children)		
has children aged under 4	−15.282∗∗∗ (2.143)	−3.978∗∗∗ (0.301)
has children aged 4 to 15	−7.125∗∗∗ (1.525)	−1.961∗∗∗ (0.249)
has children aged 16 to 18	0.551 (1.844)	1.750∗∗∗ (0.312)
# of vehicles	−6.058∗∗∗ (0.654)	−1.075∗∗∗ (0.071)
Population density 20min	63.724∗∗∗ (0.349)	40.690∗∗∗ (0.177)
SVI Socioeconomic status	5.697 (3.767)	−2.712∗∗∗ (0.601)
SVI Minority Status & Language	−112.708∗∗∗ (3.386)	−24.502∗∗∗ (0.496)
SVI Housing & Transportation	43.002∗∗∗ (2.801)	16.515∗∗∗ (0.533)
Residential built environment (base: LPA3 lowDensity)		
LPA1 publicAdmin	18.117∗∗∗ (4.226)	3.636∗∗∗ (1.054)
LPA2 highDensity	−3.074 (2.313)	4.009∗∗∗ (0.695)
LPA4 suburban	−7.840∗∗∗ (1.591)	−6.932∗∗∗ (0.231)
# of distinct visited locations	15.448∗∗∗ (1.119)	7.080∗∗∗ (0.257)
Square of # of distinct visited locations	−0.777∗∗∗ (0.127)	−0.271∗∗∗ (0.036)
VMT drive alone	−0.040 (0.028)	−0.018∗∗∗ (0.005)
VMT drive someone else	−0.075∗∗∗ (0.028)	−0.021∗∗∗ (0.007)
VMT as passengers	0.019 (0.023)	−0.015∗∗∗ (0.004)
Constant	−34.746∗∗∗ (3.447)	−15.696∗∗∗ (0.541)
Observations	114,126	114,126

Note: ∗p<0.1; ∗∗p<0.05; ∗∗∗p<0.01

6. Summary and Conclusions

In Chapter 1 we explain the reasoning and provide motivation for detailed spatiotemporal analysis of amenities provided to the California population to identify geographical pockets for which current policy may have unintended consequences. The three most important needs identified are: 1) understand the possible difference in sensitivity of travel behavior to the land use policy variables such as density of opportunities; 2) identify places in which fragmentation of time allocation makes it necessary to use the private automobile; 3) understand what motivates people to visit multiple distinct destinations in a day and the relationship of this behavior to availability of opportunities.

In Chapter 2 we explore the classification of every US Census block in California using latent Profile Analysis and a detailed inventory of business establishments. We derive in this way four distinct type that depend on the business establishment density. Then, we use the Social Vulnerability Indices developed by CDC to explore the correlation between place of residence land use characteristics and SVI. Coastal areas have the lowest SVI and in urbanized environments also higher density of opportunities. There are many rural areas that also have high vulnerability. This motivates the microanalytic model specification developed in the rest of the chapters in this report.

In Chapter 3 we develop and present a new type of regression model with vehicle miles traveled (VMT) as dependent variable. The motivation for doing this is that VMT is a primary policy variable because it is strongly correlated with greenhouse gas emissions and the density and diversity of urban environments. Policies that increase density and diversity of land use are believed to decrease VMT presumably by replacing it with non- motorized travel. This causal inference has not been conclusive and may create further disparities by gentrification and an increase in long-distance travel for lower-income commuters. In this study, we explore the heterogeneity in the relationships between personal level VMT and the built environment, accessibility to opportunities and open spaces, and people's socio-demographic traits across various levels of travel demand. The present research first develops opportunity-based measurement of residential land-use indicators at the US census block level for the whole California. This is followed by the use of the Latent Profile Analysis from Chapter 2 as the land-use indicators identifying distinct land use patterns experienced by the respondents in the 2012-2013 California Household Travel Survey. Thus, we estimate quantile regression models to understand the heterogeneity in the relationships between VMT and residential built environment characteristics and people's socio-demographic traits. The results indicate different sensitivity to land use at various travel intensities implying different response to land use policies.

In Chapter 4 using data from the 2017 National Household Travel Survey in California from 26,078 survey participants, sequence analysis is used to estimate a fragmentation indicator of people's daily schedules. Then, spatial clustering is used to find groups of observations with similarly high or low fragmentation using the longitude and latitude of their residential locations. This is followed by a hierarchical sequence clustering within each spatial cluster to identify distinct patterns of time allocation. Using the Local Indicator of Spatial Association (LISA) we find a large portion (approximately 30%) of the sample with significant spatial clustering of fragmentation. We also find systematic and significant differences in membership to these clusters based on land use, county of residence, household and personal characteristics, and travel modes used. Sequence analysis pattern recognition within LISA spatial clusters shows systematically repeating time allocation patterns that include typical work and school schedules as well as staying at home patterns. However, each spatial LISA cluster is composed of different time allocation clusters. All this analysis taken together points out substantial and measurable heterogeneity in spatial clustering of fragmentation and the need for customized policy actions in different geographies.

In Chapter 5 we explore the walking accessibility to opportunities by enumerating the distinct destinations visited in a day and the correlation between the number of destinations and accessibility. People that visit multiple locations also experience exponentially increasing with the number of locations walking accessibility. The 20 minute city is in essence composed of multiple destinations that are surrounded by many activity opportunities. We explore heterogeneity in experienced accessibility using multivariate regression for retail and education experienced accessibilities as a function of person and household characteristics, residence in one of the LPA types identified here, and SVI. Key findings include lower accessibility to retail and education opportunities for people living in places that are classified as populated by vulnerable residents (e.g., minority and lower income tracts). This analysis also indicates substantially higher and heterogeneous experienced accessibility among people that visit multiple distinct locations but with decreasing returns to the investment of visiting multiple locations. As expected living in higher density in terms of population and activity opportunities is also offering higher experienced walking accessibility even when we account for asymmetry in the distribution of accessibility indicators.

References

Akaike, H., 1987. Factor analysis and AIC. Psychometrika 52, 317–332.

Alexander, B., Ettema, D., & Dijst, M. (2010). Fragmentation of work activity as a multi-dimensional construct and its association with ICT, employment and sociodemographic characteristics. *Journal of Transport Geography*, *18*(1), 55–64. https://doi.org/10.1016/j.jtrangeo.2009.05.010

Alexander, B., Hubers, C., Schwanen, T., Dijst, M., & Ettema, D. (2011). Anything, Anywhere, Anytime? Developing Indicators to Assess the Spatial and Temporal Fragmentation of Activities. *Environment and Planning B: Planning and Design*, *38*(4), 678–70.

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, *27*(2), 93–115.

Auld, J., Mohammadian, A. (Kouros), & Roorda, M. J. (2009). Implementation of Scheduling Conflict Resolution Model in Activity-Scheduling System. *Transportation Research Record*, *2135*(1), 96–105. https://doi.org/10.3141/2135-12

Auld, J., Taha, H. R., Mahmoud, J., & Abolfazl, M. (2011). *Dynamic Activity Generation Model Using Competing Hazard Formulation*. *2254*(1), 28–35.

Batur, I., Sharda, S., Kim, T., Khoeini, S., Pendyala, R. M., & Bhat, C. R. (2019). *Mobility, Time Poverty, and Well-Being: How Are They Connected and How Much Does Mobility Matter?* 17.

Bauer, D.J., Curran, P.J., 2003. Overextraction of Latent Trajectory Classes: Much Ado About Nothing? Reply to Rindskopf (2003), MuthÅLen (2003),and Cudeck and Henly (2003). Publisher: American Psychological Association.

Bento, A.M., Cropper, M.L., Mobarak, A.M., Vinha, K., 2005. The effects of urban spatial structure on travel demand in the United States. Review of Economics and Statistics 87, 466–478.

Bhat, C. R., & Koppelman, F. S. (1999). A retrospective and prospective survey of time-use research. *Transportation*, *26*(2), 119–139. https://doi.org/10.1023/A:1005196331393

Bloomberg Businessweek (2020) https://www.bloomberg.com/news/features/2020-11-12/paris-s-15-minute-city-could-be-coming-to-an-urban-area-near-you. Accessed March 1, 2021.

Boarnet, M. G., Giuliano, G., Hou, Y., & Shin, E. J. (2017). First/last mile transit access as an equity planning issue. *Transportation Research Part A: Policy and Practice*, *103*, 296-310.

Boarnet, M., Handy, S., 2014. Impacts of residential density on passenger vehicle use and greenhouse gas emissions. Policy Brief. California Air Resources Board.

Boeing, G., 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems 65, 126–139. URL: https:// linkinghub.elsevier.com/retrieve/pii/S0198971516303970, doi:10. 1016/j.compenvurbsys.2017.05.004.

Borgers, A. W. J., Van Der Heijden, R. E. C. M., & Timmermans, H. J. P. (1989). A variety seeking model of spatial choice-behaviour. *Environment and Planning A*, *21*(8), 1037-1048.

Brown, S., Fite, P., & Poquiz, J. (2016). Moderating Effects of Gender on Outcomes Associated with Stressful Life Events Among Elementary School-Age Youth. *Child Psychiatry & Human Development*, *47*, 593–602.

Brownstone, D., 2008. Key relationships between the built environment and VMT. Transportation Research Board 7. California governor's office of planning and research, 2018. Technical advisory on evaluating transportation impacts in CEQA. Technical Report. URL: https://opr.ca.gov/docs/20190122-743_Technical_Advisory.pdf.

Callan, D. E. (1998). *An auditory feedback-based model of speech production in the developing child* [Ph.D., The University of Wisconsin - Madison]. https://www.proquest.com/docview/304489088/abstract/A3E6DBBAB2144EF3PQ/1

CalTrans. (2020). 2017 National Household Travel Survey. *Https://Nhts.Dot.ca.Gov/*.

Capasso Da Silva, D.; King, D.A.; Lemar, S. Accessibility in practice: 20-minute city as a sustainability planning goal. *Sustainability* **2020**, *12*, 129

Cervero, R., Duncan, M., 2006. 'Which Reduces Vehicle Travel More: Jobs-Housing Balance or Retail-Housing Mixing? Journal of the American Planning Association 72, 475–490. URL: https:// www.tandfonline.com/doi/full/10.1080/01944360608976767, doi:101080/01944360608976767.

Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: Density,diversity, and design. Transportation Research Part D: Transportand Environment 2, 199–219. URL: https://linkinghub.elsevier.

Chatman, D. G., Xu, R., Park, J., & Spevack, A. (2019). Does Transit-Oriented Gentrification Increase Driving? *Journal of Planning Education and Research*, *39*(4), 482–495. https://doi.org/10.1177/0739456X19872255

Chen, Y., Ravulaparthy, S., Deutsch, K., Dalal, P., Yoon, S. Y., Lei, T., ... & Hu, H. H. (2011). Development of indicators of opportunity-based accessibility. *Transportation Research Record*, *2255*(1), 58-68.

Couclelis, H. (2000). From Sustainable Transportation to Sustainable Accessibility: Can We Avoid a New Tragedy of the Commons? In D. G. Janelle & D. C. Hodge (Eds.), *Information, Place, and Cyberspace* (pp. 341–356). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-04027-0_20

Couclelis, H. (2004). Pizza over the Internet: E-commerce, the fragmentation of activity and the tyranny of the region. *Entrepreneurship & Regional Development*, *16*(1), 41–54. https://doi.org/10.1080/0898562042000205027

de Abreu e Silva, J., Morency, C., & Goulias, K. G. (2012a). Using structural equations modeling to unravel the influence of land use patterns on travel behavior of workers in Montreal. *Transportation Research Part A: Policy and Practice*, *46*(8).

de Abreu e Silva, J., Martinez, L. M., & Goulias, K. G. (2012b). Using a Multi Equation Model to Unravel The Influence of Land Use Patterns on Travel Behavior of Workers in Lisbon. *Transportation Letters: The International Journal of Transportation Research*, *4*(4).

de Abreu e Silva, J., Goulias, K.G., Dalal, P., (2012c). Structural Equations Model of Land Use Patterns, Location Choice, and Travel Behavior in Southern California. Transportation Research Record: Journal of the Transportation Research Board 2323, 35–45. URL: http://journals. sagepub.com/doi/10.3141/2323-05, doi:10.3141/2323-05.

Davis, A. W., McBride, E. C., & Goulias, K. G. (2020). Environmental correlates of travel behavior from a destination attractiveness and activity timing perspectives. In *Mapping the Travel Behavior Genome* (pp. 83-102). Elsevier.

Deutsch, K., Yoon, S. Y., & Goulias, K. (2013). Modeling travel behavior and sense of place using a structural equation model. *Journal of Transport Geography*, *28*, 155-163.

Dijst, M., & Vidakovic, V. (2000). Travel time ratio: The key factor of spatial reach. *Transportation*, *27*(2), 179–199.

Eluru, N., Pinjari, A., Pendyala, R., & Bhat, C. (2010). An econometric multi-dimensional choice model of activity-travel behavior. *Transportation Letters*, *2*(4), 217–230. https://doi.org/10.3328/TL.2010.02.04.217-230

Ettema, D., Borgers, A., & Timmermans, H. (1995b). Competing Risk Hazard Model of Activity Choice, Timing, Sequencing, and Duration. *Transportation Research Record*, *1493*, 101–109.

Ewing, R., 1997. Is Los Angeles-style sprawl desirable? Journal of the American planning association 63, 107–126. Publisher: Taylor & Francis.

Ewing, R., Cervero, R., 2001. Travel and the Built Environment: A Synthesis.Transportation Research Record: Journal of the Transportation ResearchBoard 1780, 87–114. URL: http://journals.sagepub.com/doi/10.3141/1780-10, doi:10.3141/1780-10.

Ewing, R., Cervero, R., 2010. Travel and the Built Environment:A Meta-Analysis. Journal of the American Planning Association76, 265–294. URL: http://www.tandfonline.com/doi/abs/10.1080/01944361003766766, doi:10.1080/01944361003766766.

Ewing, R., Hamidi, S., Gallivan, F., Nelson, A.C., Grace, J.B., 2014. Structural equation models of VMT growth in US urbanised areas. Urban Studies 51, 3079–3096. Publisher: Sage Publications Sage UK: London, England.

Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, *40*(4), 1. https://doi.org/10.18637/jss.v040.i04

Garikapati, V. M., Pendyala, R. M., Morris, E. A., Mokhtarian, P. L., & McDonald, N. (2016). Activity patterns, time use, and travel of millennials: A generation in transition? *Transport Reviews*, *36*(5), 558–584. https://doi.org/10.1080/01441647.2016.1197337

Gehl, J. (2011). *Life between buildings: using public space*. Island press.

Goulias, K. G. (2000). Traveler Behavior and Values Research for Human-Centered Transportation Systems. *The Pennsylvania State University, Committee on Traveler Behaviour and Values*, 6.

Goulias, K. G. (2002). Multilevel analysis of daily time use and time allocation to activity types accounting for complex covariance structures using correlated random effects. *Transportation*, *29*(1), 31–48. https://doi.org/10.1023/A:1012973004152

Goulias, K. G. (2018). Travel behavior models. *Handbook of Behavioral and Cognitive Geograph y*.https://www.elgaronline.com/view/edcoll/9781784717537/9781784717537.00010.xml

Goulias, K. G., McBride, E. C., & Su, R. (2020). Life cycle stages, daily contacts, and activity-travel time allocation for the benefit of self and others. *Mobility and Travel Behaviour Across the Life Course*. https://www.elgaronline.com/view/edcoll/9781789907803/9781789907803.00023.xml

Goulias, K. G., Pendyala, R. M., & Bhat, C. R. (2013). Keynote—Total Design Data Needs for the New Generation Large-Scale Activity Microsimulation Models. In J. Zmud, M. Lee-Gosselin, M. Munizaga, & J. Antonio Carrasco (Eds.), *Transport Survey Methods* (pp. 21–46). Emerald Group Publishing Limited. https://doi.org/10.1108/9781781902882-002

Greenfield, A. (2017). *Radical Technologies: The Design of Everyday Life*. Verso Books.

Grubesic, T. H., Wei, R., & Murray, A. T. (2014). Spatial Clustering Overview and Comparison: Accuracy, Sensitivity, and Computational Expense. *Annals of the Association of American Geographers*, *104*(6), 1134–1156. https://doi.org/10.1080/00045608.2014.958389

Hagberg, A., Swart, P., S Chult, D., 2008. Exploring network structure, dynamics, and function using NetworkX. Technical Report. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Halpern, D. (2005). How time-flexible work policies can reduce stress, improve health, and save money. *Stress and Health*, *21*(3), 157–168.

Handy, S., Cao, X., Mokhtarian, P., 2005. Correlation or causality between the built environment and travel behavior? Evidence from Northern California. Transportation Research Part D: Transport and Environment10, 427–444. URL: https://linkinghub.elsevier.com/retrieve/pii/S136192090500043X, doi:10.1016/j.trd.2005.05.002.

Harpaz, I. (2002). Advantages and disadvantages of telecommuting for the individual, organization and society. *Work Study*, *51*(2), 74–80. https://doi.org/10.1108/00438020210418791

Harvey, A. S., & Mukhopadhyay, A. K. (2007). When Twenty-Four Hours is not Enough: Time Poverty of Working Parents. *Social Indicators Research*, *82*(1), 57–77. https://doi.org/10.1007/s11205-006-9002-5

Hough, J. A., Cao, X., & Handy, S. L. (2008). Exploring Travel Behavior of Elderly Women in Rural and Small Urban North Dakota: An Ecological Modeling Approach. *Transportation Research Record*, *2082*(1), 125–131. https://doi.org/10.3141/2082-15

Hubers, C., Schwanen, T., & Dijst, M. (2008). ICT and Temporal Fragmentation of Activities: An Analytical Framework and Initial Empirical Findings. *Tijdschrift Voor Economische En Sociale Geografie*, *99*(5), 528–546. https://doi.org/10.1111/j.1467-9663.2008.00490.x

Islam, M. R., & Saphores, J.-D. M. (2022). An L.A. story: The impact of housing costs on commuting. *Journal of Transport Geography*, *98*, 103266. https://doi.org/10.1016/j.jtrangeo.2021.103266

Jacquez, G. M., Kaufmann, A., & Goovaerts, P. (2008). Boundaries, links and clusters: A new paradigm in spatial analysis? *Environmental and Ecological Statistics*, *15*(4), 403–419. https://doi.org/10.1007/s10651-007-0066-4

Jiao, J., Bischak, C., & Hyden, S. (2020). The impact of shared mobility on trip generation behavior in the US: Findings from the 2017 National Household Travel Survey. *Travel Behaviour and Society*, *19*, 1–7. https://doi.org/10.1016/j.tbs.2019.11.001

Joh, C. H., Arentze, T. A., & Timmermans, H. J. P. (2001). Multidimensional sequence alignment methods for activity-travel pattern analysis: A comparison of dynamic programming and genetic algorithms. *Geographical Analysis*, *33*(3), 247–270.

Joh, C.-H., Arentze, T. A., & Timmermans, H. J. P. (2001). A Position-Sensitive Sequence-Alignment Method Illustrated for Space–Time Activity-Diary Data. *Environment and Planning A: Economy and Space*, *33*(2), 313–338. https://doi.org/10.1068/a3323

Joh, C.-H., Arentze, T., Hofman, F., & Timmermans, H. (2002). Activity pattern similarity: A multidimensional sequence alignment method. *Transportation Research Part B: Methodological*, *36*(5), 385–403. https://doi.org/10.1016/S0191-2615(01)00009-1

Joh, K., Boarnet, M. G., Nguyen, M. T., Fulton, W., Siembab, W., & Weaver, S. (2008). Accessibility, travel behavior, and new urbanism: case study of mixed-use centers and auto-oriented corridors in the South Bay Region of Los Angeles, California. *Transportation Research Record*, *2082*(1), 81-89.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.

Kearns, R., Collins, D., & Neuwelt, P. (2003). The walking school bus: Extending children's geographies? *Area*, *35*(1), 285–292.

Kenyon, S., Lyons, G., & Rafferty, J. (2002). Transport and social exclusion: Investigating the possibility of promoting inclusion through virtual mobility. *Journal of Transport Geography*, *10*(3), 207–219. https://doi.org/10.1016/S0966-6923(02)00012-1

Kim, J.-H., Lee, C., & Sohn, W. (2016). Urban Natural Environments, Obesity, and Health-Related Quality of Life among Hispanic Children Living in Inner-City Neighborhoods. *International Journal of Environmental Research and Public Health*, *13*(1), 121. https://doi.org/10.3390/ijerph13010121

Kockelman, K., 1997. Travel behavior as function of accessibility, land usemixing, and land use balance: evidence from San Francisco Bay Area.Transportation research record 1607, 116–125. Publisher: SAGE PublicationsSage CA: Los Angeles, CA.

Koenker, R., Bassett Jr, G., 1978. Regression quantiles. Econometrica: Journal of the Econometric Society , 33–50Publisher: JSTOR.

Koenker, R., Hallock, K.F., 2001. Quantile regression. Journal of economic perspectives 15, 143–156.

Koenker, R., others, 2022. quantreg: Quantile Regression. R package version 5.88. R Foundation for Statistical Computing: Vienna) Available at: http://CRAN. R-project. org/package= quantreg .

Lee, J. H., Davis, A., Yoon, S. Y., & Goulias, K. G. (2017). Exploring daily rhythms of interpersonal contacts: Time-of-day dynamics of human interactions with latent class cluster analysis. *Transportation Research Record*, *2666*(1), 58-68.

Lei, T. L., Chen, Y., & Goulias, K. G. (2012). Opportunity-based dynamic transit accessibility in Southern California: measurement, findings, and comparison with automobile accessibility. *Transportation Research Record*, *2276*(1), 26-37.

Lenz, B., & Nobis, C. (2007). The changing allocation of activities in space and time by the use of ICT—"Fragmentation" as a new concept and empirical results. *Transportation Research Part A: Policy and Practice*, *41*(2), 190–204. https://doi.org/10.1016/j.tra.2006.03.004

Levinson, D. M. (2019). *The 30-minute city: Designing for access*. Network Design Lab.

Lin, T. G., Xia, J. C., Robinson, T. P., Goulias, K. G., Church, R. L., Olaru, D., ... & Han, R. (2014). Spatial analysis of access to and accessibility surrounding train stations: A case study of accessibility for the elderly in Perth, Western Australia. *Journal of Transport Geography*, *39*, 111-120.

Liu, F., Janssens, D., Cui, J., Wets, G., & Cools, M. (2015). Characterizing activity sequences using profile Hidden Markov Models. *Expert Systems with Applications*, *42*(13), 5705–5722. https://doi.org/10.1016/j.eswa.2015.02.057

Liu, S., Higgs, C., Arundel, J., Boeing, G., Cerdera, N., Moctezuma, D., Cerin, E., Adlakha, D., Lowe, M. and Giles-Corti, B. (2021). A generalized framework for measuring pedestrian accessibility around the world using open data. *Geographical Analysis*.

Lo, Y., Mendell, N.R., Rubin, D.B., 2001. Testing the number of components in a normal mixture. Biometrika 88, 767–778. Publisher: Oxford University Press.

Lucas, K. (2012). Transport and social exclusion: Where are we now? *Transport Policy*, *20*, 105–113. https://doi.org/10.1016/j.tranpol.2012.01.013

Mawhorter, S., Martin, A., & Galante, C. (2020). Chapter 22—California's SB 375 and the pursuit of sustainable and affordable development. In E. Deakin (Ed.), *Transportation, Land Use, and Environmental Planning* (pp. 497–521). Elsevier. https://doi.org/10.1016/B978-0-12-815167-9.00022-0

McBride, E. C., Adam, W. D., & Goulias, K. G. (2019). Fragmentation in Daily Schedule of Activities using Activity Sequences. *Transportation Research Record*, *2673*(4), 844–854.

McBride, E. C., Davis, A. W., & Goulias, K. G. (2020). Chapter 19—Sequence analysis of place-travel fragmentation in California. In K. G. Goulias & A. W. Davis (Eds.), *Mapping the Travel Behavior Genome* (pp. 371–398). Elsevier. https://doi.org/10.1016/B978-0-12-817340-4.00019-X

McBride, E. C., Davis, A. W., & Goulias, K. G. (2020). Exploration of statewide fragmentation of activity and travel and a taxonomy of daily time use patterns using sequence analysis in California. *Transportation research record*, *2674*(12), 38-51.

McBride, E.C., Davis, A.W., Goulias, K.G., 2018. A Spatial Latent Profile Analysis to Classify Land Uses for Population Synthesis Methods in Travel Demand Forecasting. Transportation Research Record: Journal of the Transportation Research Board 2672, 158–170. URL: http://journals.sagepub.com/doi/10.1177/0361198118799168,

McLachlan, G., Peel, D., 2000. Finite mixture models. New York, NY: Wiley

Merz, J., Böhm, P., & Burgert, D. (2009). Timing and fragmentation of daily working hours arrangements and income inequality–An earnings treatment effects approach with German time use diary data. *Electronic International Journal of Time Use Research*, *6*(2), 200–239.

Mitra, S. K., & Saphores, J.-D. M. (2019). Why do they live so far from work? Determinants of long-distance commuting in California. *Journal of Transport Geography*, *80*, 102489. https://doi.org/10.1016/j.jtrangeo.2019.102489

Mokhtarian, P.L., Cao, X., 2008. Examining the impacts of residential self selection on travel behavior: A focus on methodologies. Transportation Research Part B: Methodological 42, 204–228. URL: https://linkinghub.elsevier.com/retrieve/pii/S0191261507000744, doi:10.1016/j.trb.2007.07.006.

Moreno,C.;Allam,Z.; Chabaud, D.; Gall, C.; Pratlong, F. Introducing the "15-Minute City": Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities. *Smart Cities* **2021**, *4*, 93–111. https://doi.org/10.3390/smartcities 4010006

Muthen, L.K., Muthen, B.O., 1998. Mplus user's guide. Los Angeles, CA.

National Research Council, 2009. Driving and the Built Environment: The Effects of Compact Development on Motorized Travel, Energy Use, and CO2 Emissions. Technical Report.Washington, D.C.: National Academies Press.

Nelson, A.C., 2017. Compact Development Reduces VMT: Evidence and Application for Planners —Comment on "Does Compact Development Make People Drive Less?". Journal of the American Planning Association 83, 36–41. URL: https://www.tandfonline.com/doi/full/10.1080/01944363.2016.1246378, doi:10.1080/01944363.2016.1246378.

Nikitas, A., Wang, J. Y. T., & Knamiller, C. (2019). Exploring parental perceptions about school travel and walking school buses: A thematic analysis approach. *Transportation Research Part A: Policy and Practice*, *124*, 468–487. https://doi.org/10.1016/j.tra.2019.04.011

Novák, J., & Sýkora, L. (2007). A City in Motion: Time-Space Activity and Mobility Patterns of Suburban Inhabitants and the Structuration of the Spatial Organization of the Prague Metropolitan Area. *Geografiska Annaler: Series B, Human Geography*, *89*(2), 147–168. https://doi.org/10.1111/j.1468-0467.2007.00245.x

NuStats, 2013. 2010-2012 California Household Travel Survey Final Report. Technical Report. California Department of Transportation.

Nylund-Gibson, K., Choi, A.Y., 2018. Ten frequently asked questions about latent class analysis. Translational Issues in Psychological Science 4, 440–461. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/tps0000176, doi:10.1037/tps0000176.

Paez, A., & Scott, D. M. (2004). Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal*, *61*(1), 53–67.

Přibyl, O., & Goulias, K. G. (2005). Simulation of Daily Activity Patterns Incorporating Interactions within Households: Algorithm Overview and Performance. *Transportation Research Record*, *1926*(1), 135–141.

Rahn, C. (2012). *Suburban Daily Mobility – between Freedom and Sustainability*. 32nd International Geographical Congress, Session: Global Challenges & Local Responses, Köln. https://elib.dlr.de/80246/

Ravulaparthy, S., Dalal, P., Chen, Y., & Goulias, K. G. (2012). Exploratory Analysis of Spatial Hierarchical Clustering in Los Angeles County, California: Relationship of Opportunity-Based Accessibility, Reported Land Values, and Resident Characteristics. *Transportation Research Record*, *2307*(1), 132–140. https://doi.org/10.3141/2307-14

Rentziou, A., Gkritza, K., Souleyrette, R.R., 2012. VMT, energy consumption, and GHG emissions forecasting for passenger transportation. Transportation Research Part A: Policy and Practice 46, 487–500. URL: https://linkinghub.elsevier.com/retrieve/pii/S0965856411001789, doi:10.1016/j.tra.2011.11.009.

Rosenberg, J.M., Beymer, P.N., Anderson, D.J., Lissa, C.J.V., Schmidt, J.A., 2018. tidyLPA: An R Package to Easily Carry Out Latent Profile Analysis (LPA) Using Open-Source or Commercial Software. Journal of Open Source Software 3, 978. URL: https://joss.theoj.org/papers/10.21105/joss.00978, doi:10.21105/joss.00978.

Salon, D., 2015. Heterogeneity in the relationship between the built environment and driving: Focus on neighborhood type and travel purpose. Research in Transportation Economics 52, 34–45. URL: https://linkinghub.elsevier.com/retrieve/pii/S0739885915000529, doi:10.1016/j.retrec.2015.10.008.

Salon, D., Boarnet, M.G., Handy, S., Spears, S., Tal, G., 2012. How do local actions affect VMT? A critical review of the empirical evidence. Transportation Research Part D: Transport and Environment 17, 495–508. URL: https://linkinghub.elsevier.com/retrieve/pii/S136192091200051X, doi:10.1016/j.trd.2012.05.006.

Schlich, R., & Axhausen, K. W. (2003). Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation*, *30*(1), 13–36. https://doi.org/10.1023/A:1021230507071

Schwarz, G., 1978. Estimating the dimension of a model. The annals of statistics , 461–464Publisher: JSTOR.

Sclove, S.L., 1987. Application of model-selection criteria to some problems in multivariate analysis. Psychometrika 52, 333–343. Publisher: Springer.

Shahabi, C., Kolahdouzan, M. R., & Sharifzadeh, M. (2003). A Road Network Embedding Technique for K-Nearest Neighbor Search in Moving Object Databases. *GeoInformatica*, *7*(3), 255–273. https://doi.org/10.1023/A:1025153016110

Sidharthan, R., Bhat, C. R., Pendyala, R. M., & Goulias, K. G. (2011). Model for Children's School Travel Mode Choice: Accounting for Effects of Spatial and Social Interaction. *Transportation Research Record*, *2213*(1), 78–86. https://doi.org/10.3141/2213-11

Straatemeier, T., & Bertolini, L. (2020). How can planning for accessibility lead to more integrated transport and land-use strategies? Two examples from the Netherlands. *European Planning Studies*, *28*(9), 1713–1734. https://doi.org/10.1080/09654313.2019.1612326

Su, Q., 2012. A quantile regression analysis of the rebound effect: Evidence from the 2009 National Household Transportation Survey in the United States. Energy Policy 45, 368–377. URL: https:// linkinghub.elsevier.com/retrieve/pii/S0301421512001620, doi:10. 1016/j.enpol.2012.02.045.

Su, R., McBride, E. C., & Goulias, K. G. (2020). Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis. *Transportation Research Part C: Emerging Technologies*, *120*, 102796. https://doi.org/10.1016/j.trc.2020.102796

Su, R., McBride, E. C., & Goulias, K. G. (2021a). Unveiling daily activity pattern differences between telecommuters and commuters using human mobility motifs and sequence analysis. *Transportation Research Part A: Policy and Practice*, *147*, 106–132. https://doi.org/10.1016/j.tra.2021.03.002

Su, R., Xiao, J., McBride, E. C., & Goulias, K. G. (2021b). Understanding senior's daily mobility patterns in California using human mobility motifs. *Journal of Transport Geography*, *94*, 103117. https://doi.org/10.1016/j.jtrangeo.2021.103117

The City of Santa Rosa Transportation and Public Works Department, 2020. Vehicle miles traveled (VMT) Guidelines. Technical Report. URL: https://srcity.org/DocumentCenter/View/28508/ Vehicle-Miles-Traveled-Final-Draft-Guidelines-6520

Turner, J., & Grieco, M. (2000). Gender and Time Poverty: The Neglected Social Policy Implications of Gendered Time, Transport and Travel. *Time & Society*, *9*(1), 129–136. https://doi.org/10.1177/0961463X00009001007

Van, der L. H., Haaijer, R., & Willigers, J. (2019). 10—The impact of various forms of flexible working on mobility and congestion estimated empirically. In P. Coppola & D. Esztergár-Kiss (Eds.), *Autonomous Vehicles and Future Mobility* (pp. 125–139). Elsevier. https://doi.org/10.1016/B978-0-12-817696-2.00010-X

Vich, G., Marquet, O., & Miralles-Guasch, C. (2019). Green streetscape and walking: Exploring active mobility patterns in dense and compact cities. *Journal of Transport & Health*, *12*, 50–59. https://doi.org/10.1016/j.jth.2018.11.003

Vickery, C. (1977). The Time-Poor: A New Look at Poverty. *The Journal of Human Resources*, *12*(1), 27–48. https://doi.org/10.2307/145597

Wilson, C. (1998a). Analysis of Travel Behavior Using Sequence Alignment Methods. *Transportation Research Record*, *1645*(1), 52–59. https://doi.org/10.3141/1645-07

Wilson, W. C. (1998b). Activity Pattern Analysis by Means of Sequence-Alignment Methods. *Environment and Planning A: Economy and Space*, *30*(6), 1017–1038. https://doi.org/10.1068/a301017

Xianyu, J., Rasouli, S., & Timmermans, H. (2017). Analysis of variability in multi-day GPS imputed activity-travel diaries using multi-dimensional sequence alignment and panel effects regression models. *Transportation*, *44*(3), 533–553. https://doi.org/10.1007/s11116-015-9666-2

Xiao, J., Su, R., McBride, E. C., & Goulias, K. G. (2020). Exploring the correlations between spatiotemporal daily activity-travel patterns and stated interest and perception of risk with self-driving cars. *AGILE: GIScience Series*, *1*, 1–15. https://doi.org/10.5194/agile-giss-1-22-2020

Yoon, S., & Goulias, K. (2010). Impact of time-space prism accessibility on time use behavior and its propagation through intra-household interaction. *Transportation Letters*, *2*(4), 245-260.

Yuan, Y., & Raubal, M. (2012). Extracting Dynamic Urban Mobility Patterns from Mobile Phone Data. In N. Xiao, M.-P. Kwan, M. F. Goodchild, & S. Shekhar (Eds.), *Geographic Information Science* (pp. 354–367). Springer. https://doi.org/10.1007/978-3-642-33024-7_26

Zhang, W., & Thill, J.-C. (2017). Detecting and visualizing cohesive activity-travel patterns: A network analysis approach. *Computers, Environment and Urban Systems*, *66*, 117–129. https://doi.org/10.1016/j.compenvurbsys.2017.08.004

Data Management Plan

Products of Research

No new data were collected in this study. The California Household Travel Survey (CHTS) was used and is available at the Transportation Secure Data Center of the National Renewable Energy Laboratory (NREL) (https://www.nrel.gov/transportation/secure-transportation-data/index.html). The National Household Travel Survey data (NHTS) California component was also used and is available at the Transportation Secure Data Center of NREL. Also, the CDC SVI data are available in different formats at the https://www.atsdr.cdc.gov/placeandhealth/svi/index.html.

Data Format and Content

Data format and the contents of each file are available at the Transportation Secure Data Center of the National Renewable Energy Laboratory (https://www.nrel.gov/transportation/secure-transportation-data/index.html).

Data Access and Sharing

The general public can access the data through the Transportation Secure Data Center of the National Renewable Energy Laboratory (https://www.nrel.gov/transportation/secure-transportation-data/index.html) and CDC (https://www.atsdr.cdc.gov/placeandhealth/svi/index.html).

The only database used here that is not available are the NETS database that is a product of Business Establishment data collected by Dunn&Bradstreet and it is proprietary subject to a data subscription.

Reuse and Redistribution

There are no restrictions on how the data can be reused and redistributed by the general public subject to the permissions mentioned on the website of NREL and CDC.