Vertical Equity Statewide Pilot, Data Inventory, and Guidelines for Performance Based Planning (VESPI)

Adam W. Davis,
Elizabeth C. McBride,
and
Konstadinos G. Goulias

Final Report
Contract Through: USC-PSR

Submitted to Pacific Southwest
Region 9 University Transportation Center

January 6, 2018
GeoTrans Laboratory, UCSB
Santa Barbara, CA

Table of Contents

Executive Summary

In this project with tile "Vertical Equity Statewide Pilot, Data Inventory, and Guidelines for Performance Based Planning" (dubbed VESPI) we develop a method for vertical equity analysis. Vertical equity is the analysis of disadvantages groups of different incomes and other sociodemographic characteristics experience from the land use-transportation system. We first create a data inventory for the indicators needed to satisfy performance-based planning in California. In parallel, using detailed in space and time databases in the GeoTrans laboratory we create a first pilot geo-computation of equity indicators covering the entire State at fine spatial detail. We compare data available and this pilot to the literature on gentrification, equity analysis, and access to opportunities. We also review the latest travel behavior advances to identify research gaps in this research. This is done through informal discussion with experts, a presentation at an international travel behavior conference (in Santa Barbara, July 2018) together with research gap identification from workshops in the same conference (IATBR2018.org). Input from the different expert sources is then used to finalize the geo-computation of equity. The project ends with findings and next steps.

In terms of substantive policy findings, we show that timing of activity participation and travel times to work are very different among jobs of different industries and these are heavily influenced by the type of centers surrounding them. This in turn shows we need to introduce a fine grade definition of jobs and their distribution in space to gain insights about the impact of policies that are at the intersection of land use and transportation. For example, employer-based demand management strategies (e.g., staggered work hours, telecommuting) will have different impacts on different job types depending on the opportunities offered around home and job sites. This in turn also implies the mix of job types at a location will determine the impact of these policies in space. Recording the number and type of jobs in different locations will then allow us to assess the impact of policies with higher precision and accuracy.

1. Brief Literature Review

Travel behavior researchers and urban geographers often identify accessibility as a major factor in people's movement through space and the attractiveness of destinations. Originally defined as "the potential of opportunities for interaction," (*Hansen, 1959*), measures of accessibility are generally intended to capture the density and diversity of potential opportunities for people's activities, measurable as a continuous field variable over space. However most common measures of accessibility are intended to characterize the overall patterns of opportunity density and the built environment over a region, rather than providing information about the specific places in which people pursue their activities.

The majority of accessibility indicators either consider explicitly an "origin" or anchor point(s) like home and work locations around which availability and reach of opportunities are measured (Handy, 1993, *Levinson, 1998, Levinson et al., 2017*). The range of accessibility people experience throughout the day, either at specific destinations or within their activity space, is linked to social interaction, activity scheduling, and task allocation within the household (*Patterson and Farber, 2015, Shliselberg, 2015, Yoon and Goulias, 2010, Lee et al., 2016*). Diversity-centered accessibility has been measured by variables corresponding to the density of broad land use categories (*Cervero and Kockelman, 1997*) or using information entropy measures (*Davis, 2015, DeAbreu et al., 2006)* . Both methods have shown significant relationships with travel behavior, but they depend heavily on the classification scheme used.

One well-established formulation of accessibility is opportunity-based accessibility, which counts the number of potential destinations reachable within a certain period of time or by traveling a certain distance (*Paez and Scott, 2004*). These are further enhanced by accounting for congestion and the expected opening and closing hours of businesses (*Chen et al., 2011*). This is a valuable measure of the sorts of opportunities that people experience in a day but is less useful as a measure of the specific opportunities available in one place. Proximity to opportunities for complementary activities (like dining and entertainment) seems likely to affect destination choices, but a person would not, for instance, choose to go shopping in a residential area just because it was located midway between two major shopping centers that counted both of them in its accessibility. Opportunity-based accessibility is not a good measure of destination-level attractiveness. Modeling specific opportunity locations (such as stores) individually is also infeasible for several reasons including inability of surveys to capture many destinations (i.e., there are many more business establishments in a region than the locations visited by respondents), often geocoding between surveys and available databases of business establishments are mismatched, and semantic mismatching between activity type and business establishment type. Some sort of aggregation of business establishments is needed in a way that destination attractiveness is maintained. Questions remain about how best to perform the spatial aggregation of opportunity locations.

In travel demand modeling, traffic analysis zones (TAZs) are the typical spatial unit of analysis because the US Census and American Community Survey provide TAZ-level demographic and employment data, but a variety of analytical issues arise from their use. Páez and Scott (*2004*) point out two major issues with TAZs that relate to the Modifiable Areal Unit Problem (MAUP): the scale effect, by which the same analysis can lead to different conclusions depending on the resolution of the spatial units; and the zoning effect, by which different spatial partitioning leads to a wide range of possible analysis outcomes. This casts doubts about models that incorporate zone-level spatial relationships. Solutions to the problem of developing the “right” zoning system have been proposed that attempt to minimize some of the negative impacts of spatial aggregation, including some that focus on capturing travel within a single zone using a variety of scales (*Moeckel and Donnelly, 2015, Viegas et al., 2009*).

Spatial heterogeneity presents another set of issues, since the relationship between a location and the behavior of a person depends inherently on the location for reasons that are not recorded in the data, and this difference can change over the course of a day. Bhat and Zhao (*2002*) show the impact of neglecting spatial heterogeneity in the context of stop-making decisions by households, but their proposed solution uses TAZs as the spatial units of analysis. The degree to which the attractiveness of different places varies over the course of the day has been less thoroughly addressed, but it is clearly an issue. Google Maps now displays plots showing the relative popularity at different days and times for individual destinations, and ratings sites like Yelp also indicates what days and times individual bars and restaurants are best to visit. Some work has attempted to capture time of day signatures of major facilities and events (*Paul et al., 2014*), but work has been limited in travel behavior. Neglecting the variability of place attractiveness is a major issue for activity-based travel models that simulate activity participation by time of day and day of the week.

Although place-based analysis has been limited, travel behavior research has found a variety of ways to understand the mix of travel and activities people do in a day. Measures used have included the amount of time spent sleeping, eating, working, and socializing; the number, modes, and lengths of trips made; and the number of people interacted with. Lee et al. investigated the relationships among these measures by developing a three-way latent class clustering model of daily schedules (*Lee et al., 2017*) and other travel behavior researchers have noted the importance of understanding daily sequences of activities (*Bhat et al., 2013*). Geographers have focused more on the relationship between time and place; notably McKenzie et al. (*2015*) identified interesting differences between the temporal signatures of activities of locations using social media data.

The main motivation for our research is to understand how people perform different activities at different times in different types of places. We focus here on places that provide opportunities for activities like shopping, dining, socializing, entertainment, and exercise. Using a database that contains every business establishment in the study region, we identify activity anchors that we call commercial centers and categorize them by size, prominence, and diversity. Then, we demonstrate their importance as a significant control on time use and scheduling. After this we study the impact of these centers on commute time while controlling and demonstrating the effect on travel time of a variety of other factors.

The key questions we answer are:

- How can spatial aggregation be used to identify and distinguish commercial centers from a comprehensive dataset of relevant businesses?
- Do different types of business centers attract users at different times for similar activities and for which activities are these differences most significant?
- What is the relationship between these new indicators of access and destination attractiveness and commute time?

2. Data and Data Processing

This project draws data from two different sources: a comprehensive record of business establishments that we use as opportunities for shopping, dining, socializing, and other activities, and a place-based travel and activity diary that we use as a record of activities performed by California residents. Spatial coordinates in both datasets were provided in latitude and longitude referenced and were projected into the same coordinate system. Detailed place-based activity diaries provide valuable data for the analysis of what makes places attractive, but available activity metrics do not match this level of spatial detail. Trips are made to specific businesses but using individual stores as the level of analysis may not be feasible due both to data precision issues and a need to consider the spatial context and groups of potentially complementary opportunities. Commercial neighborhood and business centers present a high density of opportunities for a wide range of activities, but there are not established methods for identifying and categorizing these areas in order to investigate their differences in terms of human behavior.

The California subset of National Establishments Time Series dataset (NETS) contains a record of all business establishments (e.g. individual stores) in California from 1990 to 2013 (*Walls and Associates, 2017*). This dataset is produced from Dun & Bradstreet business establishment data. For this project, we use customer-facing businesses in the retail, food service, and entertainment categories (North American Industry Classification System 2-digit codes 44, 45, 71, and 72) located in California in 2012 with at least three employees, for a total of 193,820 businesses. Businesses with 1 or 2 employees were excluded because many of them appeared to be home addresses rather than storefronts; for instance, a home-based business that sells products online is retail, but it does not provide a place for people to physically shop.

The California Household Travel Survey (CHTS) contains demographic and travel-related information for 108,778 people in 42,431 California households in 2012-13 (*NUSTATS, 2013*). The activity-travel diary provides locations, descriptions, travel modes, and timings for all the activities performed in a single assigned day in the life of all the people in the survey. The households selected for the survey were spatially stratified by county in order to ensure the collection of sufficient data to model the behavior of rural travelers, but sampling probabilities for the Southern California residents used in our analysis are fairly consistent across counties. For the schedule comparison analysis, only out-of-home activities were considered.

To examine the correlation between time of day activity-travel behavior dynamics and the pattern recognition derived commercial centers in our method we selected Los Angeles, CA. This is motivated by the substantial diversity of the LA population, generally high levels of accessibility throughout the region, and large number of observations in CHTS.

3. Pattern Recognition Methods and Parameter Selection

Census spatial units (such as block groups and tracts) are designed to count people, so their size, shape, and boundaries may not be ideal for delineating activity clusters and commercial centers. This is especially problematic since most spatial analyses of polygon data essentially assume that everything going on in an area occurs at its centroid. Boundaries of these units are set based on “visible and identifiable features,” like arterial roads (*US Census Bureau, 2010)*, which often also house many of the businesses in an area. As a result, commercial centers built around major roads and intersections, often wind up being split across multiple units.

Figure 3.1 shows three potential cases of business aggregation to tracts from an area in central Los Angeles. The first example is the best case for small commercial centers, because all the businesses are located near the centroid of a single tract. In example 2, the businesses in an area cluster along a street dividing two tracts. In example 3, a set of businesses clusters along the intersection of two major roads that divide four tracts, which means that a tract-based spatial aggregation would effectively separate those businesses by as much as half a kilometer in an area with high population density, and much more in an area with larger tracts. The area shown by this figure was chosen because it contained a wide range of outcomes, but in the study area as a whole, far more small and medium sized clusters of businesses match cases 2 and 3 than case 1. The bubbles on the map correspond to the search radius for destinations in each of the clusters that result from our spatial clustering result.

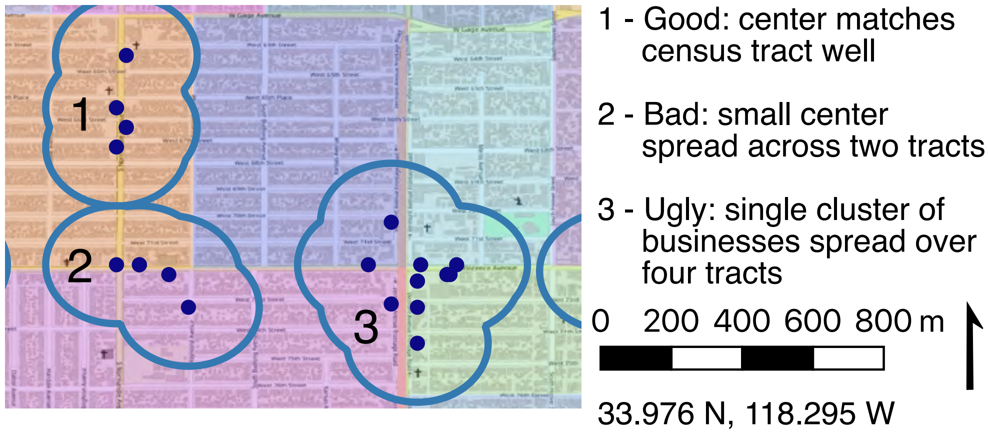

Figure 3.1 Three potential outcomes when businesses concentrated along roadways get mapped to census units. Census tracts shown in shaded colors. Final cluster search radius extents are shown in blue. Example 3 clusters along Vermont Ave, a major arterial in central Los Angeles.

For this analysis, we use Density-based spatial clustering of applications with noise (DBSCAN) (*Ester et al., 1996*), a well-established spatial clustering algorithm to identify high density clusters of Southern California retail, food service, and entertainment business locations from NETS. DBSCAN is a deterministic spatial clustering method that identifies sets of densely packed points but does not assign points in low-density areas to any cluster. DBSCAN and similar methods are used for a range of data clustering purposes including identifying distinct points of interest from large numbers of social media posts (*Maddimsetty, 2018, Orenstein et al., 2014*). A similar application of this method extracted regions of interest from large collections of geotagged photos and identified representative photos for each region of interest (*Hu et al., 2015*).

To extract clusters, DBSCAN requires a set of data points, a distance function, and two parameters: the fewest total points a point must neighbor to count as a core point (minPts), and the maximum distance between two points for them to count as neighbors ($\boldsymbol{\varepsilon}$). For this analysis, we use an efficient implementation of the DBSCAN algorithm in the R package *dbscan* (*Hashler et al., 2018*).

DBSCAN identifies clusters using a three-step process:

1. For each observation *p*, count the total points with a distance of less than $\boldsymbol{\varepsilon}$ from *p* (including *p*). Flag all *p* that have at least minPts neighbors as *core points*.
2. Identify all *core points* within $\boldsymbol{\varepsilon}$ of each other as neighbors and extend transitive neighbor-status to all the *core points* neighboring each of these, so that all core points are neighbors if they are within $\boldsymbol{\varepsilon}$ or can be connected by a string of other neighboring core points.
3. Each group of neighboring *core points* is a single cluster. All points that do not meet the minPts threshold but are neighbors of one or more *core points* count as *edge points* for all clusters containing *core points* they neighbor. All points that neither meet the minPts threshold nor are adjacent to any points that do are classified as noise.

In order to cluster points using DBSCAN, we must choose values for its two parameters: MinPts and $\boldsymbol{\varepsilon}$ (maximum distance between neighbors). Substantial efforts have been made to automate the selection of parameters for density-based clustering methods, but these parameters should still be chosen with consideration for the phenomenon being clustered (*Karami and Johansson, 2014*). The minimum points parameter controls the minimum acceptable size for a cluster and should generally be greater than two times the number of dimensions, which in this case suggests a minimum acceptable cluster size of 5 and $\boldsymbol{\varepsilon}$ should be at a scale that is meaningful for the data clustered (*Schubert et al., 2017*), which is often investigated using a distance to nearest-neighbor distance plot. DBSCAN runs quickly on the business points we used, so we test parameters over a range of $\boldsymbol{\varepsilon}$, every 50m from 100 to 1000 meters, and a range of minPts, every 5 from 5 to 100 points. Figure 3.2 shows the percent of business locations in a cluster for each of these 380 sets of parameters.

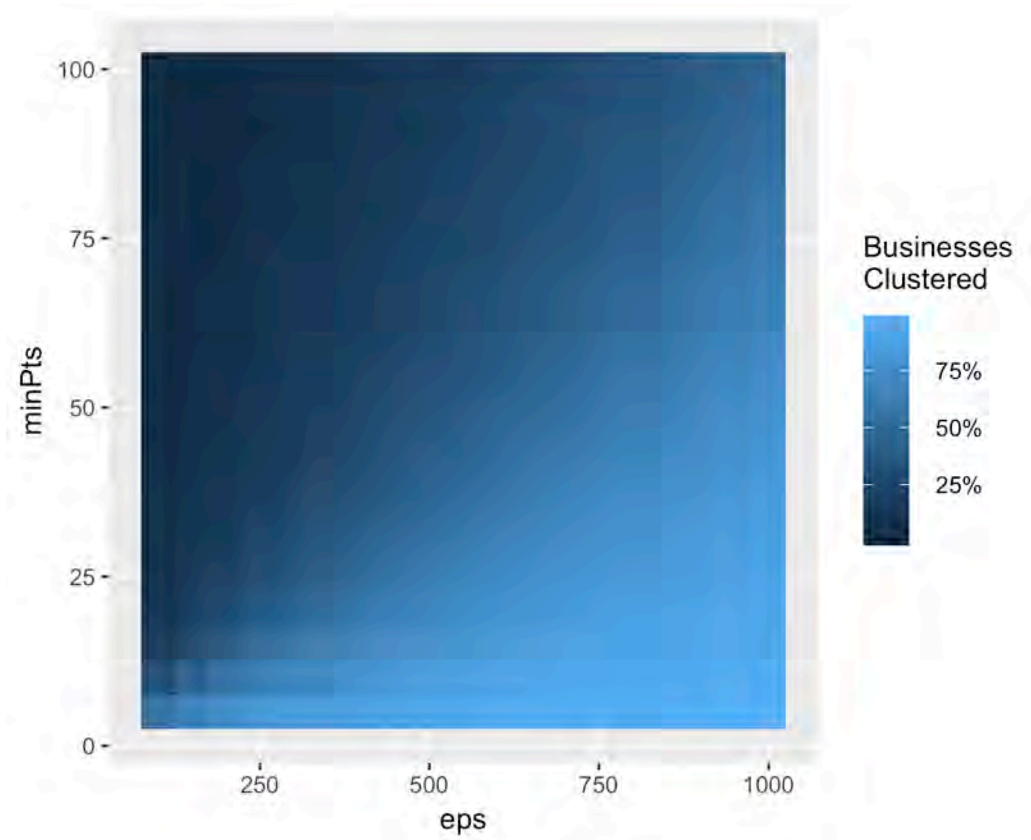

Figure 3.2 Percent of Businesses Classified using DBSCAN with 380 sets of parameters. Note: Small values of minPts paired with high values of epsilon include most relevant businesses in centers but also combine many centers across large areas.

Mapping the clusters produced by a few of these tests indicated that pairing a low $\boldsymbol{\varepsilon}$ with a high minPts (upper left in Figure 3.2) only counted the largest urban cores. Increasing the $\boldsymbol{\varepsilon}$ while maintaining a high value of minPts (upper right) expanded these major downtown clusters laterally and added a few other areas with enough businesses over a larger area. Lower values of minPts allowed for the detection of far more clusters in a much larger share of the state. High $\boldsymbol{\varepsilon}$ parameters paired with the lowest values of minPts (lower right) classified almost all developed parts of the state into a cluster and joined all nearby clusters so that, for instance, most of Los Angeles County was covered by a single cluster.

Since the goal of this research is to identify areas of densely packed opportunities for shopping, dining, entertainment, and socializing, we use classified CHTS activity locations as secondary information to aid in selecting a final clustering. For each set of parameters, we assigned every activity location to the cluster of the nearest business in a cluster as long as the distance was less than the $\boldsymbol{\varepsilon}$ used for that clustering. Essentially, this means that CHTS points were assigned to clusters that they would have been part of if treated as a NETS point. We then compared the 380 clustering results in terms of the share of shopping activities and home locations falling within a cluster (Figure 3.3). We want to maximize the inclusion of observed shopping destinations while minimizing the share of home locations in clusters, since although some people live downtown, we do not want the commercial centers to cover areas that are primarily residential.

Figure 3.3 Relative shares of CHTS shopping and residential locations falling in clusters of NETS businesses at various sets of parameters. The lowest value of minPts forms the frontier plot

As shown in Figure 3.2, the lowest value of minPts performed best in this comparison, and the lower values of $\boldsymbol{\varepsilon}$ captured far fewer home locations than did the high values. For the final analysis, we chose to use minPts=5 and $\boldsymbol{\varepsilon}$ =200m because this clustering placed most businesses into clusters, generated relatively few unreasonably large clusters, and captured nearly 80% of shopping trips and less than 20% of home locations.

It is useful to explore potential options within a range that makes sense for the data. Since our goal is to identify major destinations for non-mandatory activities, we use locations at which people pursued these activities as a piece of secondary information to compare the clustering results. For business location data, a low value minPoints threshold seems to be ideal (especially since a development with 5 stores and restaurants would count as a local center). Lower neighbor distance limits make it possible to distinguish between clusters, but very low distances capture a smaller share of relevant activity locations, in part due to a mismatch between geocoding results.

4. Business Center Classification and Activity Scheduling Results

In this section we compare activity participation and scheduling across different commercial centers classified into three categories: large, small mixed use, and small retail only. To attempt to control for the overall density and transportation infrastructure, this analysis includes only centers at least partly located in Los Angeles, Orange, Riverside, and San Bernardino counties. Commercial center classification could be done using latent class clustering or a similar model, but to simplify the analysis presented here, we create a two-step decision tree based on density and diversity of opportunities in centers. This scheme is intended to sort a roughly equal number of businesses (though different numbers of clusters) into each category. Clusters with at least 90 relevant businesses are classified as major centers. Clusters with fewer than 90 businesses but more than 40% of their businesses in categories other than retail are classified as mixed-use centers, and the remaining centers are classified as retail-focused small centers.

Figure 4.1 shows the distribution of center classifications in the western portion of the study region, which covers most of Los Angeles County. Most of downtown Los Angeles and the well-established dense corridor through western LA along Wilshire and Santa Monica boulevards (*Giuliano and Small, 1991*), as are some of the other major downtowns in the region. Smaller centers mostly cover major intersections, malls, and "main street" developments.

Table 4.1 contains descriptive statistics of the NETS businesses and CHTS out-of-home destinations and activities located in the resulting center types as well as the centers outside of the four-county Southern California focus region, and areas not clustered into any center. Activity purposes and travel modes show a strong relationship to center type, but day of the week does not. Additionally, people appear willing to travel farther to reach larger and mixed-use centers, whereas retail centers serve mainly the local community.

Commercial Centers of Greater Los Angeles

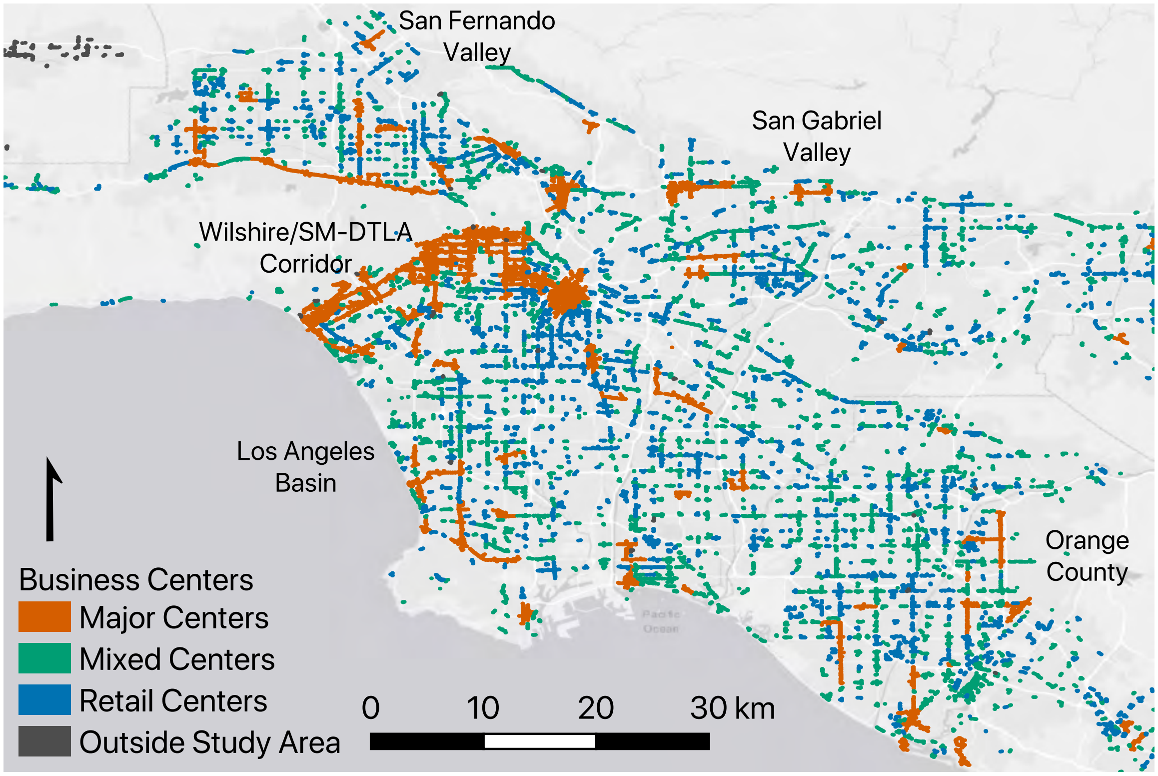

Figure 4.1 Downtown Los Angeles, large suburban downtowns, and major commercial strips are classified as major centers.

Note: Local downtowns with a range of opportunities are classified as mixed centers. Large malls and retail developments on the periphery of major centers are classified as retail centers.

Table 4.1 Summary statistics for commercial center types and non-centers in terms of their business makeup, travel to them, and out of home activities performed in them

		Major Centers	Mixed Centers	Retail Centers	Centers Elsewhere	Non-Centers
Business and Center Data	Total Centers	97	1,567	1,359	3,632	
	Largest Center	4,599	89	89	4,781	
	Median Center	137	10	9	10	
	Smallest Center	90	5	5	3	
	Total Businesses	24,172	25,032	19,964	75,694	48,958
	Retail	59.5%	47.0%	73.6%	58.1%	61.5%
	Food Service and Lodging	30.3%	44.0%	21.7%	35.3%	22.6%
	Arts and Entertainment	10.2%	9.0%	4.7%	6.7%	15.9%
Destinations	Activity Locs. (OOH)	15,217	17,505	12,226	80,238	116,092
	Total People (OOH)	8,287	11,184	8,197	35,312	63,604
	Mean distance home (km)	21.69	24.14	13.17	22.70	32.28
	Median distance home (km)	5.72	4.44	4.01	4.72	5.27
DOW	Monday	14.7%	14.1%	14.3%	11.4%	13.3%
	Tuesday	12.2%	12.3%	12.4%	17.7%	16.2%
	Wednesday	11.7%	12.7%	11.9%	18.0%	15.6%
	Thursday	12.6%	12.0%	11.9%	17.6%	16.6%
	Friday	16.9%	17.0%	17.0%	13.6%	15.0%
	Saturday	17.5%	17.4%	17.6%	11.8%	11.9%
	Sunday	14.4%	14.3%	14.8%	9.9%	11.6%
Mode	Personal Vehicle	66.6%	80.3%	76.8%	74.9%	77.8%
	Rail	2.6%	0.7%	0.8%	2.4%	0.8%
	Bus / Shuttle	8.4%	5.0%	7.2%	4.3%	4.1%
	Bike	1.0%	0.8%	0.7%	1.6%	1.6%
	Walk	19.7%	11.0%	13.2%	15.2%	10.0%
	Other	1.7%	2.2%	1.2%	1.6%	5.8%
Activity Purpose	Work	17.7%	14.6%	14.9%	15.7%	19.4%
	School	1.6%	2.0%	2.3%	1.6%	8.6%
	Personal / Medical	10.6%	13.2%	13.1%	13.5%	8.6%
	Shopping	17.6%	19.7%	23.2%	22.6%	6.1%
	Eating	13.2%	16.7%	10.8%	13.9%	3.3%
	Entertainment	3.3%	2.6%	1.7%	2.2%	3.2%
	Social / Community	6.8%	7.4%	7.5%	6.0%	17.5%
	Other	5.6%	6.6%	5.3%	6.2%	10.8%
	Loop/Dropoff and Mode change	23.5%	17.2%	21.2%	18.3%	22.5%

5. Pilot Case Study – Activity Participation Curves in the Los Angeles Metropolis

In order to compare the activity and time use dynamics of the different types of business centers we identified, we produce activity participation curves for different activities in the various types of centers. We generate these curves with a temporal resolution of 15 minutes for visual clarity and because a very large share of activities in the activity diaries are recorded as having started and ended on an even quarter hour, which suggests increased precision may be inappropriate.

Different statistics can be extracted from the overall process, but for Figures 5.1 to 5.6, we use the share of ongoing activities of a given type at every time point in each type of commercial center normalized by the total number of activities of that type that take place in that commercial center over the course of the day. To calculate this, start by creating a comprehensive list of activities in a certain category (such as *work* or *dining*), and note their location, start, and end times. Classify each activity by the type of center it falls into (if any) and convert the start and end times to the temporal resolution for the analysis. For this paper start times not on an even quarter hour are rounded *down* to the nearest 15 minute mark, and uneven end times are rounded *up*. This ensures that every activity will be counted for at least one time.

The calculation for percent of relevant activities (subscripted i) happening at center type c at time t is as follows: first identify how many total activities of the type take place at any time during the day in t; then for each time point, identify how many of those activities started on or before that time and end after it. The ratio between these quantities is the share of relevant activities active at a given time point in a given center type.

$$Percent\ Active_{ct} = 100 \times \frac{\sum_i \left(inclust(i,c) \times active(i,t)\right)}{\sum_i inclust(i,c)} \qquad \text{(Eq. 5.1)}$$

$$inclust(i,c) = \begin{cases} 1 \mid location_i \in center_c \\ 0 \mid location_i \notin center_c \end{cases}$$

$$active(i,t) = \begin{cases} 1 \mid start_i \geq t \ \cap end_i < t \\ 0 \mid start_i < t \ \cup end_i \geq t \end{cases}$$

To get a sense of the measurement accuracy of this calculation, we use a bootstrapping process to generate 100 new sets of activities by resampling from the original set of activities. The calculation for percent active was then performed for every time center type and time point for each of these bootstrap activity schedules.

Timing Comparisons and Discussion

For each of these plots, the observed pattern is shown as solid line, and the results of 100 bootstrap runs of schedule regeneration are shown in semitransparent lines to provide a sense of the uncertainty of these measures. The vertical axis in these plots shows what share of relevant activities that take place in a particular type of center during the entire day are ongoing at a given time, as shown in Equation 5.1. If one curve is consistently higher than another, the activities it contains must last longer, on average, since they get counted at multiple time points. When differences are reported as significant in the text, but a specific number of runs is not provided, assume the comparison held in at least 95 of the runs (equivalent to a $p<0.05$), unless it is clear that none of the paths overlap, in which it is safe to assume that the comparison is significant at a level equivalent to $p<<0.01$.

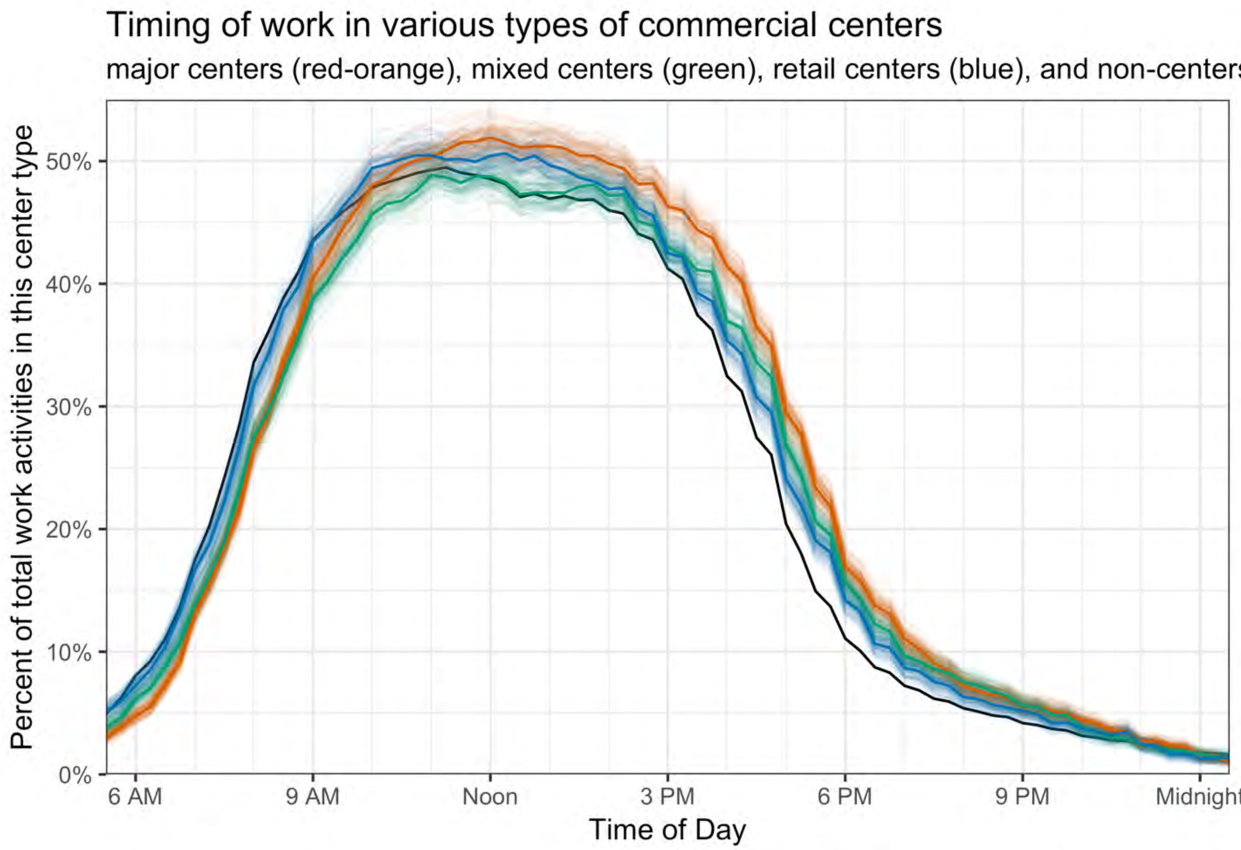

Figure 5.1 Timing of Work

Figure 5.1 shows that work activities follow a relatively consistent pattern across all three center types, but the overall timing of work activities in major centers is shifted roughly half an hour later than in the other center types and roughly 45 minutes later than in non-centers. People work at roughly the same time in all three center types, but the hours are not exactly the same everywhere: the peak of work activity participation appears to start about 30 minutes earlier in retail centers, which are ahead of both other center types from 5:15-9:15 in 90 of the runs and from 6:45-9:00 in all of the runs. More jobs extend into the late afternoon in major centers, which have the highest rate of work activities in at least 95 runs from 14:45-17:45 and at least 90 runs until 19:00. Schooling (not shown) is even more consistent, with a consistent peak running from about 8:00 to about 15:00 in all cluster types. The overall pattern of work taking place later in major centers is even more pronounced when compared to non-centers.

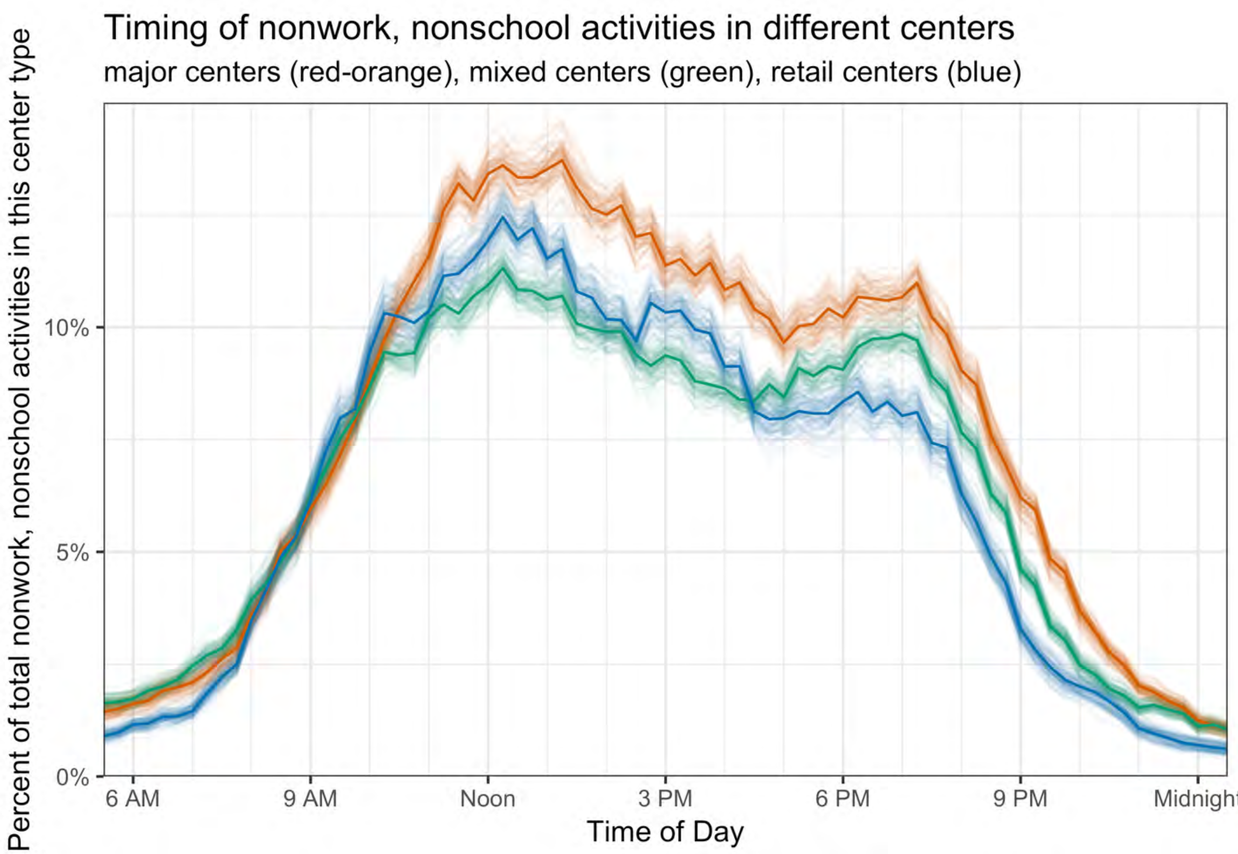

Figure 5.2 Timing of Non-work Activities

The overall timing of work activities is fairly consistent across the center types, but as Figure 5.2 shows, non-work patterns suggest that there are substantial differences in the timing and duration of non-work activities in the different centers. Based on the overall higher curve for major centers, non-work activities there generally last longer. Additionally, these activities persist later into the evening in both major centers and mixed centers, whereas retail centers draw a larger share of their trips in the morning. These differences are likely due in part to the overall differences in mix of activities in different places – namely the increased density of office jobs in major centers may lead to overall higher rates of activity participation, but the time use curves for several activities, including shopping, dining, and entertainment show considerable differences from place to place as well. Instead of comparing all three center types on a single plot, activity-specific plots contain major centers and mixed-use centers, as well as the observed time pattern for all activity locations not located in commercial centers.

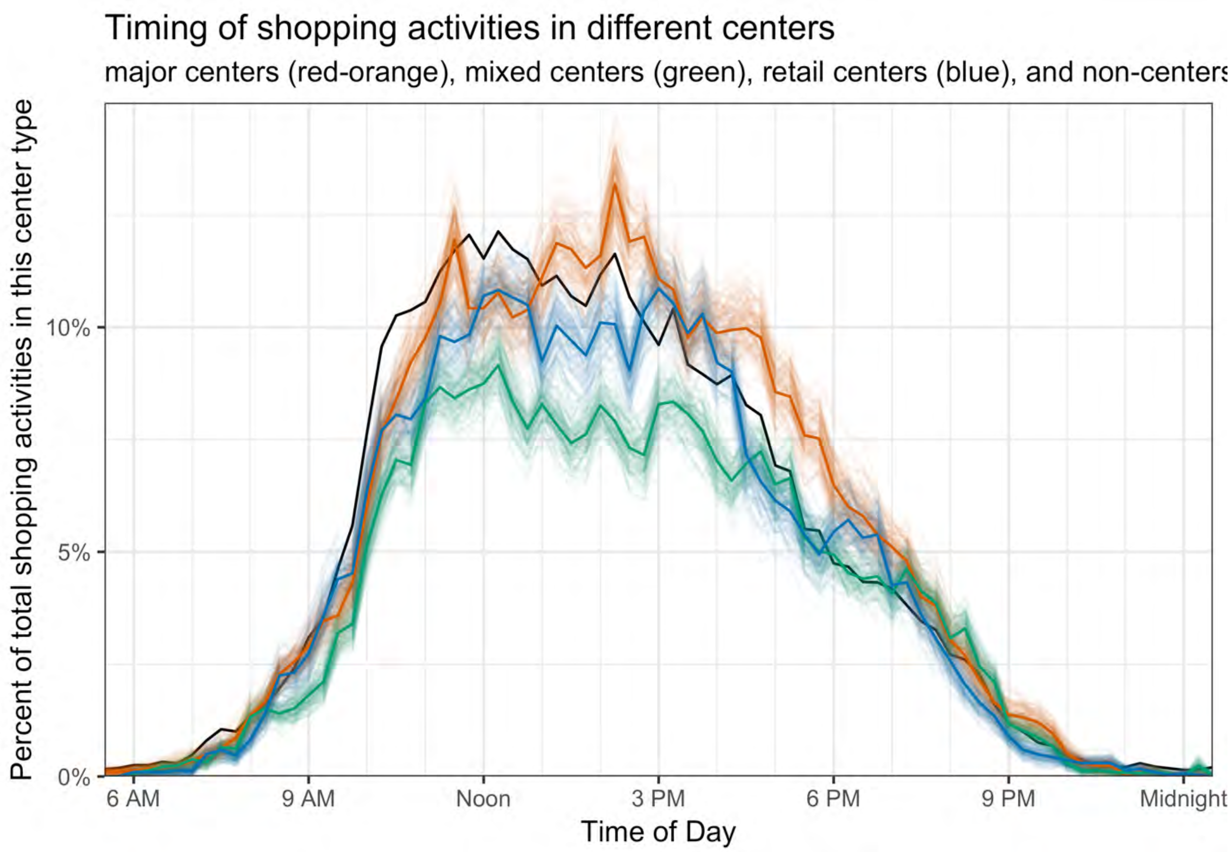

Figure 5.3 Timing of Shopping Activities

Figure 5.3 shows that trips are spread over roughly the same hours in all types of centers but tend to last longer in major centers. People go shopping at roughly the same times of day everywhere, but the gap between activity participation rates in the two center types shown in Figure 5.3 suggests that there is real variation in the duration of these activities from place to place. A greater share of shopping activities is taking place in major centers than in mixed centers at every time point from 9:00 to 18:00 in nearly every simulation (always at least 95; usually in all 100), and major centers lead small retail centers significantly through most of the day as well (particularly 13:00 to 14:45 and 16:30 to 17:45, when it leads in at least 99 simulations at all times). Inspecting the durations of specific shopping activities confirms this result: shopping activities in major centers have a mean duration of 52.4 minutes, nearly 20% longer than in retail centers (44.4 minutes) and 40% longer than in mixed centers 37.2 minutes). This suggests that people either do fundamentally different types of shopping at different center types or enjoy shopping for longer in major downtowns and that using a single model for duration of shopping activities erases important distinctions between places.

The activity pattern for non-centers appears somewhat more similar to that of major centers, but it decreases earlier in the afternoon. Retail opportunities in non-clustered areas may correspond to small corner stores in residential neighborhoods or big box stores that take up enough space to have few neighbors within 200 meters of their geocoded location and thus be counted as noise by DBSCAN. Other clustering algorithms may be used to rectify this result in the future. It is also worth noting that the shopping trips used in this analysis are drawn from two different activity purposes listed in CHTS, which distinguishes between shopping for a major purchase and everyday/routine shopping. These activities do have moderate differences in average duration, but their spatial distribution is relatively even between different center types, and the durations for major shopping are also longer at major centers, which suggests that segregating the two activity types may lead to different findings.

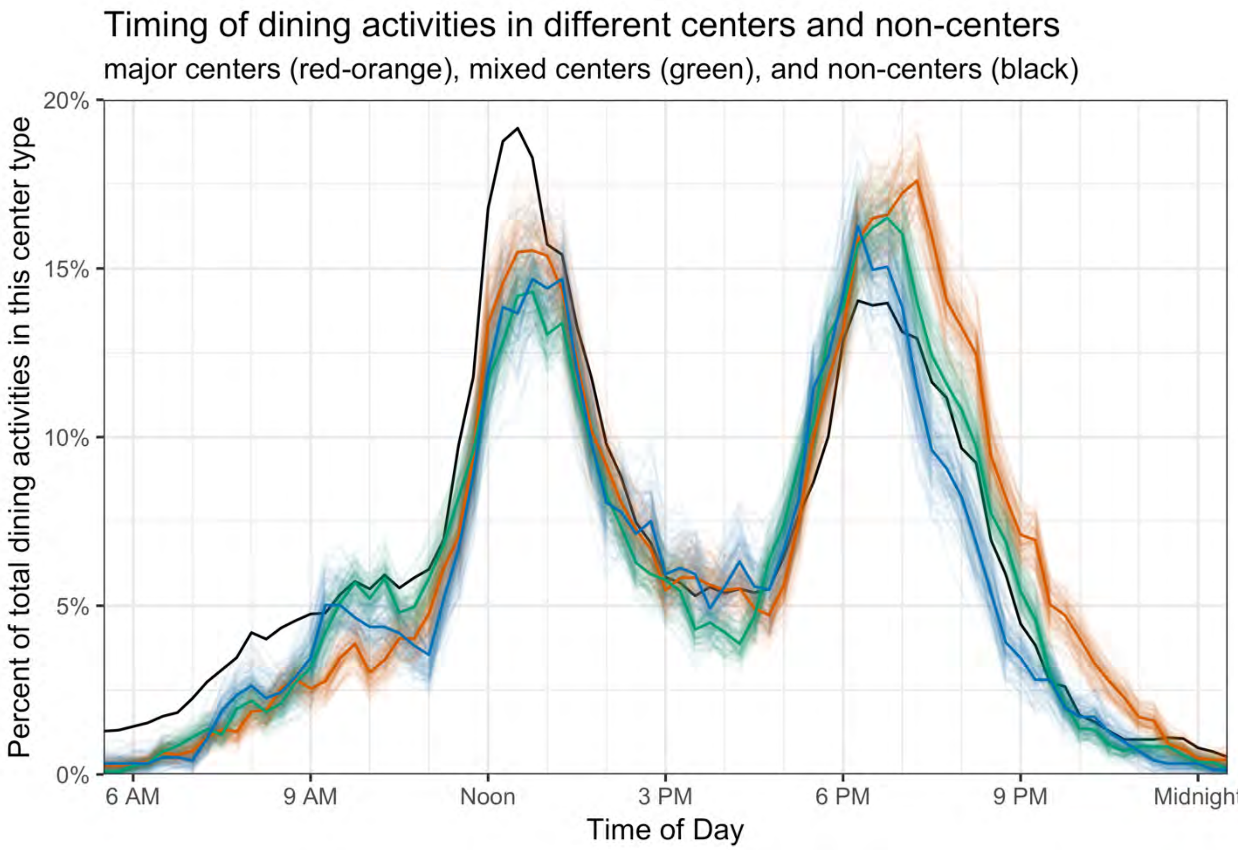

Figure 5.4 Timing of Dining Activities

Unsurprisingly, dining activities are most common around lunch and dinner times across all business center types as well as less dense areas, but there are some notable differences, as shown in Figure 5.4. Despite the various center types' similarities in terms of total activities, eating at restaurants and diners is much less common in the small retail centers than in either the mixed or major centers, as Table 4.1 shows, and as a result the bootstrap shows considerably more variability in these temporal patterns than in the patterns for the centers that present a more diverse range of opportunities. All of the centers diverge from the non-center areas in terms of the relative significance of lunch and dinner to the overall totals, as well as with later starts in the morning. Both categories of smaller centers draw substantial numbers of people for breakfast, whereas the large centers have a much smaller bump at this time (and have the lowest activity participation share in all but one or two runs continuously from 9:15-10:15). Major centers appear to have a slightly larger draw at lunch, but the bootstrap results suggest this may be random.

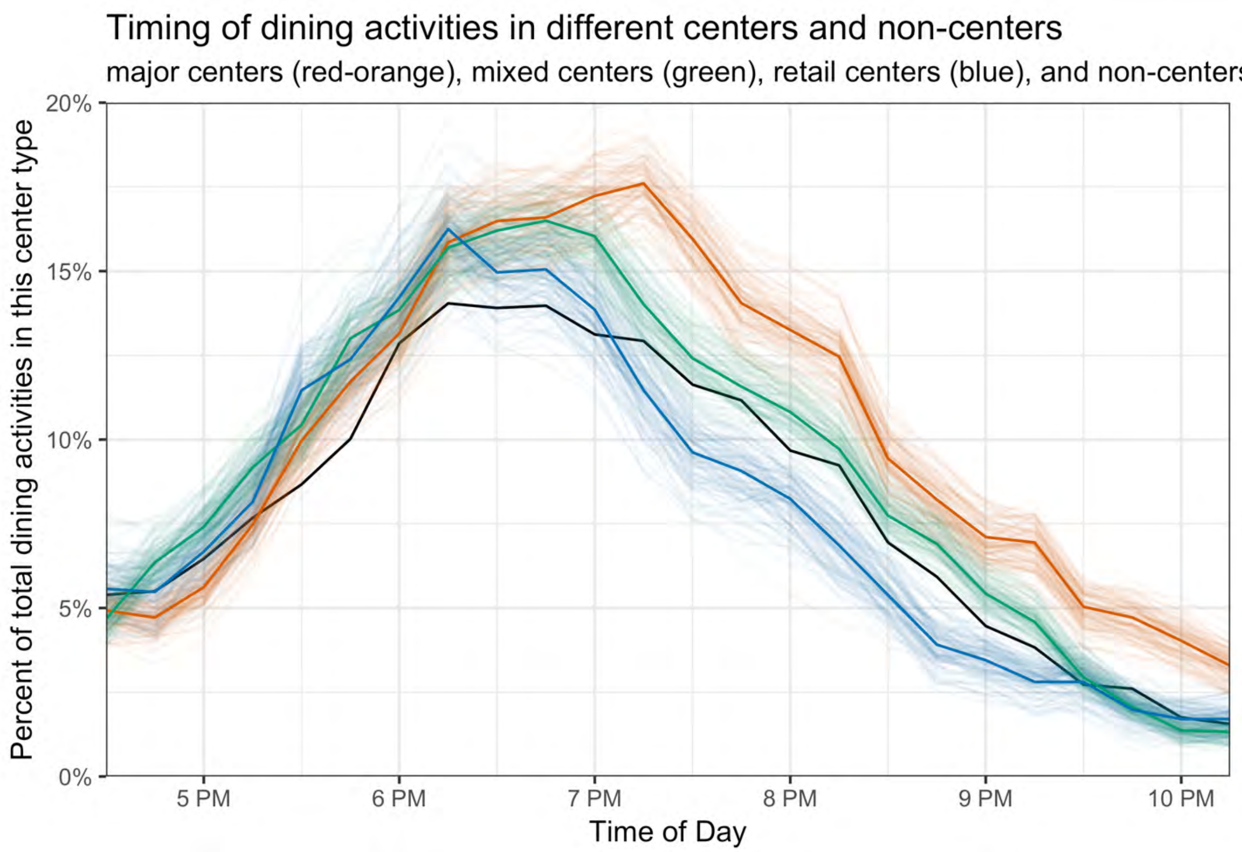

Figure 5.5 Timing of PM Dining activities

Dinner is the main meal for which people go to all center types, unlike non-center areas, but the timing of these events is somewhat different, as shown in Figure 5.5. Both of the smaller center types reach their maximum concentration of eaters between 18:00 and 18:30 pm, whereas the major centers do not till after 19:00 pm. In addition, the number of people getting dinner in major centers begins to decline roughly an hour later than it does in the smaller centers, an effect that is borne out by the bootstrapping: major centers are significantly higher than both other center types for most times from 19:15-23:15, usually in all 100 simulations. In addition to starting later, diners in major centers also spend slightly longer at meals starting between 17:00 and 22:00 (77.7 minutes, on average) than in mixed (71.2 minutes), or retail (66.5 minutes) centers, though these differences are not as substantial as the differences in retail activity duration noted above. Interestingly, despite the later starts, 12.6% (105/843) of people who get a meal at any time after 16:00 in major centers list an entertainment activity at a later destination, whereas the rate is 8.1% (83/1024) for mixed centers. This potential for activity pairing may represent a substantial pull to major downtowns, which present a wider range of entertainment opportunities.

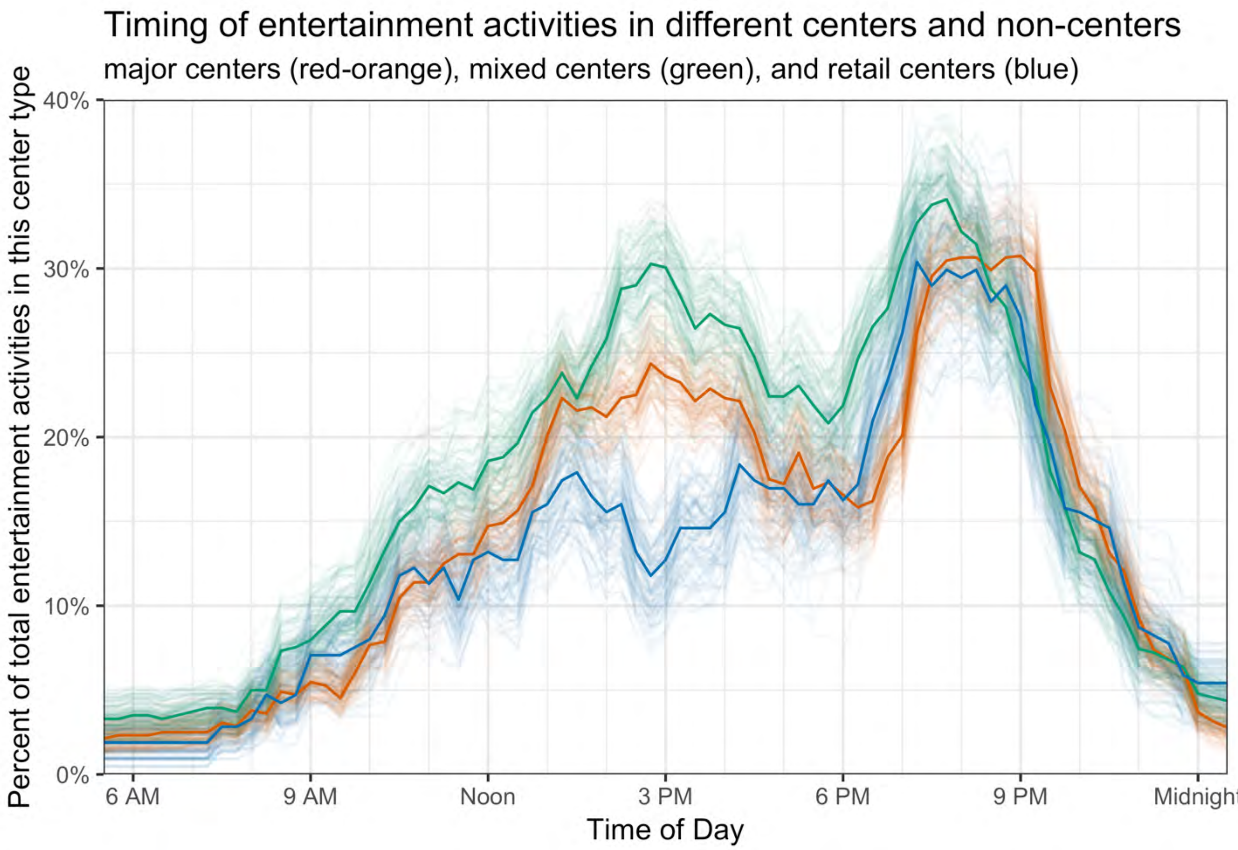

Figure 5.6 Timing of Entertainment Activities

Figure 5.6 shows entertainment activities (e.g., movies, watch sports, live music). This is one of the less common activities in the sample, which results in a somewhat noisier plot than for the other activities. Entertainment is primarily an afternoon and evening activity in all clusters (Figure 5.6), but the mix of afternoon and evening activities is different from place to place. Entertainment activities last similar lengths (about 2.5 hours, on average, with a slightly lower median, suggesting a mix of movies and longer activities) in all three center types, but substantially more of these activities take place in the major centers. There appear to be two main temporal patterns of these activities: early afternoon, which is most common in mixed centers and least common in retail centers, and evening, which begins earlier in mixed clusters than the other two types.

6. Commuting Travel Time and Access to Jobs

Time spent commuting has been tied to lower levels of social capital and happiness as measured by the amount of time people are able to spend in activities like socializing, recreation, and exercise (*Besser, Marcus, & Frumkin, 2008*), most people prefer some amount of commuting to none at all, generally because it provides an opportunity to spend time alone and separate home and work life (*Redmond & Mokhtarian, 2001*). Commute time is driven in large part by the distance between home and work, but people consider many other attributes of specific homes and neighborhoods when choosing a home (*Wachs et al., 1993*). Additionally, the spatial mismatch theory suggests that long-duration commutes often reflect changes to job and residential location that exclude low-income people who live in cities, particularly those who do not have access to personal vehicles and parents (especially mothers) of young children who do cannot take that much time off from their unpaid labor at home (*Blumenberg, 2004; Gobillon, Selod, & Zenou, 2007*).

For this analysis, we investigate commute durations in the California Household Travel Survey for 19,499 people who report working at a fixed location, who went to work on their survey day, and who started and ended the survey day at home. To extract total commute times, we calculate the total travel time for all trips made between the last departure from home before a period at a work location and between the last departure from work to the arrival back at home. Time spent at destinations visited on the way from home to work or work to home are not included, nor is travel time from work-based tours that take place between the first arrival at work and the last departure from work. Descriptive variables include household and personal characteristics like income, gender, commute mode, and job sector, and two sets of land-use characteristics, the employment density of the home Public Use Microdata Areas (PUMA) and an indicator for the presence of each type of commercial center within 5 km of home and work. For this task, commercial centers were identified and classified statewide using the same method that was used for the Southern California case study in the previous section.

People in the study spend an average of 30.7 minutes traveling from home to work and an average of 36.0 minutes on the commute home, likely because more people make at least one stop on the trip home (31.8%) than on the way to work (22.4%). Figure 6.1 shows the sample distribution of commute times, which range from 0 to over 4 hours per day. Table 6.1 provides a breakdown of the value frequencies for variables used in the regression models that follow.

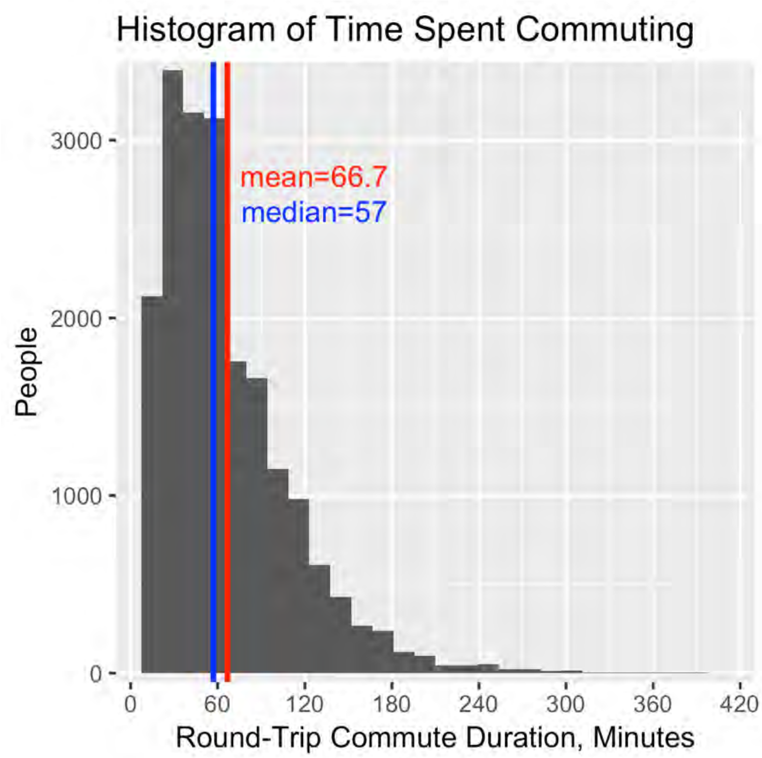

Figure 6.1 More than half the people in the sample spend less than an hour commuting each work day, but some commutes are much longer.

Results for a set of nested linear regression models are presented in Table 6.2; in all four models, coefficients can be interpreted as the number of minutes added to a person's total round-trip commute on the survey day. The first model accounts for only income, commute mode, and gender (with the mode changes and total stops variables included to account for any time out of the way required by incidental destinations added to the commute). The second model adds employment sector and county of residence (not shown on table), which diminish the effect of income. The third and fourth models add county-by-county dummy variables for each mode and income category, respectively, which further addresses the spatial variability of commute conditions in California. In general, the models indicate that commute mode and employment sector are the strongest indicators of commute time.

Although walking and biking are generally slower than other modes, the effort they require means that people are generally willing to use them only for very short commutes. Bus commutes are generally similar in length to car commutes, but their relative speed has a great deal of spatial variability (Figure 6.2), that appears to track population density and public transit quality. San Francisco rail commuters stand out to a similar degree (also roughly 30 minutes faster than car commutes), which may indicate that the city's public transit offerings are considered preferable to cars for all but the longest commutes. Chartered shuttle and vanpool commutes are generally the longest.

People appear to be willing to accept longer commutes for higher paying jobs, although this relationship varies somewhat between different parts of the state (Figure 6.3). The small but consistent difference between the commute times of men and women across these models may reflect differences in home and childcare duties between the parents.

The effect of income on commute time is substantially lower in the three models that include employment sector than in the one that does not, which suggests that the spatial distribution of job sites may play a substantial role in explaining the differences in people's commute times. While people are not willing to commute as far for low-paying jobs, they are especially unwilling to travel for low paying jobs that offer unstable hours (accommodation, food service, entertainment, and retail) or that are often available throughout a region (those and education), rather than being clustered in city centers. Jobs in remote areas (agriculture, mining) tend to require the longest commutes. Construction may be a special case, since workers may report a "work" location that is a job site rather than their regular office.

Table 6.3 shows the results of the final model, which adds a set of land use variables. The general employment density of the area around people's homes appears to have a slight negative bearing on their commute durations, but the six variables associated with the presence/absence of each kind of center within 5km of home/work provide a more complete story and have a stronger effect on commute travel times. People who work in densely developed areas like major city centers tend to have substantially longer commutes (at least 8 minutes per day, depending on whether the other sorts of centers are present as well). Having more opportunities around the home (whether those opportunities are present in a downtown or a mixed center or both) decreases the length of the commute by an even larger amount. These results suggest that people are willing to travel further to get to jobs downtown and also likely add side trips to their commute before or after work in order to access the opportunities available nearby. In contrast, people with a wider range of opportunities near home do not need to link as many trips for shopping, socializing, and dining to their commute.

The models presented here explain a relatively small share of commute travel time (R^2=0.217 in the model that allowed the effect of mode to vary county-by-county and accounts for distance to centers and land use), but the coefficients on dummy variables (all variables except Mode Changes and Extra Stops on Commute) capture differences that are highly significant and often quite substantial. The survey includes only a single day's diary for each person and because a large number of other factors affect commute duration (namely the choice of home and work location, which likely are not chosen with commute time as the primary factor).

Table 6.1 Summary Statistics for Explanatory Variables (unweighted survey data)

Var	Value	Count	Percent
Income	Under $50K	4,060	20.8%
	$50-100K	6,325	32.4%
	$100-150k	4,050	20.8%
	Over $150K	3,641	18.7%
	Don't Know	410	2.1%
	Refuse	1,013	5.2%
Gender	Female	9,270	47.5%
	Male	10,195	52.3%
	Refuse	34	0.2%
Commute Mode	Personal Vehicle	16,631	85.3%
	Air	1	0.0%
	Boat	23	0.1%
	Bus	712	3.7%
	Rail	526	2.7%
	Shuttle/Vanpool	499	2.6%
	Bike	532	2.7%
	Walk	450	2.3%
	Other	121	0.6%
Total Stops	0	11,696	60.0%
	1	3,612	18.5%
	2	1,758	9.0%
	3	753	3.9%
	4	832	4.3%
	5	295	1.5%
	6	258	1.3%
	7 +	295	1.5%
Mode Changes	0	17,652	90.5%
	1	399	2.0%
	2	794	4.1%
	3	122	0.6%
	4	415	2.1%
	5 +	117	0.6%

Var	Value	Count	Percent
Job Sector	Accommodation/Food	861	4.4%
	Retail	1,710	8.8%
	Education	2,995	15.4%
	Real Estate	264	1.4%
	Management	119	0.6%
	Health Car	2,175	11.2%
	Wholesale	172	0.9%
	Finance	878	4.5%
	Mining, etc.	98	0.5%
	Arts, Entertainment, Rec.	564	2.9%
	Manufacturing	1,343	6.9%
	Transportation	620	3.2%
	Professional Services	1,774	9.1%
	Public Administration	2,002	10.3%
	Information	971	5.0%
	Administrative	271	1.4%
	Agriculture	476	2.4%
	Utilities	339	1.7%
	Construction	374	1.9%
	Other Services	786	4.0%
	Other	305	1.6%
	Don't Know	185	2.9%
	Refuse	217	1.1%
PUMA Density (Jobs/km2)		*mean*	998
		median	444
		minimum	1
		max	26,247
Major Center Near Home		10,972	56.3%
Mixed Center Near Home		17,466	89.6%
Retail Center Near Home		17,047	87.4%
Major Center Near Work		12,509	64.2%
Mixed Center Near Work		18,397	94.3%
Retail Center Near Work		18,057	92.6%

Table 6.2 Model Results Comparison, coefficients in minutes of commuting per day

Category	*Variable*	*Mode Income*	*Mode Income Sector*	*Mode+ Income Sector*	*Mode Income+ Sector*
	(Intercept)	49.1	41.1	38.1	38.5
Income (vs <50)	50-100K	3.0	2.2	2.4	2.2
	100-150k	6.9	4.9	5.1	4.7
	Over $150K	7.9	5.0	5.2	5.0
	Don't Know	7.1	4.6	5.0	4.5
	Refuse	8.4	5.8	5.9	5.8
Gender	Male	7.8	6.1	6.2	6.1
	Refuse	2.1	0.7	1.6	0.5
Commute Mode (vs Drive)	Air	85.9	74.6	76.0	75.4
	Boat	0.1	1.1	5.1	1.2
	Bus	-6.7	-4.8	3.1	-4.8
	Rail	-2.0	-1.0	11.3	-1.0
	Shuttle / Vanpool / Paratransit	20.8	18.6	21.0	18.8
	Bike	-15.5	-12.2	-11.2	-12.3
	Walk	-23.7	-19.1	-17.7	-19.3
	Other	-3.9	-4.1	-1.1	-3.8
	Mode Changes	5.3	4.4	4.0	4.5
	Extra Stops on Commute	9.1	9.2	9.4	9.2
Job Sector (vs Accommodation / Food)	Retail		2.4	2.4	2.4
	Education		2.9	2.8	2.8
	Real Estate		4.7	4.7	4.7
	Management		8.1	8.3	8.0
	Health Care		9.3	9.2	9.2
	Wholesale		10.2	9.9	10.2
	Finance		10.8	11.0	10.8
	Mining, etc.		11.1	10.9	10.8
	Arts, Entertainment, Recreation		12.1	11.9	12.0
	Manufacturing		12.6	12.6	12.7
	Transportation		12.7	12.6	12.7
	Professional Services		12.8	12.8	12.8
	Public Administration		13.3	13.3	13.2
	Information		13.6	13.6	13.6
	Administrative		14.3	14.0	13.8
	Agriculture		14.4	14.4	14.4
	Utilities		15.2	15.2	15.3
	Construction		21.0	21.1	21.0
	Other Services		6.1	6.1	6.1
	Other		8.8	9.0	8.8
	Don't Know		15.5	15.3	15.7
	Refuse		11.6	11.8	11.7

Table 6.3 Results for Model accounting for travel mode (allowed to vary by county) and distance to types of business centers, all coefficients in total minutes spent commuting per day

		Est.	Std. Err.
Income (vs <50)	(Intercept)	45.5	2.5
	50-100K	2.1	0.9
	100-150k	4.5	1.0
	Over $150K	4.7	1.1
	Don't Know	4.7	2.2
	Refuse	5.2	1.6
Gender	Male	6.1	0.7
	Refuse	4.2	7.4
Commute Mode (vs Drive)	Air	80.3	43.8
	Boat	6.1	10.8
	Bus	4.9	3.1
	Rail	13.3	4.0
	Shuttle	21.3	2.9
	Bike	-8.7	2.9
	Walk	-15.2	2.9
	Other	-0.6	4.6
	Mode Changes	3.8	0.6
	Other Stops	9.3	0.3

		Est.	Std. Err.
Job Sector (vs Accommodation / Food)	Retail	1.6	1.8
	Education	2.2	1.7
	Real Estate	3.3	3.0
	Management	7.7	4.2
	Health Care	8.4	1.8
	Wholesale	9.4	3.6
	Finance	9.8	2.1
	Mining, etc.	11.1	4.6
	Arts, Entertainment, Recreation	11.0	2.4
	Manufacturing	12.2	1.9
	Transportation	11.7	2.3
	Professional Services	12.2	1.8
	Public Administration	12.2	1.8
	Information	13.1	2.1
	Administrative	14.0	3.0
	Agriculture	12.1	2.5
	Utilities	14.0	2.8
	Construction	20.2	2.7
	Other Services	5.5	2.1
	Other	8.9	2.9
	Don't Know	14.3	3.5
	Refuse	11.1	3.3
	PUMA Density (1000 jobs/km^2)	-0.5	0.2
Centers	Major Center Near Home	-5.6	0.8
	Mixed Center Near Home	-11.9	1.6
	Retail Center Near Home	-7.8	1.5
	Major Center Near Work	8.3	0.8
	Mixed Center Near Work	2.0	2.0
	Retail Center Near Work	5.6	1.7

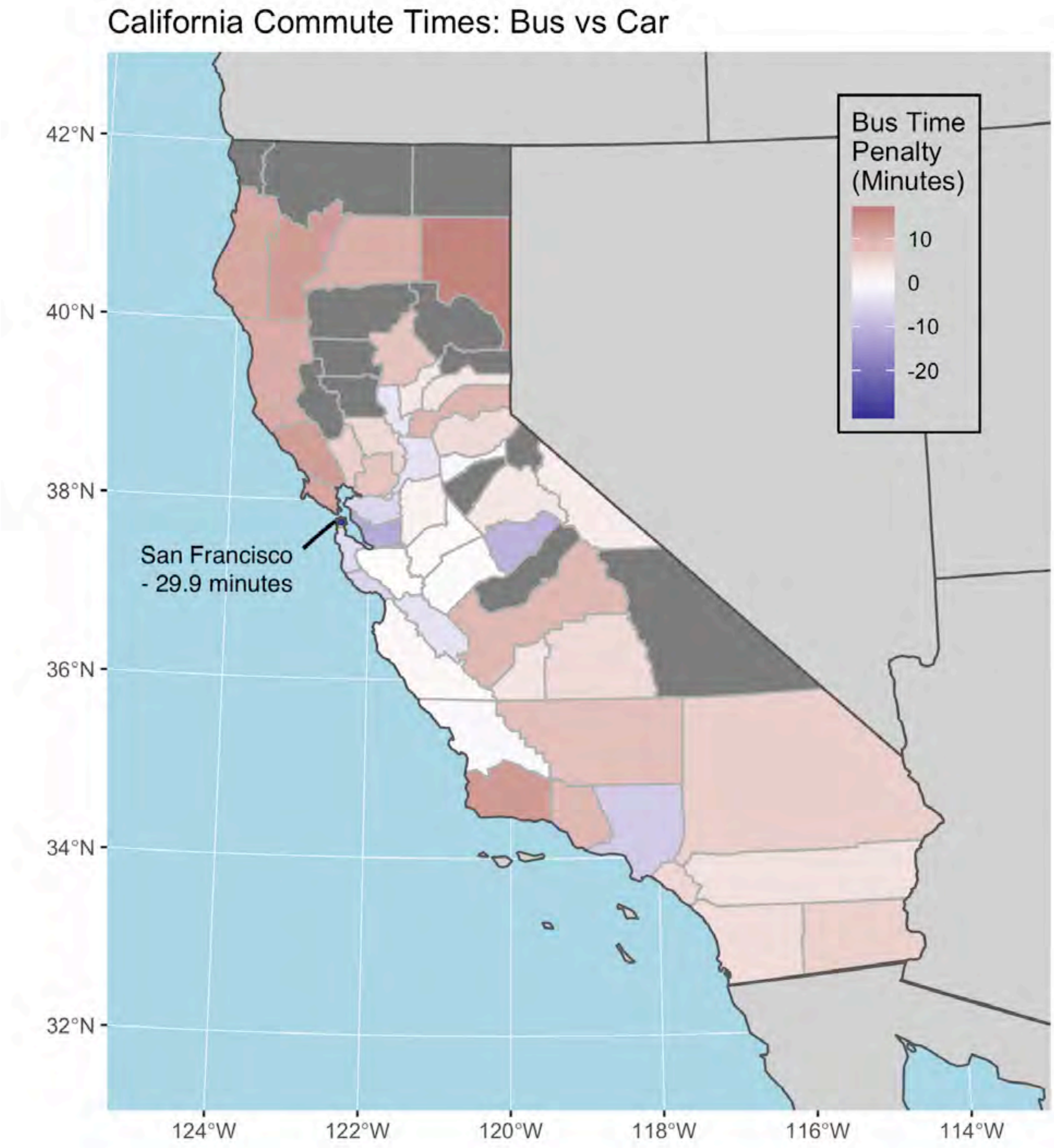

Figure 6.2 Statewide, bus commutes take slightly longer than car commutes, but the effect is reversed in in densely developed areas with relatively high-quality public transit, like Los Angeles, the East Bay counties, and especially San Francisco.

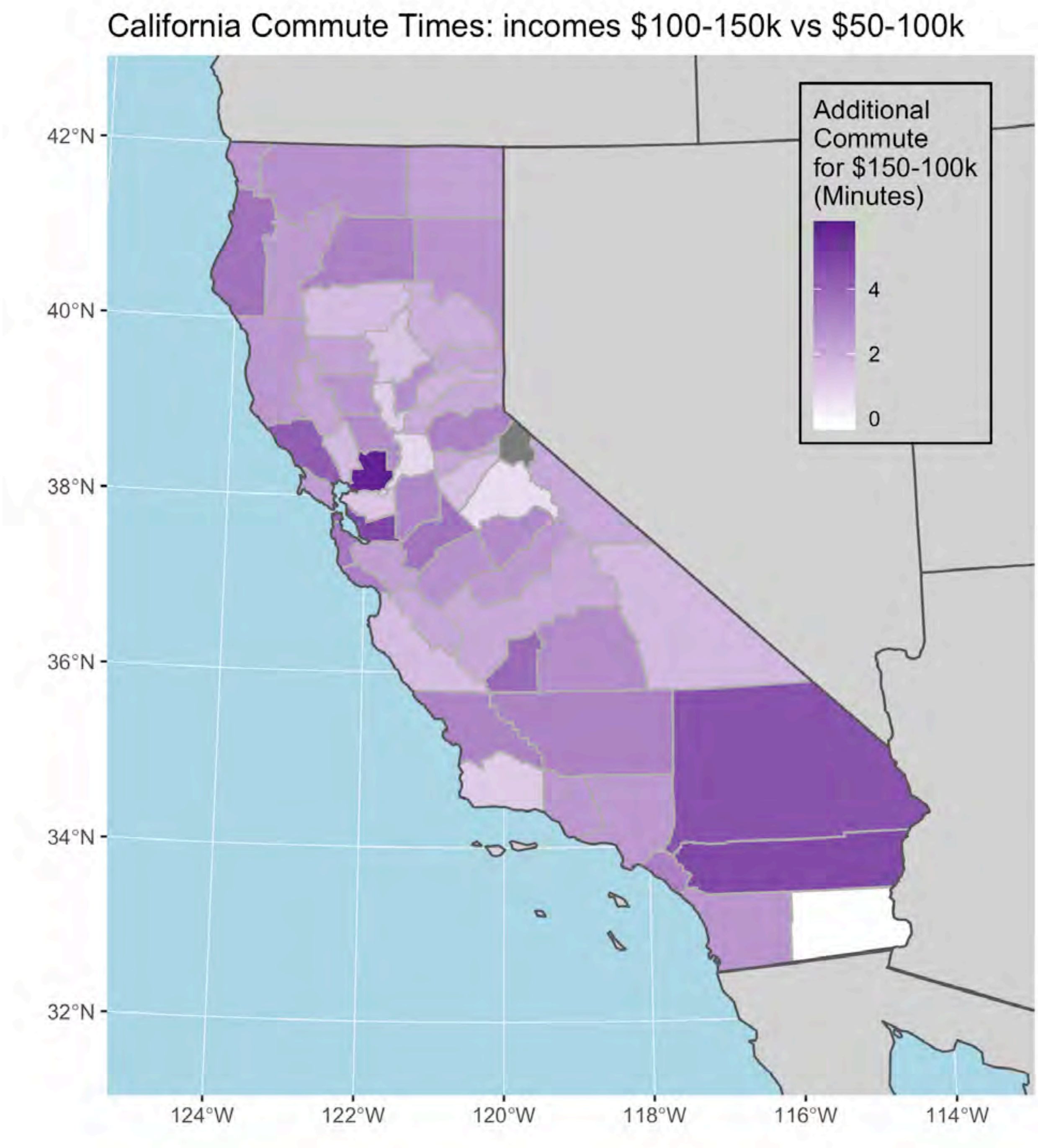

Figure 6.3 Wealthier people have longer commutes in every county, but the difference is largest in suburban counties.

7. Summary of Findings and Travel Behavior Research

In this research, we apply spatial clustering methods to a business establishments dataset in order to identify commercial centers and demonstrate a method to parameterize this clustering and develop destination attractiveness indicators. Many of these commercial centers are located along major infrastructure links, which means the activity opportunities they present would be poorly modeled by aggregating to census units or TAZs, which are often divided by these infrastructure links. We find that these different types of centers attract different kinds of activities at different times of day, and we confirm the significance of these differences using bootstrapping.

Among our specific findings, we identify substantial differences in the duration of shopping activities in different types of centers, as well as in the timing of work, eating, and entertainment. In most cases, people spend more time on ostensibly similar activities in centers that present a larger and more diverse range of opportunities, and these activities tend to extend later into the evening. Smaller centers that spread throughout the region's suburbs seem to attract shorter duration and earlier evening activities. Since many of these timing differences occur around major congestion times of day in the region, spatial differences in activity scheduling are important for travel behavior modelers and transportation planners to understand.

The commute travel time analysis shows fundamental differences in travel time among people working in different industries but also differences across California. This analysis also demonstrates clearly the need to identify centers of activity in proximity of home and work locations separately. It also shows the presence of these centers in proximity of home and work are associated with both a higher travel time and lower travel time depending on the type of center. This is the single most important indication that we need a different taxonomy than the simpler density and diversity of environments around home and work. Although we demonstrate here spatial heterogeneity in behavior for activity participation and work travel time, the study is not a complete representation of behavioral determinants.

Are these differences in what people do when at different places better understood in terms of type or in terms of timing/sequencing? This study suggests that more specificity may be needed for recording some types of activities in travel surveys; shopping for groceries or a quick stop at a convenience store are very different from shopping for clothes or browsing a bookstore, but they would be recorded as the same activity purpose in the CHTS. In contrast, the difference between an early dinner in a local restaurant on the commute home and a late dinner before going to a club in Hollywood may be more about what one plans to do before and after dinner than the dinner itself.

These findings are also particularly useful for understanding destination choice. Basing spatial analysis on the distribution of opportunities rather than on units derived from counting jobs and residences may be a more informative way to represent spatial opportunities for activity-based forecasting and destination choice models. Studies of spatial heterogeneity (*Paez and Scott, 2004, Bhat and Zhao, 2002, Bhat et al., 2013*) that include commercial centers as part of their specification may find improved results because different places have inherently different time of day signatures. It is important to recognize that specific activities and trips occur as parts of tours and that people interact with each other. In this project, we briefly address the relationship between dining and entertainment activities, but further consideration of other potential activity pairings is vital to understanding the relationship between place characteristics and destination choice.

In the course of this project we received a variety of comments about the method used here. One of the comments is about the difficulty researchers face in identifying and using fine grained data about business establishments. The solution to this is usually to accept the aggregation error due to the use of traffic analysis zones readily available through the US Census. A few regional agencies define their own analysis zones (dubbed micro-analysis zones) to counter the detrimental impacts of traffic analysis zones. We address this issue using inventories of business establishments from private sources and developing centers instead of zones. Researchers perceive the creation of different types of spatial units as a difficult task and we give empirical proof that many analyses can be done starting from business establishments that provide not only new insights but also key performance indicators for planning (i.e., spatial distribution of centers of different sizes and mix of opportunities).

We turn now to feedback we received using one type of crowdsourcing technique about the new methods here. In the 15th International Conference on Travel Behavior Research (July 15-20, 2018) in Santa Barbara that hosted approximately 320 researchers from around the world we also had the opportunity to present the center identification and timing of activities. We also had the opportunity to review how other researchers approach access to transportation and accessibility. This is an ideal setting for crowdsourcing because on the one hand we offered a research presentation using material from this project and on the other hand we collected input from eight research gap workshops that took place throughout the conference by first directing participants to presentations and then through brainstorming sessions define research gaps and specific actions to bridge these gaps.

We present here a brief version of these research gaps and explain the connection with our project. In the first workshop on *Mobility as a Service* (see iatbr2018.org) the only explicitly spatial reference was made for research on the first mile/last mile potential of ride-hailing without a clear identification of land use policies. In the second workshop on *Time Use and Travel*, special generators are identified as in need of further study. The centers we define in this project can be used to develop a further taxonomy of these generators. In addition, this workshop identified as key direction further understanding of human interactions in space and time. Our project and its extension studying sequences of activities is moving along this direction of closing part of this research gap. In the third workshop on *Transport for Healthy, Happy, and Holistic Living*, understanding the role of the built environment on subjective well-being is identified as a key aspect but with emphasis on human interactions and jointly with social and virtual environments. Moreover, this workshop also calls for developing geographic taxonomies enabling the study of differences and commonalities in perceptions and behavior. This workshop also pointed out to the need for characterizing places by the type of opportunities they offer. The methods used to identify centers in this project are characterizing location by the type of activity opportunities (including job types) and can be used as a stepping stone for identification of different geographies and their affordances. The fourth workshop on *Automation and Self-Driving* did not address diversity and mix of opportunities but points out to the potential this technology has in shifting job and residential locations but also destination choices. In the fifth workshop on *Data-Driven Learning and Travel* a variety of methods and their potential to fill research gaps were reviewed. DBSCAN in the way we use it in this project is a data-driven learning technique that is then enhanced by studying patterns of time of day in activity and travel. The sixth workshop on *Life Course and Dynamics* addressed a time scale that is not central to this research project. In the fifth workshop on *Big Data and Travel* spatial analysis was only addressed in terms of aggregation of movement traces during an observation period but did not lead to any direct input to this project and spatial databases did not emerge as a Big Data source of information in the workshop. The last workshop on *Connected Freight* addressed issues of supply chain and logistics at multiple scales (from the global to the local/last mile) and the potential of a variety of technologies including automation. A variety of issues were identified explaining lack of behavioral research about goods movements. However, the participants pointed out the potential automation has in changing supply chain and logistics that are heavily dependent on the spatial distribution of shippers and receivers. Identification of centers in the way we do in this project can facilitate modeling and simulation of goods movement in regional planning.

We find this type of crowdsourcing a more efficient method than originally planned for a variety of reasons including: a) as organizers of the IATBR2018 conference we were able to strategically insert questions about our work in different settings such as presentations and workshop brainstorming; b) reached a wider audience from many different countries and different research backgrounds; and c) were able to compare our research work with the state of the art in travel behavior research in themes that expanded the original repertory of the VESPI project.

8. Performance Based Planning and Next Steps

In this research project we demonstrate the usefulness of characterizing the type of destinations people seek in activity participation. We also show the relationship between these destinations and timing and duration of activities as well as commuting travel time. This study is a feasibility proof that finer spatio-temporal resolution can lead to a new family of key performance indicators for an integrated land use-transportation system. These indicators can be used to test policies of changing land use and assessing their impacts on daily travel behavior. We also show that timing of activity participation and travel times to work are very different among jobs of different industries and these are heavily influenced by the type of centers surrounding them. This in turn shows we need to introduce a fine grade definition of jobs and their distribution in space to gain insights about the impact of policies that are at the intersection of land use and transportation. For example, employer-based demand management strategies (e.g., staggered work hours, telecommuting) will have different impacts on different job types depending on the opportunities offered around home and job sites. This in turn also implies the mix of job types at a location will determine the impact of these policies in space. Recording the number and type of jobs in different locations will then allow us to assess the impact of policies with higher precision and accuracy.

Immediate Next Steps

The work in this project shows how we can represent activity opportunities in space and time to capture the spatial fragmentation in opportunity locations and hints on the need to also include temporal fragmentation of people's activities. Research on spatial fragmentation shows that work is more likely to be fragmented in space when people use information and communication technologies (*Alexander et al., 2010*). In addition, accessibility may also be better defined by representing the multiplicity of locations that can be accessed to participate in activities (see the review by *Couclelis, 2000*). Missing from the activity-travel behavior research is conclusive evidence that spatial fragmentation and temporal fragmentation are strongly related and one determines the other. Spatial fragmentation (Figure 4.1 is representative of spatial fragmentation) is indicative of urban sprawl and motivates the need to travel longer distances by car. Temporal fragmentation may lead to less free time and lower subjective well-being. Moreover, the advent of many new technologies such as automated vehicles and Mobility as a Service (including ubiquitous ride-hailing) may increase both fragmentations nullifying efforts to build sustainable communities as in Senate Bill 375.

Manifestation of activity participation flexibility is a schedule with a sequence of activities with multiple switching between different activities in a day (e.g., the sequence of escort a child- go to work- eat meal-run an errand-got to work- go to a social event- go back to work- go shopping - escort a child to soccer practice – do some work using mobile technologies – escort child back home - work at home) leading to increased transport demand because many activities are no longer bound to specific times and specific places. In a new PSR project that follows VESPI we first study the relationships among activity-travel fragmentation, social interaction, and then move to consider the role accessibility plays in fragmenting a daily schedule. We examine fragmentation of activities and travel using a fine-grained temporal analysis of 1440 minutes in a day. In this way, patterns of switching among activities and travel are analyzed as transitions in a minute-by-minute basis. Patterns of transitions are then correlated with social interaction and accessibility. *Social interaction* is included in that project and defined as the propensity of people to participate in activities and travel with other people and their propensity to serve other people. For example, escorting children to school, meeting other people for work, or meeting friends are all social interactions. Accessibility is a key indicator explaining travel behavior when it is computed as the access to services at fine spatial levels measuring the opportunities offered to residents as we see in the findings of this VESPI project. The immediate next step is to use this fine-grained (in space) accessibility and correlate it with activity-travel fragmentation.

Figure 8.1 shows minute by minute sequences of activities for a family. In a 1440 minute schedule this is a series of time points at which a subject can move from one discrete "state" to another. People with many “states” in their daily schedule have fragmented schedules. Our intend is to analyze sequences of activity participation to study fragmentation of respondents' days using a minute-by-minute time series, in which every minute of the day contains a specific "state" for each person in the study. These states are based on types of places individual people visit during their diary day: Home, Work, School, Other, and Travel. Then, we test the hypothesis that spatial fragmentation and temporal fragmentation are related and identify different relationships between these two fragmentations. In essence, we seek to identify the spatio-temporal correlation between Figure 4.1 and Figure 8.1 and then relate this to policy analysis.

In this way, we will be able to show if persons that live in a dense and diverse environment (e.g., center of cities) tend to break down their activities and move from one place to another more often in a day than persons living in residential suburbs that are far from dense environments. Findings from this analysis are directly applicable to the California SB375 policies that motivate increases in density and diversity of urban environments as a policy tool to decrease vehicle miles travelled (VMT) and greenhouse emissions. These ideas are at the forefront of research and the IATBR2018 workshop directions are indicative of this. Our most recent paper (*McBride et al., 2019*) sets the stage for the study of activity fragmentation and it was accepted for presentation at the forthcoming 98th Annual Meeting of the Transportation Research Board (Washington D.C., Jan 13-17, 2019) receiving positive comments. This provides further proof that the VESPI project and its extension through temporal fragmentation is a new idea worthy exploring further.

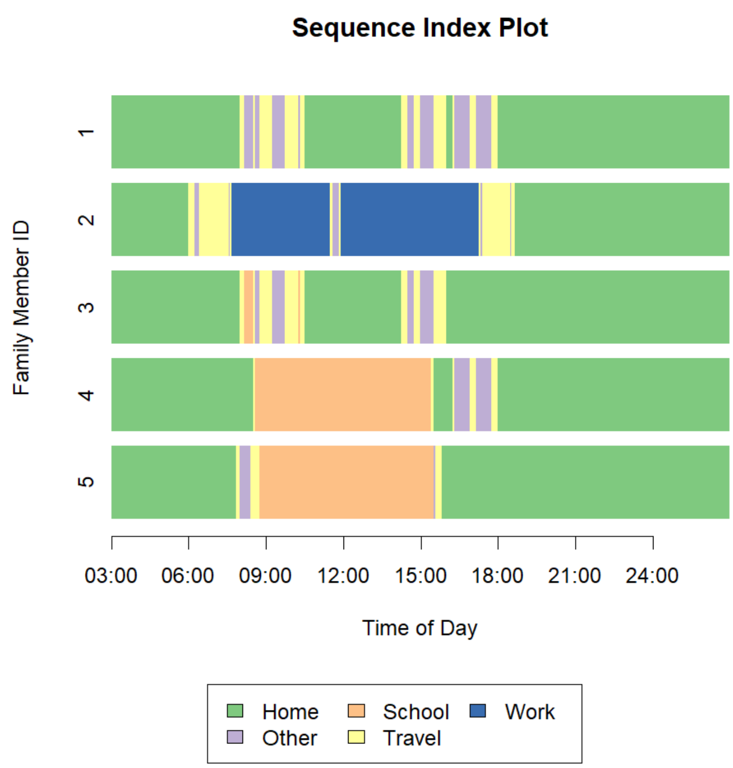

Figure 8.1 One Family's Sequence in a Day
(Reproduced from MacBride et al., 2019)

Bibliography

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.

Besser, L. M., Marcus, M., & Frumkin, H. (2008). Commute Time and Social Capital in the U.S. American Journal of Preventive Medicine, 34(3), 207–211. https://doi.org/10.1016/j.amepre.2007.12.004

Bhat, C. R., Goulias, K. G., Pendyala, R. M., Paleti, R., Sidharthan, R., Schmitt, L., & Hu, H.-H. (2013). A household-level activity pattern generation model with an application for Southern California. Transportation, 40(5), 1063–1086. https://doi.org/10.1007/s11116-013-9452-y

Bhat, C.R., & Zhao, H. (2002). The spatial analysis of activity stop generation. Transportation Research Part B: Methodological, 36(6), 557–575. https://doi.org/10.1016/S0191-2615(01)00019-4

Blumenberg, E. (2004). En-gendering Effective Planning: Spatial Mismatch, Low-Income Women, and Transportation Policy. Journal of the American Planning Association, 70(3), 269–281. https://doi.org/10.1080/01944360408976378

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. Transportation Research Part D: Transport and Environment, 2(3), 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6

Chen, Y., Ravulaparthy, S. K., Deutsch, K., Dalal, P., Yoon, S. Y., Lei, T., … Hu, H.-H. (2011). Development of Indicators of Opportunity-Based Accessibility. Transportation Research Record: Journal of the Transportation Research Board, 2255, 58–68. https://doi.org/10.3141/2255-07.

Couclelis, H. (2000). From sustainable transportation to sustainable accessibility: Can we avoid a new tragedy of the commons?. In Information, Place, and Cyberspace (pp. 341-356). Springer, Berlin, Heidelberg.

Davis, A. W. (2015). Investigating Place Attitudes in Santa Barbara, CA. University of California, Santa Barbara, Santa Barbara, CA.

de Abreu e Silva, J., Golob, T. F., & Goulias, K. G. (2006). Effects of Land Use Characteristics on Residence and Employment Location and Travel Behavior of Urban Adult Workers. Transportation Research Record: Journal of the Transportation Research Board, 1977, 121–131.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise (pp. 226–231). AAAI Press.

Giuliano, G., & Small, K. A. (1991). Subcenters in the Los Angeles region. Regional Science and Urban Economics, 21(2), 163–182. https://doi.org/10.1016/0166-0462(91)90032-I

Gobillon, L., Selod, H., & Zenou, Y. (2007). The Mechanisms of Spatial Mismatch. Urban Studies, 44(12), 2401–2427. https://doi.org/10.1080/00420980701540937

Hahsler, M., Piekenbrock, M., Arya, S., & Mount, D. (2018). dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms (Version 1.1-2). Retrieved from https://CRAN.R-project.org/package=dbscan

Handy, S. (1993). Regional Versus Local Accessibility: Implications for Nonwork Travel. Retrieved from https://escholarship.org/uc/item/2z79q67d

Hansen, W. G. (1959). How Accessibility Shapes Land Use. Journal of the American Institute of Planners, 25(2), 73–76. https://doi.org/10.1080/01944365908978307

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. Computers, Environment and Urban Systems, 54, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001

Karami, A., & Johansson, R. (2014). Choosing DBSCAN Parameters Automatically using Differential Evolution. International Journal of Computer Applications, 91(7).

Lee, J. H., Davis, A. W., Yoon, S. Y., & Goulias, K. G. (2016). Activity space estimation with longitudinal observations of social media data. Transportation, 43(6), 955–977. https://doi.org/10.1007/s11116-016-9719-1

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 Using Latent Class Cluster Analysis. Transportation Research Record: Journal of the Transportation Research Board, 2666, 58–68. https://doi.org/10.3141/2666-07

Levinson, D. M. (1998). Accessibility and the journey to work. Journal of Transport Geography, 6(1), 11–21. https://doi.org/10.1016/S0966-6923(97)00036-7.

Levinson, D., Marshall, W., Axhausen, K.: Elements of Access. Network Design Lab, Sydney (2017)

Maddimsetty, R. P. (2018, March 19). You Are (Probably) Here: Better Map Pins with DBSCAN & Random Forests. Retrieved July 27, 2018, from https://engineering.foursquare.com/you-are-probably-here-better-map-pins-with-dbscan-random-forests-9d51e8c1964d

McBride E.C., A.W. Davis, and K. G. Goulias (2019) Fragmentation in Daily Schedule of Activities Using Activity Sequences. Paper accepted for presentation at the 98th Annual Meeting of the Transportation Research Board (Jan 13-17, 2019).

McKenzie, G., Janowicz, K., Gao, S., & Gong, L. (2015). How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest. Computers, Environment and Urban Systems, 54, 336–346. https://doi.org/10.1016/j.compenvurbsys.2015.10.002

Moeckel, R., & Donnelly, R. (2015). Gradual rasterization: redefining spatial resolution in transport modelling, Gradual rasterization: redefining spatial resolution in transport modelling. Environment and Planning B: Planning and Design, 42(5), 888–903. https://doi.org/10.1068/b130199p

NUSTATS. (2013). 2010-2012 California Household Travel Survey Final Report (p. 137). California Department of Transportation. Retrieved from http://www.dot.ca.gov/hq/tpp/offices/omsp/statewide_travel_analysis/Files/CHTS_Final_Report_June_2013.pdf

Orenstein, D. E., Frenkel, A., & Jahshan, F. (2014). Methodology Matters: Measuring Urban Spatial Development Using Alternative Methods, Environment and Planning B: Planning and Design, 41(1), 3–23. https://doi.org/10.1068/b38017

Páez, A., & Scott, D. M. (2004). Spatial statistics for urban analysis: A review of techniques with examples. GeoJournal, 61(1), 53–67. https://doi.org/10.1007/s10708-005-0877-5

Patterson, Z., & Farber, S. (2015). Potential Path Areas and Activity Spaces in Application: A Review. Transport Reviews, 35(6), 679–700. https://doi.org/10.1080/01441647.2015.1042944

Paul, B. M., Vovsha, P. S., Hicks, J. E., Livshits, V., & Pendyala, R. M. (2014). Extension of Activity-Based Modeling Approach to Incorporate Supply Side of Activities: Examples for Major Universities and Special Events. Transportation Research Record, 2429(1), 138–147. https://doi.org/10.3141/2429-15

Redmond, L. S., & Mokhtarian, P. L. (2001). The Positive Utility of the Commute: Modeling Ideal Commute Time and Relative Desired Commute Amount. University of California Transportation Center. Retrieved from http://escholarship.org/uc/item/4mc291p2

Rosenthal, S. S., & Strange, W. C. (2003). Geography, Industrial Organization, and Agglomeration. The Review of Economics and Statistics, 85(2), 377–393.

Schubert, E., Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (2017). DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. ACM Transactions on Database Systems (TODS), 42(3).

Shliselberg, R. (2015). Accessibility and Spatial Interaction. Transport Reviews, 35(6), 814–816. https://doi.org/10.1080/01441647.2015.1058302

US Census Bureau. 2010 Geographic Terms and Concepts - Census Tract. https://www.census.gov/geo/reference/gtc/gtc_ct.html. Accessed Jul. 27, 2018.

Viegas, J. M., Martinez, L. M., & Silva, E. A. (2009). Effects of the Modifiable Areal Unit Problem on the Delineation of Traffic Analysis Zones, Effects of the Modifiable Areal Unit Problem on the Delineation of Traffic Analysis Zones. Environment and Planning B: Planning and Design, 36(4), 625–643. https://doi.org/10.1068/b34033

Wachs, M., Taylor, B. D., Levine, N., & Ong, P. (1993). The Changing Commute: A Case-study of the Jobs-Housing Relationship over Time. Urban Studies, 30(10), 1711–1729. https://doi.org/10.1080/00420989320081681

Walls & Associates. (2013). National Establishment Time-Series (NETS) Database: 2012 Database Description. Retrieved July 22, 2017, from http://exceptionalgrowth.org/downloads/NETSDatabaseDescription2013.pdf

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. https://doi.org/10.3328/TL.2010.02.04.245-260