Report

Multiday GPS Travel Behavior Data for Travel Analysis

FHWA-HEP-015-026

Prepared for:

Federal Highway Administration

Submitted by:

RSG

In Cooperation with:

Jennifer Dill, PhD & Joseph Broach, PhD (candidate) (Portland State University)
Kate Deutsch-Burgner, PhD (Data Perspectives Consulting)
Yanzhi "Ann" Xu, PhD & Randall Guensler, PhD (Trans/AQ, Inc.)
David M. Levinson, PhD & Wenyun Tang, PhD (candidate) (University of Minnesota)

2200 Wilson Boulevard, Ste. 205
Arlington, VA 22201-3324
888.774.5986
www.rsginc.com

Table of Contents

List of Figures

List of Tables

List of Equations

Glossary of Terms

AB	Activity Based
ANOVA	Analysis of Variance
ARC	Atlanta Regional Commission
CATI	Computer-Assisted Telephone Interview
CDF	Cumulative Density Functions
CHTS	California Household Travel Survey
EM	Expectation Maximization
GIS	Geographic Information System
GTFS	General Transit Feed Specification
HOT	High-Occupancy Toll
LCCA	Latent Class Cluster Analysis
MHT	Multiple Hypothesis Technique
ML	Maximum Likelihood
MTA	Metropolitan Transportation Authority
MTC	Metropolitan Transportation Commission
NREL	National Renewable Energy Laboratory
NSR	Not-Same-Route
OLS	Ordinary Least Squares
PM	Posterior Mode
PSRC	Puget Sound Regional Council
PSTP	Puget Sound Transportation Panel
RP	Revealed Preference
SP	Stated Preference
SR	Same-Route
SRNSP	Same-Route NOT Shortest-Path
SRSP	Same-Route-Shortest-Path
TBI	Travel Behavior Inventory
TIAS	Trip Identification and Analysis System
TRB	Transportation Research Board
TSDC	Transportation Secure Data Center
TTAPS	Toronto Travel-Activity Panel Survey
TTAPS	Toronto Travel-Activity Panel Survey
VOR	Value of Reliability

VOT	Value of Time

Introduction to Report

By Mark Bradley (RSG)

The use of GPS devices to collect trip-specific data as part of household travel surveys has increased steadily in recent years, and will likely become the main mode of travel survey data collection in the future as smartphone-based platforms for collecting travel data come into use. Compared to diary-based methods, the advantages of GPS data capture include the following:

- The time and location of each trip end can be captured with more precision.
- There is less potential for respondents to omit entire trips or activities from the survey.
- The data can be used to trace the route traveled for any particular trip.
- It becomes more cost-effective to capture multiple days of travel for each respondent.

These unique aspects of GPS data enable new types of behavioral analysis relative to those conducted with more traditional travel survey data. In particular, multiday data capture, in combination with more precise and complete travel data on each day, allows researchers to investigate day-to-day variability in travel behavior at the individual and household level. Such analyses can provide more insight into peoples' travel patterns at a broader level, and guide future efforts in modeling and predicting travel behavior and designing transportation policies.

Large-sample, multiday GPS datasets from household travel surveys are still relatively limited in quantity, as is the expertise required to process point-by-point GPS trace data into trip-level data that can be used by most analysts. To address these issues, the US Department of Transportation and the National Renewable Energy Laboratory (NREL) have created the Transportation Secure Data Center (TSDC).[1] The TSDC allows researchers to access preprocessed data from almost one dozen different multiday GPS travel datasets from across the United States; it also allows researchers to analyze these data in a secure environment that ensures the protection of data privacy.

The two main objectives of this project are: 1) to provide new examples of the type of valuable research that can be done using multiday GPS travel survey data; and 2) to demonstrate that such research can be conducted in the TSDC research environment. Each of the following four chapters describes a research project that was funded and carried out as part of this project. The four research topics were originally specified by RSG, with input from FHWA, and then further refined by the authors during the course of their research.

[1] TSDC: http://www.nrel.gov/transportation/secure_transportation_data.html

In "**The Effect of Day-to-Day Travel Time Variability on Auto Travel Choices**," Jennifer Dill, PhD, and Joseph Broach, PhD (candidate), of Portland State University address the important research topic of measuring the effect of auto network reliability on drivers' choices. Using data from a 7-day vehicle-based GPS survey in the Atlanta region and a longer-duration vehicle-based GPS survey in the Seattle region, the authors identified several cases where respondents made multiple car trips between the same origin-destination (O-D) pairs during the survey period, and measured the actual experienced day-to-day travel time variation for those O-D pairs. The authors report several interesting analyses showing that such variability is related to trip and traveler characteristics, including trip purpose, distance, and household income.

In "**Multiday Variation in Time Use and Destination Choice in the Bay Area Using the California Household Travel Survey**," Kate Deutsch-Burgner, PhD, of Data Perspectives Consulting, investigates day-to-day variation in the number, types, and level of dispersion (distance) of destinations visited during the specific days of a 3-day person-based GPS survey in the California Bay Area. Using the technique of latent class cluster analysis (LCCA), she is able to distinguish clearly different patterns of variability in terms of number of trips and type and dispersion of destinations. This analysis method shows promise for addressing the complexity of multiday travel data, and may become even more useful as future person-based (e.g., smartphone-based) GPS datasets include a greater number of travel days and a potentially wider variety of different patterns across the days.

In "**Capturing Personal Modality Styles Using Multiday GPS Data—Findings from the San Francisco Bay Area**," Yanzhi "Ann" Xu, PhD, and Randall Guensler, PhD, of Trans/AQ, Inc., analyze the same multiday GPS dataset from the Bay Area that was used for the analysis described in the preceding chapter. In this analysis, however, the focus is on day-to-day variation in mode choice—research for which person-based, rather than vehicle-based, GPS data collection is clearly necessary. The authors were able to identify distinct groups of individuals in terms of whether they always used the same mode or used a variety of modes, and in terms of whether auto or alternative modes were used more often. They were also successful in relating these groupings to different person and household characteristics. The propensity to use multiple modes would benefit standard travel modeling methods, as someone who usually uses auto but also uses transit one or two days per week may be more likely to increase his or her transit use in response to service changes, as compared to someone who never uses transit at all.

Finally, in "**An Empirical Study of the Deviation between Actual and Shortest-Travel-Time Paths**," Wenyun Tang, PhD (candidate), and David Levinson, PhD, of the University of Minnesota, use multiday person-based GPS data from the Minneapolis region Travel Behavior Inventory (TBI) to determine how often drivers use the shortest path for their home-to-work trip, and look at patterns in the deviation in travel time between the shortest path and the actual path. In terms of day-to-day variability, the authors were not able to identify many cases where respondents made the same direct home-to-work auto trip on multiple days. This outcome indicates that analyses that measure travel behavior across multiple days (rather than simply treating them as separate single days) will tend to require large sample sizes, particularly when the analysis focuses on a specific type of behavior (e.g., direct home-to-work auto trips).

The research presented in the following four chapters provides interesting findings in their own right, and insights into the types of research designs and methods that will be valuable in analyzing multiday GPS data as it becomes more ubiquitous and accessible in the future. The authors generally recognize that their methods could benefit from larger sample sizes, in terms of the number of respondents, and particularly in terms of the number of days per respondent. (For example, use of 7-day GPS data capture periods would allow analysis of patterns, including both weekdays and weekends.) The authors also note the critical importance of how the GPS trace data are processed into trip-level data, and the need for evolving practices and standards in GPS data processing. Finally, the authors describe the value of the TSDC in making these unique datasets available while providing a secure and productive research environment.

Chapter 1.0 The Effect of Day-to-Day Travel Time Variability on Auto Travel Choices

By Jennifer Dill, PhD[2] & Joseph Broach, PhD[3] (candidate) (Portland State University)

1.1 Introduction and Literature Review

> "Quantitative research into [value-of-time] variability and value of reliability (VOR) has lagged because of the lack of data on the day-to-day travel time variability that drivers face for particular trips."(1)

This paper explores the potential of archived GPS data to expand the understanding of travel-time reliability. While reliability is often observed and considered at the system or segment level, travel-time uncertainty is also experienced at the household and trip level. Any move toward incorporating reliability into regional travel models will necessitate a re-examination of travel-time variation at more disaggregate levels. This chapter presents some observations of reliability at the household level using multiday vehicle-based GPS data analyzed within the Transportation Secure Data Center (TSDC).

The research team embarked with three major goals for the chapter. The first goal was to consider the ways in which multiday GPS data could be translated into data on reliability. The second goal was to explore relationships between trip- and household-level travel-time reliability and related trip, household, network, and urban location factors. The third goal involved use of the topic of reliability as a case study to test the usefulness of the TSDC in its current form for academic research.

Travel time has long been a central measure of both system performance and project benefits in urban transportation systems. Extensive literature has been developed around quantifying the value of time (VOT) spent traveling.(2) More recently, a consensus has formed that travel-time variability is often as costly to travelers as average trip times.(3–6) Meanwhile, work is just beginning on incorporating reliability into travel modeling practice.(1) While empirical work to date has established some rough guidelines for the value of reliability VOR, only a limited slice of what real-world reliability—or unreliability—looks at the level of everyday household travel.

Most travel-time reliability studies to date have been either stated preference or have studied specific facilities, such as high-occupancy toll (HOT) lanes. Even when individuals have been the unit of analysis, measures of reliability have usually been in the aggregate (loop detectors) or by proxy (floating cars). The emergence of GPS

[2] Jennifer Dill, PhD, Director, Transportation Research and Education Center (TREC) and Professor, Nohad A. Toulan School of Urban Studies and Planning, Portland State University, PO Box 751, Portland, OR 97207.
[3] Joseph Broach, PhD (candidate), Nohad A. Toulan School of Urban Studies and Planning, Portland State University, PO Box 751, Portland, OR 97207.

data allows measurement of both reliability and response at the individual trip level. Understanding travel-time reliability at a more disaggregate level is a necessary step toward its incorporation in travel demand models. Furthermore, there are potentially broader contributions to an understanding of individual decision-making under uncertainty, information search, and habit formation.

When travel between two points is not reliable, the unpredictability may force a traveler to leave earlier to minimize the risk of arriving late.[(7)] If the morning commute averages only 20 minutes, but once a week takes twice as long, then the commuter might allocate 40 minutes to make the commute every day. Even when the precise arrival time is less important (e.g., for a shopping trip) there are costs to the uncertainty and having to spend longer than expected in the car (or bus or train).[(7)]

Both stated preference (SP) and, more recently, revealed preference (RP) methods have been used to measure value of reliability (VOR) almost exclusively in a route choice framework. In SP work, comparing pairs of alternatives by mean travel time and a short sample of early and late arrivals has become the preferred technique.[(5)] RP data has mostly been limited to semi-controlled experimental settings, primarily HOT lanes in California and Minnesota.[(8–12)]

A few studies to date have used GPS data to examine both travel-time reliability and response at the individual level. The goal in each case was to estimate a route-choice model that included experienced reliability. Carrion and Levinson collected 8 to 13 weeks of GPS data for auto commuters after the collapsed I-35 bridge in Minneapolis had just reopened.[(10)] Carrion and Levinson also collected 6 weeks of GPS data (though participants were instructed where to drive during the first 4 weeks) for users of three competing routes in Minneapolis, one of which was tolled.[(11)] Another recent study used 12 months of auto GPS data from the Seattle Traffic Choices Study that set up virtual toll roads that participants could pay to use from an allowance fund.[(1)] While data were similar to that used here, the focus in each case was on particular facilities and not the reliability experienced across all trips.

Unique in the literature is a study by Bachman et al.[(13)] The authors used vehicle trip GPS data collected as part of a Denver-area household travel survey to measure delay at the network-link level. They then considered, among other things, household (e.g., size, vehicle ownership, income, etc.) and urban-form (e.g., CBD, fringe, urban, suburban, rural, etc.) variables related to link delay. While their focus was congestion and not reliability, their use of GPS data to link network performance with household experience is similar to this research.

The contribution of the present work is to propose measures of reliability based on multiday GPS data collected at the household level. Instead of sampling by specific roadways or points, the entire range of household travel on any road, for any purpose, at any time of day is considered. This addresses existing reliability questions from a different perspective and prompts new questions that cannot be answered with traditional reliability data. Publicly accessible GPS data archived at the TSDC has been used to conduct the analysis.

The rest of this chapter included the following: an overview of data selection and processing methods; a descriptive overview of household-level travel and reliability; an examination of household, trip, and urban environment correlates of reliability; a description of experience using the TSDC environment; and a discussion of findings, limitations, and suggestions for future work in this area.

1.2 Data Selection and Processing

The TSDC had nine different datasets available for analysis. The research team used two datasets to answer questions. The first dataset was the 2004–2006 Puget Sound Regional Council's Traffic Choices Study (PSRC); this dataset includes 18 months of data, though the total sample size is only 275 households. The second dataset used was the Atlanta Regional Commission (ARC) 7-day dataset because of the relatively large sample size (911 households with valid GPS data). The Seattle data are well suited to analyzing many common origin-destination (O-D) pairs from the same household, but likely have limited variation in terms of spatial patterns and other factors because of the household sample size. The Atlanta sample should provide greater variation, but a limited number of repeated O-D pairs for each household.

Analysis Samples

Table 1-1 provides an overview of trips included in the analysis sample. The PSRC data collection period stretched over 17 months. Because of the sheer size of the dataset and the complication of including the experimental phase (GPS-based tolling), analysis focused on the 3-month control period, from April to June 2005. This was the period after recruitment had been completed and during which participants were instructed to travel normally. The 3-month period included nearly 145,000 trips (over 5 million GPS points). Trips that started or ended outside of the region were further filtered. Based on analysis samples, a clustering algorithm (explained in a subsequent section) was applied to origins and destinations to identify repeated trips.

Table 1-1: Analysis Sample Description

	Puget Sound (PSRC)	Atlanta (ARC)
Households	272	911
Vehicles	405	1,648
Collection Period	3 months Apr. to Jun. 2005	7 days Mar.-May, Aug.-Sep. 2011

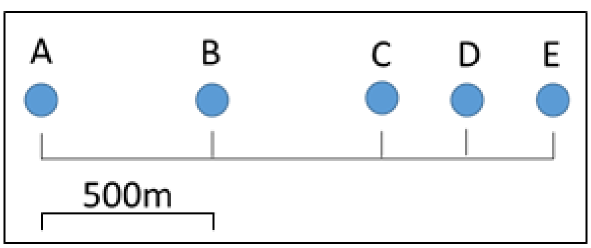
	Puget Sound (PSRC)	Atlanta (ARC)
Vehicle Trips Per Day (mean)	3.9	4.2
GPS Sample Interval	10 seconds	1 second
Total Trips	144,951	47,992
Trips to Repeated O-Ds (3+ times)	53,269	9,356
Repeated OD Pairs	9,172	2,366

Identifying Repeated Trips

Reliability measures require a travel-time expectation. While anticipated driving time could be derived from sources beyond direct experience (e.g., a navigation service or an individual's dead reckoning), the most natural sources of expected travel time are previous trips between the same (or similar) points. For GPS data without auxiliary travel diaries, true origins and destinations are unknown to the analyst, and common trip ends must be implied based on proximity.

While grouping trip ends may appear simple, it actually is fairly complex. Consider the five trip ends in Figure 1-1. Given a threshold of 500m, then C, D, and E should be grouped given their relative proximity in the figure, but what about A and B, which are set apart in the figure? B is within 500m of both point A and C. Point B cannot belong to both an AB and a BCDE group or the corresponding trip would be double counted. Fortunately, clustering techniques exist to handle complex groupings systematically.

Figure 1-1: Grouping Challenge—If a 500m threshold is set, which of these five trip ends should be grouped?

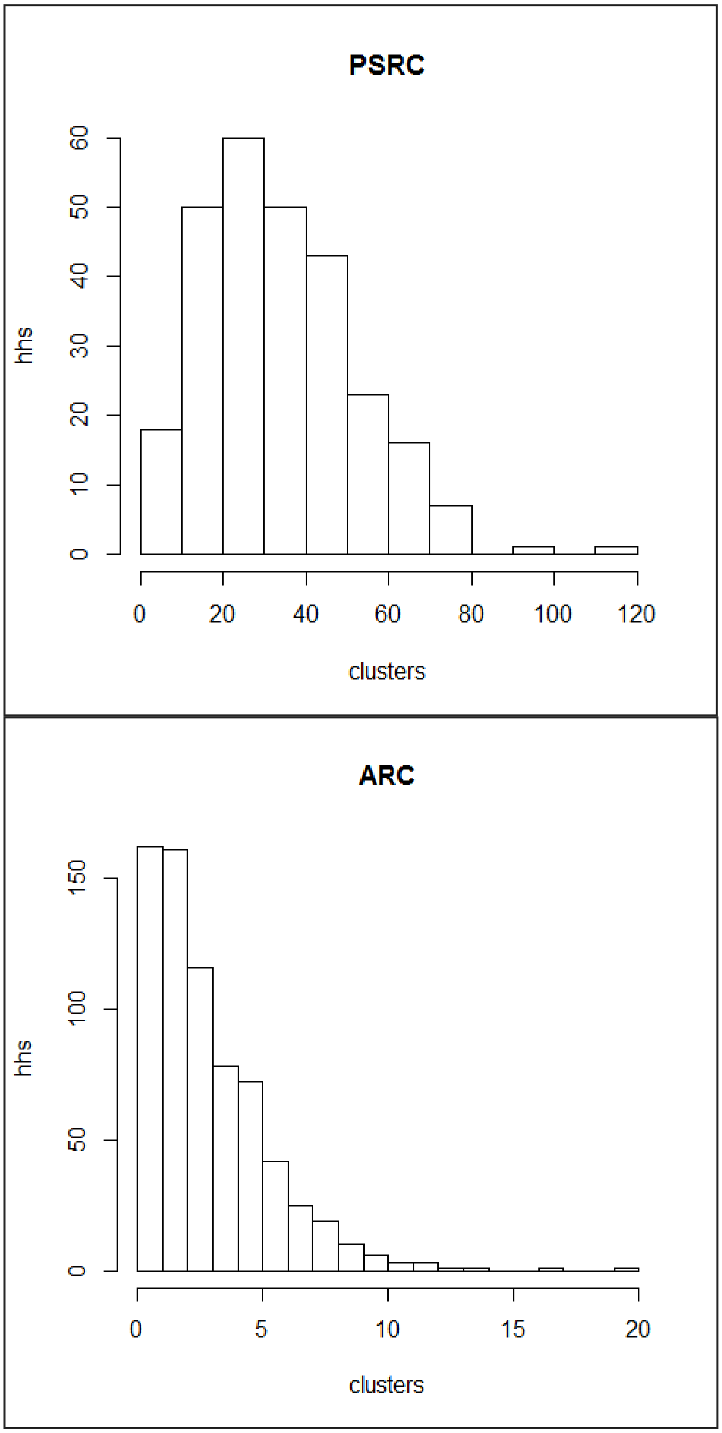


The research team applied an agglomerative clustering technique using Ward's method. This algorithm has been used in other cases to identify common trip ends in GPS data.[14,15] The clustering method begins with each origin point for a given household in its own group. With each iteration, the lowest-cost merge between nearby clusters is applied until a cost threshold is reached. The method is then repeated for destination points within each household. A cutoff cost was set at 500m (0.31 miles). This cutoff determines the maximum distance at which any two points will be grouped. This cost was found to be a reasonable tolerance to account for trips to and from the same or nearby locations, accounting for both parking search and GPS error. A 1000m threshold was also tested for the PSRC dataset and the number of repeated trips identified increased by less than 10%, suggesting the cutoff method was not overly sensitive to threshold choice.

Using the clustered origins and destinations, an O-D pair was considered repeated if at least three trips were recorded where both the origin and destination fell in the same origin and destination clusters. It was thought that setting the threshold at three instead of just two repeated trips would filter out some temporary common destinations (e.g., driving a child from home to an away soccer game, then returning later to pick the child up). Travelers on single-repetition trips seem more likely to view them as exceptions to household travel routines, and less likely to form travel-time expectations for them. Because the driver on a specific trip was not recorded, vehicles in multivehicle households were pooled. Trips with a total distance below the clustering threshold of 500m were excluded from the analysis, since these trips could represent travel within a single-trip end cluster.

As shown in Figure 1-2, this clustering method identified 53,269 repeated trips between 9,172 clustered O-D pairs in the PSRC analysis dataset. Trips repeated three or more times accounted for just under 37% of all trips recorded over the 3-month period. Households had, on average, less than 34 identified common directional O-D pairs. In the shorter 7-day ARC panel, 9,356 trips were identified as repeated three or more times between 2,366 unique O-D pairs. Repeated trips accounted for a smaller share of the ARC data—about 19.5% of all trips with an average of less than three repeated O-D pairs per household. Using the research team's definition of repeated trips, and assuming the PSRC and ARC data are comparable aside from collection duration, it appears that only approximately half of repeated O-D travel was captured in a 1-week collection period compared with a 3-month study. Figure 1-2 displays the variation in identified repeated O-D pairs across sampled households.

Figure 1-2: Repeated Trip O-D Pairs per Household for Each Sample



Comparable data on repeated trips are difficult to find, but there have been GPS studies of common destinations. Buliung et al. reported that 70% of destinations in a 7-day GPS study were visited at least twice.[16] If two trips is considered the threshold, 57% of PSRC and 34.5% of ARC trips were repeated. In a 5-day GPS sample, Dill and Broach found that 57% of trips ended at destinations visited three or more times.[15] They excluded trips ending at home. These are clearly not perfect comparisons, since in this instance both origin and destination are considered in repeated trips measure. The differences likely are due in part to variations in day-to-day trip chains.

The research team did consider time of day as an additional repeated trip criterion, but rejected the idea for two reasons. First, while a driver's expected travel time might change for the same trip taken at different times of day, literature suggests there might still be a reliability "frustration" cost in knowing that travel time has been much different in the past. Second, in order for this to work, the research team would have to assign a somewhat arbitrary time threshold in addition to distance beyond which a person's travel-time expectations would be assumed to reset. The research team was reluctant to assign a time threshold compared with one for distance, and clustering would have required either an extra step or some weight assignments to distance and time.

For the remainder of this chapter, the terms repeated O-D pairs, repeated trips, common O-D pairs, and common trips will be used interchangeably. Each refers to the definition of a trip between the same origin and destination cluster observed at least three times during the study period.

Map Matching GPS Data to Travel Networks

A second major exercise was to identify the actual network routes traveled on each common trip. The process of joining GPS tracks to the most likely series of links along a network is known as map matching. The PSRC data had not been map matched as part of the original processing, and the TSDC's in-house algorithm was not designed to handle such coarse data (10-second minimum interval between points). As a result, the research team developed a modified version of the Multiple Hypothesis Technique (MHT) map-matching algorithm, enhanced by a technique to "densify" the sparse GPS points.[17] The modified algorithm was designed to function completely within the TSDC analysis environment.

The ARC GPS data were recorded at much shorter intervals and had already been map matched in the TSDC. Unfortunately, a licensing issue with the network data provider prevented the TSDC from sharing the matched links in time for this analysis. A workaround is currently being developed so that "anonymized" network-link identifiers can be shared along with summary statistics. This would allow comparisons between different routes for the same trip without violating the license agreement.

The MHT applied to the PSRC data is well suited to matching GPS data to dense, complex travel networks. One of the challenges inherent in using GPS data on all household trips is the need for a complete network of local streets. Research focused on specific facilities, especially freeways and toll roads, can greatly simplify map-matching complexity by narrowing the search to just the relevant facilities. However, the research team's aims necessitated matching to the full network of local streets in the Puget Sound region. A suitable network had been provided to the TSDC by the regional government at the time of collection. It consisted of nearly 215,000 undirected network links and over 180,000 intersection nodes.

The matching algorithm is optimized to match GPS data with a density of at least one point per network link traversed. At the available ten-second-or-greater collection interval, a car traveling 30 mph (48 km/h) would cover approximately 440 feet (134 m) between recordings in the best-case scenario. That distance would cover two blocks in dense parts of the region, and the algorithm would not be able to predict the traveled path accurately.

To solve the point-density problem, the research team interpolated false GPS points along a straight line between each consecutive pair of actual points. The points were then evenly spaced so that the maximum distance between coordinate pairs would not exceed 200 feet (70m), roughly the length of the shortest blocks. This preprocessing step leveraged the predictive power of the MHT matching algorithm.

The MHT works by building a set of hypothesized paths and then updating the set with each new actual (or pseudo, in this case) GPS point. The set is updated at each step by either joining the new point to an existing hypothesized path, or else extending a path via a feasible travel maneuver. A cost function is then applied to each new candidate path, and a specified number of least cost paths are retained for the next iteration. The average point to matched link distance was the cost function, starting with the 25 nearest network links to the origin point, and keeping the 20 best candidate paths between each iteration.

An advantage of the MHT technique is the enforcement of feasible network paths. Topology is strictly enforced at each step so that, for instance, a GPS track that "jitters" between an elevated freeway and a surface street underneath will not result in a route that bounces impossibly between the two as can happen with purely proximity-based matching procedures. The MHT paths remember that a ramp to the elevated freeway was taken many miles before and the surface street option is ignored.

Original work on the MHT suggested an average cost threshold of 100 meters and a maximum of three "odd" links.(17) An odd link was defined as any link with a nearest matching point 75 meters or more away.(17) Because of the point interpolation, many links were flagged as odd around turns where the interpolated path followed a diagonal instead of the actual street network. In addition, the original work had included an additional speed limit deviation term that was not able to be duplicated due to a lack of speed limit data. Based on these differences, and using average point cost and odd link distributions as guides, the average point cost limit was reduced from 100 meters to 50 meters. The odd link threshold was relaxed to include matched routes with fewer than 10 odd links. Using these modified criteria, the research team successfully identified routes for 84% of trips between common O-D pairs. If the original cutoffs were used, the figure would have dropped to 75%, which is still a reasonable match rate given the data and conditions. For comparison, the original MHT work identified only about 53% of trips, though on a national (Swiss) network.

1.3 Household Travel and Reliability

The research team assumed that reliability at the household level manifests itself as travel-time variability for common trips. Common trips are those where one can reasonably expect the household to have an experience-based expectation of how long the trip will take. Common trips were operationalized as any trip between clustered O-D pairs taken three or more times during the study period, regardless of trip purpose, day of week, or time of day. Of course, this definition implies different minimum repeat rates for different sampling durations. In the PSRC data, a once-per-month trip would qualify as repeated, while in the ARC dataset, only a trip taken at least three times in one week would be identified as common. Definitions of common travel are far from established and the research team relied largely on judgment and data availability.[15] Other potential criteria to identify an O-D pair as common include day of week, time of day, trip purpose, and route. Route similarity is particularly interesting, since route switching might indicate either a strategy to improve reliability or an action attributable to reasons unrelated to reliability (e.g., variety seeking). The research team chose to focus on route switching as a strategic behavior and included trips between the same O-D pair as common regardless of the route used.

General Travel Patterns on Common Trips

The unit of analysis for experienced reliability is the common household trip O-D pair. Individual trips between the clustered trip ends had to be aggregated to arrive at group-level statistics. Trip purposes were generated based on O-D location type pairs as shown in Table 1-2.

Table 1-2: Trip-Purpose Distribution for Common Trip O-D Pairs

	Puget Sound (PSRC)*		Atlanta (ARC)*	
	ODs	Trips	ODs	Trips
Home-based work	21.0 %	25.5 %	21.1 %	18.7 %
Home-based non-work	56.6 %	54.2 %	56.0 %	58.5 %
Work-based other	11.9 %	11.8 %	6.0 %	4.9 %
Non-home non-work	10.4 %	8.5 %	6.7 %	7.0 %
Home-based school	n/a	n/a	8.7 %	9.4 %
School-based other	n/a	n/a	1.4 %	1.4 %

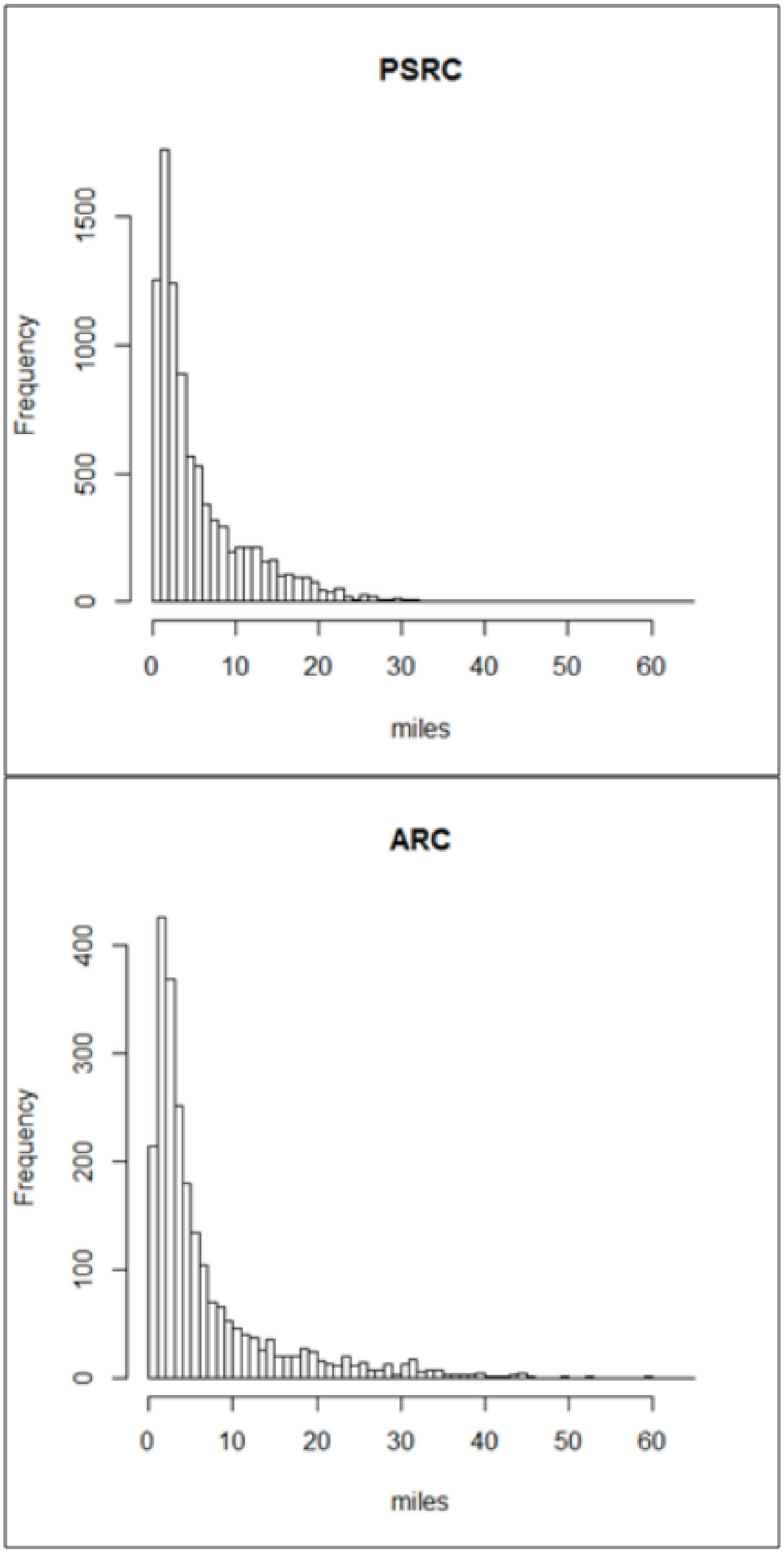
* percentages may not add to 100 due to rounding

When calculating common trip purpose, it was possible for trip ends falling in the same cluster to have been assigned different types, either due to coding errors or multiple location types in close proximity (within 500m). The research team imposed a location type hierarchy to ensure each cluster had a single location type and each cluster pair would have a single purpose. In the case of multiple location types assigned to a trip end cluster, the research team used the order Home → Work → School → Other.

Trip-purpose distributions were surprisingly similar for the two datasets despite the large difference in collection periods and any regional distinctions. The great majority of repeated O-D pairs were anchored either at home or work (nearly 90% of repeated O-Ds in both samples). In the longer sample, a smaller fraction of common O-Ds were non-home-based. School trip data were only available with the ARC dataset.

Mean GPS distance was used to aggregate trips within each common O-D group. As shown in Figure 1-3, average distance distributions for common trips are similar for both datasets. Most common are short, 1-2 mile trips. Based on the shape of the distance distributions, the research team categorized trips into three distance tiers for further analysis: 1) short trips less than 4 miles; 2) medium-length trips between 4 and 10 miles; and 3) long trips greater than 10 miles. Trips with an observed distance of more than three times straight-line distance were excluded from analysis, since these circuitous routes were probably the result of coding errors or nonutilitarian travel.

Figure 1-3: Common Trip O-D Mean Distance



Reliability Measures

It is important to make a distinction between unreliability and expected delays. A highly congested route is not necessarily an unreliable one if the congestion merely results in longer but consistent travel times. Measures of travel-time reliability are based on the variability around a traveler's travel-time expectation. Since actual driver expectations for a given trip would be difficult to capture, travel-time distributions for similar trips are commonly used to measure both expected travel time and travel-time variability. More research is needed to understand how additional sources of information might affect travelers' reliability perceptions. For instance, if a driver uses the radio or other sources of current travel information to adjust his or her travel-time expectation (before or perhaps even during a trip) then to what extent are the costs of additional travel time reduced, if at all?

Most existing work has used data from only a portion of a trip along a set network segment (e.g., a common stretch of freeway). Travel times along a segment are normally measured using roadway-based traffic sensors such as loop detectors.(8,12) This method allows for a large sample and the ability to examine systematic variation due to factors such as time of the year, week, or day. At least one study used floating cars to improve temporal and spatial resolution.(9) Aside from aggregation errors, a principal disadvantage of these segment-level techniques to examine trip reliability is their capture of only a portion of each trip. Drivers may experience completely different levels of reliability before and after they traverse a measured segment.

Multiday GPS data allow for true trip-level observations of travel time and reliability and promise to bring us closer to measuring experienced rather than predicted reliability. Two studies in Minnesota have successfully used trip-level GPS reliability data in route choice models for specific facilities.(10,11) So far, no published work has used multiday GPS data to examine the overall reliability experience of households for repeated trips on all facilities.

Several reliability measures have been proposed and used in studies that measure trip-level behavior. A number of studies have used percentile offsets from the median (e.g., 75th minus 25th percentile travel time or 90th percentile minus the median).(5,8–10,12) The standard deviation of travel time around the mean is also commonly used.(8,11) A recent summary report stated a preference for standard deviation adjusted for distance, citing improved model performance.(1) The report suggested that longer trips are likely to have longer delays in absolute terms, but each minute of unexpected delay on a shorter trip may represent a greater reliability cost to the traveler.(1)

The research team measured trip-level reliability as the standard deviation in travel time recorded for common trip O-D pairs divided by the mean GPS distance for the O-D pair. The resulting ratio summarizes the expected variation in travel time per unit distance.(1) The research team felt that this measure was more appropriate than percentile offsets and raw standard deviation given that the observations varied considerably in both distance and number of observations.

Reliability at the Common Trip Level

Trip reliability varied significantly by purpose, time of day, and distance as shown in Figure 1-4, which presents mean reliability values averaged across common trip O-D pairs. The standard deviations provided reflect reliability variance within a given common trip category. For all common trips, expected standard deviation of day-to-day travel time for the same trip ranged from 0.38 minutes per mile (0.24 min/km) in the ARC data to 0.69 minutes per mile (0.43 min/km) in the PSRC data. As a result, for a typical 5-mile (8 km) trip, one would expect 95% confidence intervals from about plus or minus 3.8 to 6.9 minutes.

Results for purpose and time of day emphasize the fact that congestion and reliability are different phenomena. A consistently congested route may be quite reliable, while one with only sporadic traffic snarls may be less so. Commute trips and peak-hour trips were generally more reliable than midday trips. It may also be the case that such trips are better planned in terms of route or departure time compared with (potentially more flexible) midday trips.

Longer-distance trips in both samples are progressively more reliable than shorter trips on a per mile basis. This is perhaps not surprising, since random delays have more time to balance out on a longer-distance trip. Similar to commute and peak period trips, longer trips might also benefit from better route and departure time planning.

The contrast between the two samples is consistent across every trip breakdown. The PSRC sample exhibited poorer reliability and more variation across common trips compared with the ARC data. There are several potential explanations for this. First, the contrast could represent a true reliability difference between the two regions; there certainly are considerable geographic and network structure differences between Seattle and Atlanta. Second, the timing of the studies could have played a part. The PSRC data were collected before the Great Recession in the United States, and one result of that event was a well-chronicled decline in vehicle travel. Third, the longer collection period of the PSRC sample (3 months vs. 1 week) could have resulted in a broader range of irregular—but repeated—trips being captured. Such trips may be subject to greater variation in scheduling and perhaps less planning than routine trips.

The differences seem worthy of further study, especially in regard to optimal collection periods for GPS reliability data. One possibility would be to compare the full ARC dataset with random 7-day samples from the PSRC data to see whether the reliability differences persist, but that is beyond the scope of the present study.

Table 1-3: Common Trip Reliability Means

(standard deviations in parentheses)	Puget Sound (PSRC) SD(minutes) / mile	Atlanta (ARC) SD(minutes) / mile
All common trips	0.69 (0.84)	0.38 (0.40)
Home-based work	0.39 (0.36)	0.26 (0.29)
Home-based non-work	0.71 (0.81)	0.39 (0.37)
Work-based other	0.85 (1.18)	0.49 (0.57)
Non-home non-work	0.89 (1.02)	0.47 (0.45)
Home-based school	n/a	0.45 (0.51)
School-based other	n/a	0.47 (0.42)
6AM – 10 AM	0.36 (0.52)	0.28 (0.31)
10 AM – 3 PM	0.78 (1.09)	0.36 (0.38)
3 PM – 7 PM	0.56 (0.59)	0.33 (0.34)
7 PM – 6 AM	0.59 (0.85)	0.20 (0.26)
Multiple times of day	0.76 (0.87)	0.43 (0.43)
0-4 miles	0.91 (1.0)	0.52 (0.46)
4-10 miles	0.45 (0.43)	0.26 (0.28)
10 miles and up	0.31 (0.27)	0.19 (0.19)

Reliability at the Household Level

Aggregating common household trip O-Ds produces measures of experienced household-level reliability over the sample period. Mean reliability across all common O-Ds provides a sense of travel-time predictability for the entire range of a household's driving. Since one response to poor reliability on a trip might be to reduce travel or use a different mode, travel-time variation for less-frequently observed trips would not necessarily have a smaller influence on perceptions of overall reliability than for more-frequent trips. For example, imagine a household that has a reliable commute driven frequently, but also has a number of other common trips that are unreliable and for which they avoid driving whenever possible. The household might well have a negative view of travel-time reliability, even though most of its observed travel is predictable. On the other hand, when considering reliability burdens, it seems important to weight by frequency, since some households will be unable to shift travel away from unreliable trips. For this reason, the research team also aggregated reliability weighted by trip frequency. As GPS reliability data becomes more common, an interesting area of further research will be comparing observed versus perceived reliability at the household level.

Figure 1-4 presents the distribution of common O-D travel-time reliability across households in each sample. There is considerable variation in reliability experienced at the household level. In each sample, the majority of households encounter a similar level of reliability while a smaller number of households experience substantially more travel-time variability.

There are also differences between the two samples in terms of household outcomes. PSRC households could expect a standard deviation in repeated trips of about 42 seconds per mile (26 s/km). ARC households exhibited a lower variability in travel time with a mean standard deviation of 23 seconds per mile (14 s/km). Weighting O-Ds by trip frequency made little difference at the sample level. To what extent the better performance in the Atlanta region reflects an actual performance difference or just an artifact of the samples and collection duration would require further study.

Figure 1-5 plots frequency of driving trips between common O-Ds and trip-weighted reliability relative to other households in the sample. If households confronted with lower reliability respond by driving less, one would expect a downward trend in the data with households traveling more when travel-time variability is low. In neither dataset is such a trend observable. A linear regression model confirmed no significant relationship between auto travel frequency to common destinations and experienced reliability, controlling for household sociodemographics and urban/rural residential location.

Figure 1-4: Mean Unweighted Reliability of Common Trips at the Household Level

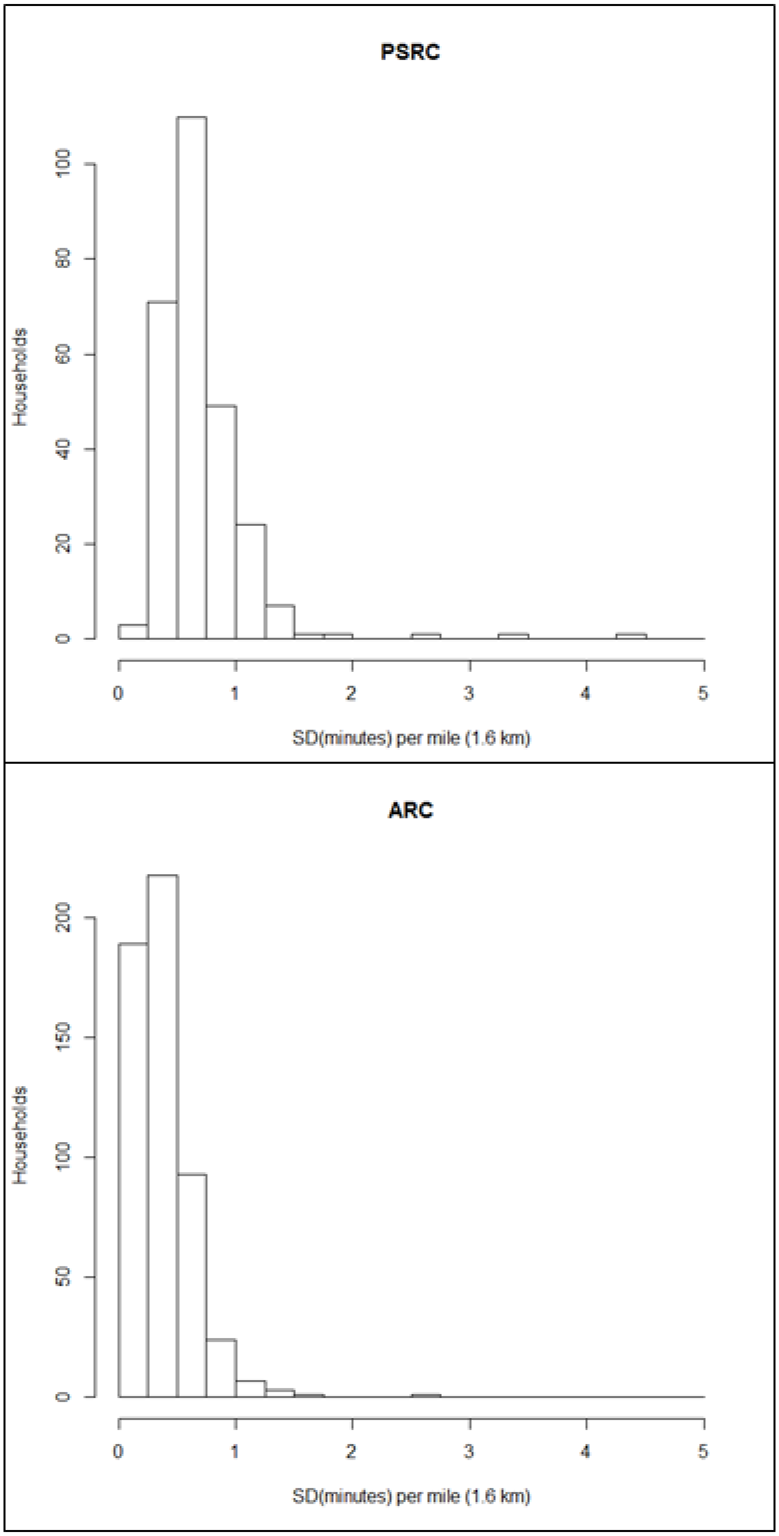


Figure 1-5: Households by Relative Reliability and Travel Frequency

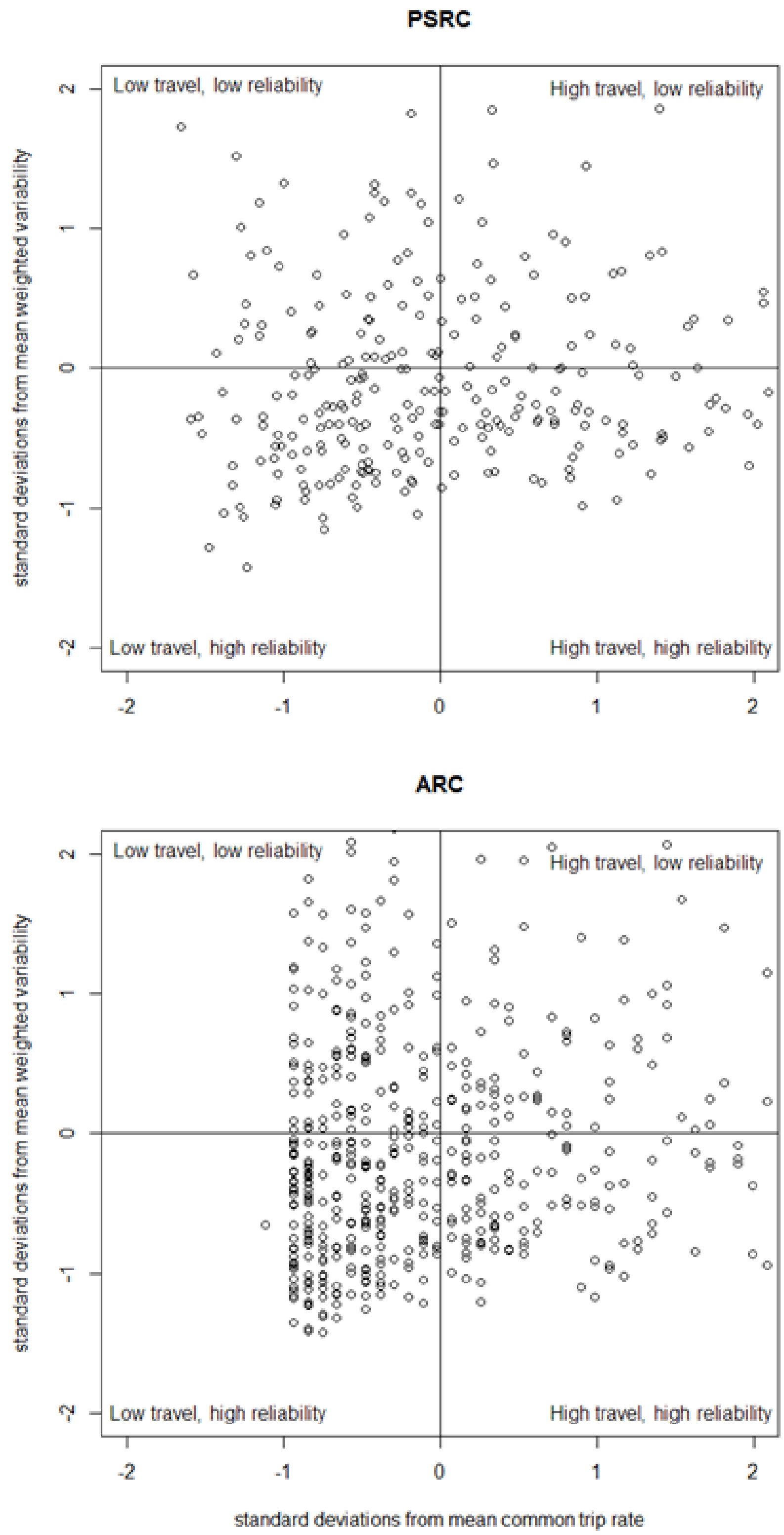


Figure 1-5 groups households into four classes. These classes are as follows, beginning in the first (upper-right) quadrant and moving counterclockwise: 1) high frequency, low reliability; 2) low frequency, low reliability; 3) low frequency, high reliability; and 4) high frequency, high reliability. Group 1 (high frequency, low reliability) is perhaps the most perplexing; despite experiencing poor reliability, Group 1 drove at high rates. One explanation is that this group represents captive households that lack the flexibility, resources, or travel options to shift away from unreliable auto travel. Another possible explanation is that many high-frequency travelers are self-selected from a population that is not particularly sensitive to driving time variability.

Table 1-4 explores differences in household characteristics across the four reliability groups for each dataset. In this instance, travel frequency and reliability refer only to driving trips between identified common O-D pairs. Both regional samples display the same trend with regard to household income, though the differences are greater in magnitude in the PSRC data. Both low-travel groups have relatively lower income—virtually identical regardless of reliability. The high-travel, high-reliability group has the highest income, although the difference from Group 1 is only significant in the PSRC dataset ($p<0.05$). An interpretation that matches findings in other reliability work is that low-income travelers place a lower value on reliability and are therefore relatively insensitive to experienced reliability. An alternative explanation—one with different policy implications—is that the lower-income, low-travel groups lack the means or flexibility to adjust travel (e.g., time of day) to their reliability environment. That the income difference is much more drastic in the Puget Sound region merits further study.

Table 1-4: Characteristics of Different Household Auto Travel-Reliability Groups

		2. Low Auto Travel		1. High Auto Travel	
		PSRC (n=61)	ARC (n=126)	PSRC (n=45)	ARC (n=90)
Low Reliability	Income	54,296 (36,061)	87,595 (43,845)	71,182 (38,165)	94,342 (38,511)
	Vehicles/driver	1.02 (0.19)	1.0 (0.35)	1.00 (0.19)	0.90 (0.41)
	Number of kids	0.55 (1.02)	0.73 (1.0)	0.93 (1.02)	1.23 (1.06)
	Seattle city resident (PSRC) or Urban home location (ARC)	52.5 %	12.2 %	35.6 %	1.2 %
	Own home	73.2 %	82.6 %	79.5 %	95.1 %

		3. Low Auto Travel		4. High Auto Travel	
		PSRC (n=96)	ARC (n=215)	PSRC (n=66)	ARC (n=105)
High Reliability	**Income (midpoints)**	54,615 (35,044)	86,016 (39,360)	92,333 (63,039)	99,755 (39,938)
	Vehicles/driver	1.01 (0.20)	1.06 (0.44)	1.04 (0.22)	0.98 (0.48)
	Number of kids	0.22 (0.55)	0.72 (0.97)	1.10 (1.24)	1.51 (1.47)
	Seattle city resident (PSRC) or Urban home location (ARC)	49.0 %	1.0 %	27.3 %	1.0 %
	Own home	75.3 %	89.9 %	92.1 %	94.8 %

*Standard deviations in parentheses

The only significant difference in vehicle sufficiency (vehicles/drivers) is found between the ARC groups 1 and 3. The high-travel, low-reliability households in Atlanta are more likely to share a car than the low-travel, high-reliability group. Splitting time with a single vehicle may allow for less flexibility to avoid times of low reliability.

High-travel groups are significantly more likely to have children than low-travel-frequency households. Children add common trips to a household schedule, but there does not appear to be a clear link to reliability in either sample.

As proxies for urban environment, the research team examined reliability groups by home location. In the PSRC data, no simple classifications were available, so home addresses in the City of Seattle were used as an urban/suburban proxy. Approximately 42% of households lived in Seattle city limits. Seattleites were overrepresented in both low-travel groups. This may be due in part to greater accessibility to nonauto modes in the city. The high-travel, high-reliability group was the most likely to live in "suburban" areas, suggesting self-selection of frequent travelers to areas of higher travel-time reliability.

The ARC data had an urban and suburban classification of home location. Only 3.5% of the sampled households were in "urban" areas. This trend was somewhat different than in the PSRC data. Urbanites were overrepresented only in the low-travel, low-reliability group, with all other groups being overwhelmingly suburban. This again raises the question of whether Group 2's overall travel is restricted due to a poorer reliability environment, or whether they are merely able to leverage more reliable nondriving options.

Finally, the research team examined housing tenure by travel group. In the PSRC sample, only Group 4 had a significantly higher home ownership rate. Housing choice may reflect a self-selection trend similar to the one found for urban versus suburban location. Those households with preferences for more travel locate in suburban areas with higher rates of home ownership. In the Atlanta region, higher rates of home ownership were apparent in both high-travel groups, and the differences across groups were smaller in magnitude.

Available evidence points to some interesting socioeconomic differences across travel and reliability groups. In the PSRC data, a picture emerges of relatively wealthy, suburban families taking advantage of reliable networks to travel frequently by private vehicle. Relatively lower-income households more likely to live in the city and with fewer kids fall more or less equally into the two low-auto-travel groups. Further investigation is needed to understand to what extent lower auto travel rates are a response to reliability.

In the ARC data, the picture is perhaps less clear, and the groups are generally more homogeneous. The high-travel, high-reliability group is also somewhat wealthier and more likely to have kids, but the differences are only significant relative to the low-travel groups. An interesting finding is the large overrepresentation of urban households in the low-travel, low-reliability group. While there may not be a clear overall trend between reliability and auto travel frequency, there is evidence that reliability conditions and travel outcomes vary significantly across different socioeconomic groups.

1.4 Route Choice and Reliability

When travel time along a particular corridor becomes unreliable, a household has several options. It might switch mode, adjust departure time, or reduce travel between the pair of locations. Given the option, a potentially less disruptive adjustment might be to explore alternative routes. For example, a freeway might be chosen initially due to expected lower travel time, but if the freeway becomes unreliable, a slower but more consistent parallel surface street might be preferred. Reliability has consistently been found to affect choices among competing routes, even if a more reliable route requires a toll payment.[4,8–12]

Existing work has usually been framed as a quasi-experiment with a known or artificially restricted choice set of alternative routes. The research team did not have that luxury here. Without a detailed route choice model, the research team's information on route alternatives had to be derived from observed travel. A route was considered an alternative if it was used for a common trip at least once during the 3-month PSRC study period. Since the research team's reliability measure was derived from the same trips, an ambiguous causality problem was encountered.

Two hypotheses seemed plausible regarding the relationship between observed route variety and travel-time reliability for common trips. First, an increase in observed route variety could reflect the common finding that drivers will switch routes to improve reliability—deemed the search or response effect. This could be a medium- or long-term choice, or simply a short-term response to temporary conditions or information like traffic reports. Second, more route options available for a trip could indicate a more resilient network between a given O-D pair, increasing expected reliability.

Identifying Alternative Routes

Routes between identified common trip end pairs are subject to several sources of randomness and error. Since the research team was more interested in trips considered spatially similar rather than in identifying identical origin and destinations, a fairly large clustering threshold of up to 500 meters was specified. As a result, otherwise identical routes might start and end on different sequences of network links. In addition, GPS units are subject to initial recording delays (cold start) and general spatial error. Finally, the map-matching technique employed likely introduced occasional small errors, especially where true data points were sparse.

Given all of the potential sources of error at the unusually fine resolution level attempted here, the research team used a strong criterion to define an alternative route. As shown in Figure 1-6, the research team adopted a simple technique that grouped observed routes for a given common trip into groups of similar routes considered to be approximately the same. The goal was to place individual routes into the smallest number of groups within which no pair of routes differed along more than 25% of their length. The research team ignored the end links for each route, since these are often subject to the greatest noise in GPS data, the map-matching algorithm, and actual behavior (e.g., parking on different streets).

Figure 1-6: Identifying Unique Route Alternatives from Observed Travel

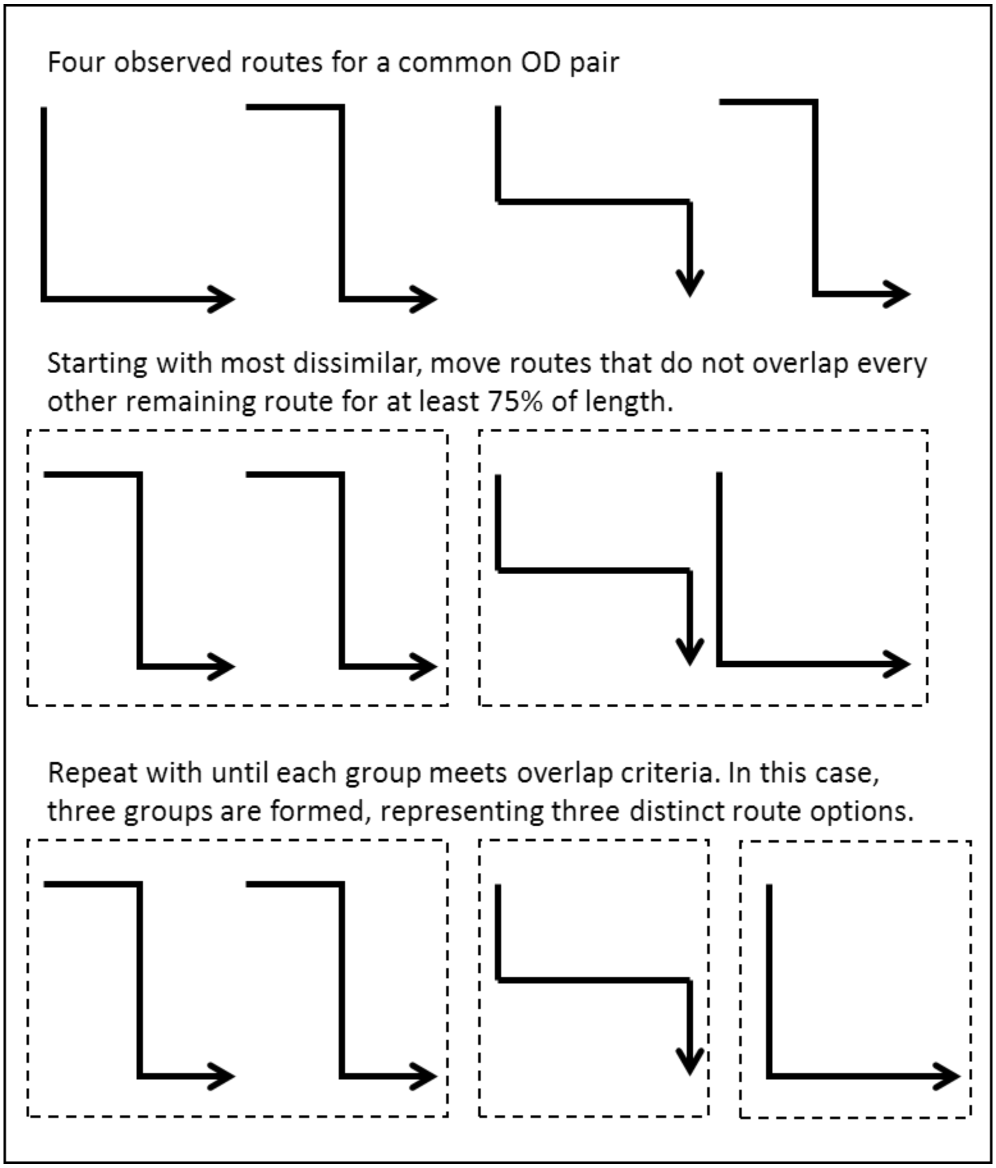
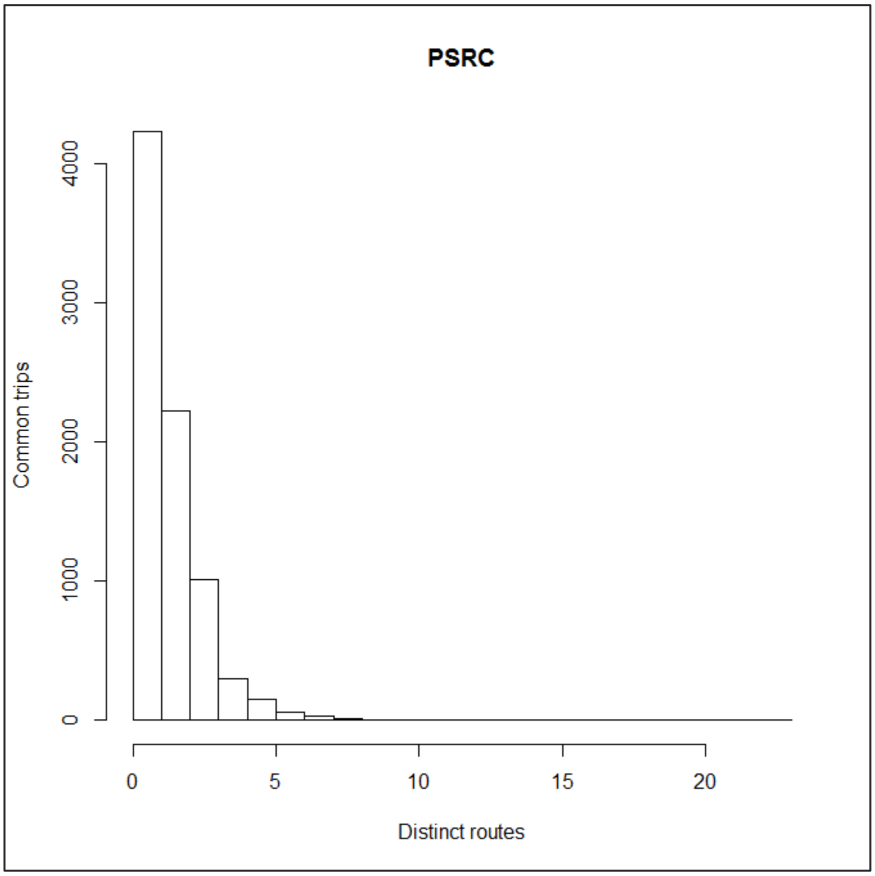


Figure 1-7 shows the distribution of the number of distinct routes across all common O-D pairs. For approximately half of sampled repeated trips, only a single route option was observed in the data. Most remaining trips had between two and five distinct route options.

Figure 1-7: Distribution of Distinct Routes for Common Trips (overlap≤75%)



Route Variety and Reliability

To test the hypothesis that greater route variety might be a response to poor travel-time reliability, the research team estimated a simple linear regression model. The unit of analysis is an observed trip with at least two valid, map-matched routes. The dependent variable is the number of distinct routes identified using the rule that considers any two routes with less than 75% overlap to be distinct.

Three independent variables were specified in the regression model: 1) the mean distance of observed trips between the common O-D pair; 2) the number of times the trip was observed; and 3) the measure of trip reliability—the standard deviation of travel time normalized by trip distance.

The coefficient for frequency was positive and significantly related to the number of distinct route alternatives, as expected. The frequency of observed travel sets an upper bound on the number of distinct routes observed. Mean trip distance had a negative and significant coefficient; this may be due to more randomness in route selection for short trips. While it was thought that longer trips may provide more opportunity for route variation, deviating from a known route on a long trip would require knowledge about a wider area of the network. It is also possible that the distance coefficient just corrects for a bias in the distinct route characterization. Since percent overlap was observed, a 4-mile (6.4 km) trip would require only 1 mile (1.6 km) of difference to classify a route as a distinct alternative; whereas, a 10-mile (16.1 km) trip would require a route difference of 2.5 miles (4 km). Determining at what point individual drivers consider a route unique over different-distance trips would be a useful area for further research.

Controlling for trip distance and frequency, the estimated coefficient for trip reliability was positive and significant. This is consistent with the hypothesis that travelers respond to increased travel-time variability by switching routes to optimize travel. An increase in travel-time standard deviation of about 4 minutes per mile would be expected to correspond with an additional distinct route. Model results are summarized in Table 1-5.

Table 1-5: OLS Regression of Distinct Routes per Trip (PSRC)

	Coefficient	SE	t-stat	p-value
Number of trips	0.08	0.003	29.9	<0.01
Average distance (miles)	-0.03	0.002	-12.3	<0.01
SD(minutes) / miles	0.24	0.02	16.0	<0.01
(Intercept)	1.4	0.03	51.6	<0.01
R^2	0.16			
n	8,015			

1.5 Using the TSDC

The TSDC is a unique resource—permitting secondary analysis of GPS travel data that otherwise would be difficult, if not impossible, to access. Acquiring large sets of survey data like those used in this paper from multiple agencies and their data partners is an uncertain task, at best. When supplementary data are needed, such as the decade-old PSRC travel network, the hurdles can be insurmountable. Even if acquired, such archived data are often poorly documented and in formats that require considerable work to convert.

The TSDC provides a relatively standardized archive of GPS travel data that allows researchers to access multiple datasets without engaging multiple agencies or consulting firms. TSDC is able to offer access on behalf of the original collecting agencies by allowing access and analysis within a specialized virtual environment. The spatial data itself remains on a centralized server in order to protect the privacy of respondents.

The tradeoff for easy access is an unusual workflow necessitated by the sandboxed environment. After applying for a computing account and data access, researchers access and analyze the data within a virtual machine, similar to a remote desktop session. All data are stored in a series of PostgreSQL (psql) databases on a remote server. Access to the databases is read-only, requiring that any new variables be created outside the database structure. It is also not possible to modify or add indexes to database tables to accelerate complex data queries, although the TSDC technical staff are receptive to alteration requests. Even the usual commands to export queried data for analysis (i.e., COPY TO) are blocked. As a result, most work must be done in intermediate plain-text data files. A possible future solution would be to allow researchers to apply for user space on the database server that would allow them to work on subsets of the spatial data within psql.

The TSDC virtual environment includes a useful suite of software tools preloaded and configured. Included are PGAdmin, Python(x,y), R, QGIS, Notepad++, ArcGIS, and the standard MS Office suite. Researchers do not have permission to modify the software tool configurations, but these were found to be smartly configured with all necessary extensions and connections needed to complete spatial and statistical analyses. Text files can be added to the environment, and the research team was able to import existing, Python-based map-matching modules by pasting them into blank text files in the virtual machine.

Researchers that normally rely on software not found in the included suite, such as SPSS, SAS, Stata, or any of the travel demand modeling software packages, may find analysis within the TSDC more challenging. The technical staff suggested that temporarily loading software may be possible, if the license allows it. Another option is to request that anonymized data without disaggregate spatial data be transferred out of the TSDC, but they try to avoid this. In this instance, the research team was able to complete all analysis within the virtual environment.

For researchers accustomed to running analyses on powerful, dedicated modeling workstations or servers, resources allocated to the TSDC virtual machines may be limiting. The environment currently provides only two processor cores and four gigabytes of memory. Since the databases are remote from the virtual environment, there is added overhead in executing large, complex queries. The map-matching algorithm employed by the research team, for instance, ran an order of magnitude more slowly than it would on even a low-end modeling workstation and required nearly one week and 13 manual restarts to complete. As demonstrated in this chapter, there is exciting potential to not only access—but to extend—existing GPS data through the TSDC, but the limited processing available currently presents a limitation.

One option discussed with technical staff was to allow researchers to request temporary increases to resource levels in the virtual machine, similar to the way organizations check out limited software licenses to those requiring them. Another option discussed for noninteractive scripts is to potentially submit them as jobs to be run on local modeling machines at the TSDC. This would require staff time to vet and run the scripts, and researchers would need to spend extra time developing the software to run without supervision.

In addition to archiving data, the TSDC reprocesses GPS data using a standard set of routines. These routines identify trips and tours and calculate distance, speed, and consistently calculate other drive-cycle statistics. This should improve comparability across datasets. Researchers then have the option of using the original data or the normalized, TSDC-processed GPS data. Since GPS travel data processing is far from standardized, the original and normalized data can differ considerably, and researchers must decide which version best fits their analysis needs.

The research team discovered large differences between the original and normalized (TSDC-processed) ARC data. The original processing split trips more frequently such that there were approximately 12,000 (30%) more trips in the original dataset. Manual inspection suggested that the original trip breaks may be more accurate. There were also differences in the trip-purpose distributions. For instance, there are 1,444 (2.8%) home-to-work trips in the original data, but only 774 (2.0%) in the normalized data. In addition, a substantial number of normalized trips were missing data for home, work, and school trip ends, while the original dataset had complete data. Given the importance of precise trip identification, the research team elected to use the original ARC data.

Differences in the original and normalized PSRC data were minor based on inspection, and the research team chose to use the normalized data. There was no reason to think the original data would be any more comparable to the original Atlanta data, and the normalized data already had spatial attributes.

Although the TSDC tries to obtain full documentation on submitted data, the research team did encounter some documentation gaps. The original trip-splitting algorithms and trip-end-matching procedures were unknown for both datasets. The PSRC network data used included a coded facility type variable, but the code definitions were unavailable. These issues were generally minor, and overall documentation was sufficient. The TSDC staff was also willing to go back to agencies for more documentation.

A final observation is that data and processing currently only flows one way, from the TSDC to researchers. There is no systematic way to report data errors, system bugs, or to contribute additions to the data as a researcher. If such a system were established, other researchers could leverage this team's additions to the data—such as the map matching of the PSRC data or reliability classifications of households—and avoid fixing the same errors once again. This would require additional staff time at the TSDC, but the benefits to the archive and to future researchers might be worth the cost.

Without the TSDC's central archive, gaining access to the data sources used would have been much more difficult. The archive encourages the use of multiple datasets by greatly reducing the marginal cost of adding additional GPS data sources to an analysis. This permits comparisons that generate interesting new questions and potentially improve the generalizability of results. Though GPS data collection has become less costly to collect, it is still a rare commodity in transportation research. That a team from Portland can apply new analysis techniques to GPS data collected in the Puget Sound and Atlanta regions beginning nearly a decade before, and that all of the analysis was conducted on data and machines located in Colorado, demonstrates the potential of the TSDC to further research. The accessible archive extends the useful life of valuable travel data and significantly expands the potential pool of analysts.

1.6 Study Limitations

The highly disaggregate GPS travel data used in this analysis allowed us to extend questions about reliability at a finer resolution than is usually possible. The data have limits, though. Many of the analysis variables had to be derived, and while the sensitivity of results to the assumptions was considerable (e.g., cluster method and tolerance for identifying common trip ends, distinct route thresholds), there is no doubt that different assumptions or techniques could change the results. There is also additional noise in data at this scale, from GPS signal errors to processing errors. With no ground truth available, it must be acknowledged that this random noise diluted the research team's ability to detect underlying patterns.

A particular limitation was the use of vehicle-based GPS data from the PSRC project. The research team was able to observe driving patterns in household-owned autos but could only speculate about the use and availability of other modes. Given the variety of car-sharing services now available, vehicle-based GPS data may soon not even capture a household's driving behavior sufficiently. Since the GPS units stay with the vehicle, these data are also more difficult to link to initial origins and final destinations. This influenced the research team's decision to use a fairly broad clustering tolerance when identifying common trip ends. Finally, the research team had no way of linking particular household members to a vehicle or trip. Therefore, no observations could be made at the individual level. Person-based GPS addresses some of these problems but also introduces new problems, such as imputing mode of travel from the constant stream of data.

While some comparisons were made between the ARC and PSRC datasets, it was recognized that there were considerable differences between the samples beyond location. The PSRC dataset was longer, conducted several years earlier, and used different GPS technology that recorded data at much coarser intervals. Some or all of these differences might explain part of the observed reliability differences between the datasets. On the other hand, the differences in data collection make the overall similarities in travel and reliability patterns observed even more surprising.

Finally, the usual caveats about sample representativeness apply—even more so when also trying to generalize across time. As noted at the outset, the research team considers this work to be exploratory. Methods and definitions for fine-scale GPS travel data are still in the early stages of development. With each additional study, true patterns and useful concepts and techniques will become more easily distinguishable from the noise.

1.7 Conclusion

Multiday GPS travel data can expand the field of view regarding the everyday experience of reliability. Instead of focusing on specific facilities and their use, it is recognized that trips, and the reliability of those trips, does not start and end at freeway ramps or toll barriers. Such comprehensive data over time also have the potential to reveal different responses to travel-time reliability (or unreliability). Households may adjust travel frequency, destination, time of day, route, or mode as they confront their unique travel landscapes.

The research team developed a method to identify repeated trips and used this method to calculate observed trip- and household-level reliability measures for both a 3-month vehicle-based dataset from the Puget Sound region and a 7-day dataset from the Atlanta region. The overall consistency of patterns in the results across the different regions, despite considerable differences in collection technique and duration, suggests that the research team's approach may generalize fairly well.

The research team found noticeable variation in reliability based on trip and household characteristics. No strong evidence was found in the data samples that reliability directly correlates with travel frequency at the household level. However, interesting groupings of household types into combined frequency and reliability classes were found.

The PSRC data was map matched to a travel network to allow route comparisons for common trips. The research team identified the number of distinct routes for each trip and found that the number of route alternatives tends to rise as travel time reliability for a trip decreases. This result agrees with previous findings from facility-specific studies that one response to high travel time variability is to seek out and use alternative routes to mitigate costs.

While the data provided a more complete picture of household reliability experiences, that picture was limited to travel in a household's private vehicles. This study clarified that a truly complete record of household reliability would require multiday person-based GPS data so that travel outcomes on all modes and trips could be considered.

Another natural extension of this work would compare trip-based GPS data with link- or segment-level data from other sources (e.g., from magnetic loop detectors). Fully incorporating reliability in regional travel demand models will require estimates of travel time variation between O-D pairs. Segment-based data capture only a small portion of any given trip, but they are collected across a broader sample of segment users and times. It would be interesting to examine how much of the total variation in trip travel times could be predicted from segment data from specific links along a trip's route. The two sources might be combined for travel model input (e.g., segment data could provide a base reliability measure, with GPS trip data used to create adjustment factors for different trip contexts).

The TSDC makes accessing archived GPS travel data, especially datasets from multiple agencies and regions, much easier. The TSDC analysis environment generally worked well for this research, though currently there are some limitations for resource-intensive spatial processing. The technical staff is working to address these limitations in the future.

Three pressing research needs emerged from this project as the study of multiday GPS travel data continues to develop. First, GPS data collection and processing methods must be continually formalized and improved so that data in archives such as the TSDC can be truly comparable. Second, further discussion, testing, and refinement of definitions are needed for phenomena like common travel, reliability, and distinct routes. Otherwise, research in this area will continue to be ad hoc and difficult to compare and generalize. Finally, efforts should be made to qualitatively link emerging definitions of reliability to individual perceptions. Both for research and policy, concepts like common trips and unreliability need to ring true when presented back to actual travelers. Multiday GPS data are poised to make strong contributions to understanding travel behavior and transportation experience. If moving from system and specific segments to household and complete trips is not without its difficulties, neither is it without its rewards.

1.8 References

1. Transportation Research Board (2014) *Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand*. Washington, D.C.: Transportation Research Board of the National Academies.

2. Mackie PJ, Jara-Dıaz S & Fowkes AS (2001) The value of travel time savings in evaluation. *Transportation Research Part E: Logistics and Transportation Review* **37**, 91–106.

3. Noland RB & Polak JW (2002) Travel time variability: A review of theoretical and empirical issues. *Transport Reviews* **22**, 39–54.

4. Brownstone D & Small KA (2005) Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice* **39**, 279–293.

5. Carrion C & Levinson D (2012) Value of travel time reliability: A review of current evidence. *Transportation Research Part A: Policy and Practice* **46**, 720–741.

6. Li Z, Hensher DA & Rose JM (2010) Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. *Transportation Research Part E: Logistics and Transportation Review* **46**, 384–403.

7. Bates J, Polak J, Jones P, et al. (2001) The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review* **37**, 191–229.

8. Lam TC & Small KA (2001) The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review* **37**, 231–251.

9. Small KA, Winston C & Yan J (2005) Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability. *Econometrica* **73**, 1367–1382.

10. Carrion C & Levinson DM (2012) A model of bridge choice across the mississippi river in minneapolis. In *Network Reliability in Practice*, pp. 115–129. Springer.

11. Carrion C & Levinson D (2013) Valuation of travel time reliability from a GPS-based experimental design. *Transportation Research Part C: Emerging Technologies* **35**, 305–323.

12. Ghosh A (2001) Valuing time and reliability: commuters' mode choice from a real time congestion pricing experiment. University of California, Irvine.

13. Bachman W, Oliveira M, Xu J, et al. (2012) Household-Level Global Positioning System Travel Data to Measure Regional Traffic Congestion. *Transportation Research Record: Journal of the Transportation Research Board* **2308**, 10–16.

14. Sharman BW & Roorda MJ (2011) Analysis of Freight Global Positioning System Data: Clustering Approach for Identifying Trip Destinations. *Transportation Research Record: Journal of the Transportation Research Board* **2246**, 83–91.

15. Dill J & Broach J (2014) Travel to Common Destinations: An Exploration Using Multiday GPS Data. *Transportation Research Record: Journal of the Transportation Research Board* **2413**, 84–91.

16. Buliung RN, Roorda MJ & Remmel TK (2008) Exploring spatial variety in patterns of activity-travel behaviour: initial results from the Toronto Travel-Activity Panel Survey (TTAPS). *Transportation* **35**, 697–722.

17. Schüssler N & Axhausen KW (2009) *Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT)*. Zürich: IVT, ETH.

Chapter 2.0 Multiday Variation in Time Use and Destination Choice in the Bay Area Using the California Household Travel Survey

By Kate Deutsch-Burgner, PhD[4] (Data Perspectives Consulting)

2.1 Abstract

Understanding variability in daily behavior is of the utmost importance in travel behavior modeling. Traditionally, this variability has mostly been assessed using travel diaries encompassing one day. However, conclusions reached from analyzing data from a one-day observation period could be incorrectly attributing variation seen in the sample to interperson variation (across people) rather than to possible intraperson variation (same person behaving differently across days) due to data limits. The research presented in this chapter examines the existence of intraperson variation in behavior. Using a sample of Bay Area residents who participated in the 2012 California Household Travel Survey, an individual's behavior across 3 days is examined. These data have been accessed using the Transportation Secure Data Center, housed in the National Renewable Energy Laboratory. Variations in day-to-day travel behavior are explored using data from respondents who carried a personal, wearable GPS data logger for 3 consecutive days. Exploratory analysis and summary statistics are first presented. Following this descriptive analysis, a latent class cluster analysis of the sample is performed. Results of this analysis are presented and allow for differentiation between "variability types" of individuals and behaviors. Aspects of variation in behavior across clusters are examined using spatial context of activity locations. A second latent class cluster analysis is used to develop a further understanding of variation in day-to-day behavior using frequencies of destination types. Sociodemographic indicators are used to explain cluster membership. Findings suggest that certain sociodemographic indicators—such as gender, employment, age, and others—are correlated with different "variation types" of individuals.

2.2 Introduction

Fundamental to the activity-based modeling paradigm is the need to understand and model the daily time use and activity patterns of individuals. Measuring and statistically explaining variation that occurs in human activity is fundamental to computationally modeling behavior. However, in order to accurately capture and model behavior, the proper data must exist. The standard practices of travel demand modeling have progressed over the past several decades to include more highly detailed activity data for modeling the complexities of behavior, which naturally places higher demands on the data collection process. However, traditional methods of paper and pen or pencil diaries are cumbersome and have high respondent

[4] Kathleen E. Deutsch-Burgner, PhD, Owner, Data Perspectives Consulting, www.dpccal.com, (805) 270-5360, kate@dpccal.com.

burden. Even with the integration of computer and web-based surveys, respondents are required to log their daily activities and describe the details (e.g., what, where, when, etc.). However, with the progression of the sophistication in methods used for analysis has come a progression in useful technology for data collection purposes. For instance, the use of GPS in household-level travel surveys reduces some of the respondent burden during the data collection process. By utilizing GPS, a portion of the data collection is automated, allowing for the passive collection of certain behavioral aspects, such as where and when the activity or trip occurred (provided that a respondent carries the GPS device, or it is installed in his or her vehicle). This technology enables researchers to expand many aspects of the survey (e.g., the duration of the survey or depth of detail asked about specific portions of the survey period). The collection of data across multiple days is the focus of the research presented in this report. GPS has also facilitated the collection of data for longer durations. However, these data usually include fewer details, as respondents are often not asked to record or annotate the activities that were logged for each day of a multiday survey; this limits the data to only that which can be derived using the GPS and a series of algorithms. Although this practice might be further advanced by integrating smartphones into data collection methodologies, most large-scale household travel surveys that utilize multiday GPS data collection are currently limited in this area.

Although the need for multiday and multiperiod analysis has been discussed within the travel behavior analysis and travel demand modeling community for over 40 years, the research is limited. This is primarily due to the limited datasets that include observations across multiple days or multiple periods. The history of multiday and multiperiod discussions and research will be presented in the literature review in the following section. Following the literature review, the data are described. Another section will present the analysis approach to exploring variation in behavior across days, and findings will be discussed. First, a descriptive analysis of key variables and findings related to variation in behavior across days will be presented. Following this, a latent class cluster model of day-to-day variation in travel attributes (and destination choice) is presented. The variability observed through this cluster model is further explored by examining the geographic extent of activities using a measurement of geographic point dispersion, which is known as standard distance. In addition, activity types are examined using the GPS coordinates and used to create a second latent class cluster analysis based on change in the composition of destinations from day to day. It is important to note that these two cluster analyses differ significantly in the objective. The first cluster analysis considers all trips made by respondents living in the Bay Area. This includes long-distance trips; as such, the cluster analysis highlights these aspects. However, in the second cluster model, long-distance trips are omitted to focus on the activity types in a more specific geographic area. Finally, the broader findings will be discussed along with conclusions.

2.3 Review of Relevant Literature

The necessity of collecting multiday or multiperiod data has been a topic of discussion for the past several decades. A fundamental shift occurred in the early 1970s that transitioned efforts away from merely planning and developing infrastructure to meet capacity needs to a focus on transportation systems that more holistically meet the demand of users (Pendyala and Pas, 2000). While it was recognized during this time that longitudinal research efforts were needed, there were few research initiatives to explore the concept. With time, the subject of observed variation in travel behavior using longitudinal surveys gained traction. By the 1980s, the discussion was active. Although there was a lack of longitudinal datasets, several empirical analyses utilizing the limited data sources were presented by a variety of researchers, which will be discussed in subsequent sections. With the improvement in technology over the last 30 years, and the survey methodologies presently available, the topic of longitudinal data analysis and model improvement can again be discussed, this time with a new perspective on possibilities. The increased ubiquity of technological devices and their relatively low cost have enabled large-scale data collection efforts to occur with longer durations of data collection.

Variation in behavior can be observed through an examination of a variety of temporal scales. It has been noted that the collection of data for a one-day period is a sensible and rational practice, as it follows a natural physiological repetition, and it is a convenient time unit while administering surveys (Kitamura, 1987). However, although the 24-hour day is a convenient and well-defined temporal unit, the ability of a one-day data collection effort to adequately capture behavioral differences is compromised. Without the ability to include repeated observations across days, it is impossible to assess whether the variation observed is due to interpersonal variation or intrapersonal variation. This concern, and the implications of day-to-day variability on behavioral modeling, are addressed by Hanson and Huff in a series of papers (Hanson and Huff, 1982; Huff and Hanson, 1986; Hanson and Huff, 1986; Hanson and Huff, 1988). In these papers, the authors deconstruct the commonly held and often unquestioned assumption of the "typical day." When these papers were published, the authors used one of the only existing datasets containing longitudinal data at the time—a 35-day diary from Uppsala, Sweden.

The early discussions and empirical analysis of multiday variation were primarily based on only a few datasets. The Uppsala dataset, collected in 1971, was an activity diary collected in Uppsala, Sweden, that consisted of a 35 consecutive-day data collection effort. An explanation of this dataset can be found in Hanson and Hanson (1981). A second dataset, the Reading dataset, conducted in 1973, was a 7-day data collection based in Reading, England (Shapcott, 1978). A third dataset, conducted by the Dutch Ministry of Transport, beginning in 1984, was a 7-day travel diary that was also conducted in several waves. This dataset is known as the Dutch National Mobility Panel study, and a description can be found in Pas (1988). Although a little later, many of the early empirical studies of variability also came from the Puget

Sound Transportation Panel (PSTP). This longitudinal panel survey was conducted by the Puget Sound Regional Council (PSRC), and was started in 1989 and ended in 2002. Research on this panel includes variation in both day-to-day behavior and panel year-to-panel year (see, for instance, Ma and Goulias, 1997; Goulias, 2002). The panel consisted of 10 waves of data, and contained an activity diary for each respondent over a 2-day period and additional respondent data (PSRC panel summary). Ten years after the start of the PSTP, a data collection was conducted in Germany in 1999. This data collection, in association with the project Mobidrive, consisted of a 6-week continuous travel diary for 361 persons in the German cities of Karlsruhe and Halle/Salle (Axhausen et al., 2000).

It is important when discussing longitudinal data to distinguish between multiperiod data and multiday data. As Pendyala and Pas (2000) describe, multiperiod data is collected over a longer time span, with one or more days of data collected consecutively, whereas multiday data is collected in a consecutive series of days. The collection of multiperiod data allows researchers to explore the variation across larger time scales (e.g., seasonality or across life stages). The analysis of multiday data facilitates the exploration of variation in an individual's behavior from one day to the next. Pendyala and Pas (2000) also discuss the disadvantages of 1-day, cross-sectional data collection efforts and address shortcomings of these types of data collection. The authors mention two sources of variability: 1) the day-to-day variability in a person's or household's needs; and 2) variability as a result of feedback from the transportation system. This known variability and the day-to-day dynamics render 1-day, cross-sectional data inadequate for modeling some aspects of travel behavior. After a detailed review of previous work, Pendyala and Pas outline considerations, challenges, and strategies for overcoming challenges presented while collecting repeated observation data. It is important to note that this paper, written in 2000, was addressing concerns and needs mentioned and explored in papers as early as the 1970s. Although there were a number of research studies, practical application in large-scale surveys that would address this hole in travel behavior data and analysis remained scarce between 1980 to 2000. It was only due to the reduced respondent burden by utilizing GPS technology that these types of data are now more regularly collected. Although limited in size, several foundational papers have been published exploring a variety of aspects of multiday variation. Hanson and Hanson (1981), using the Upsala dataset, condensed 51 measures of variability in travel from day to day into seven principal components of travel variability. They then used regression models to identify key sociodemographic indicators correlated with different factors of variability, finding that both role-related variables (e.g., employment, life cycle, sex, household size, and marital status), and socioeconomic aspects (occupation, education, and income) both explain variability; however, role-related variables were found to offer more explanation than sociodemographics. Kitamura (1987), using the Dutch Mobility Panel, has proposed a model of multiday travel patterns with the inclusion of latent patterns using a stochastic modeling approach. These latent patterns are essentially types of daily patterns, and it is assumed that individuals have more than one type of pattern and therefore have variation in behavior. Kitamura's analysis has shown that the existence of a latent pattern is a function of previous day's patterns. Pas (1988) on the other hand, using the Reading dataset, has proposed a methodology for analyzing variation by identifying travel activity type patterns and grouping individuals within clusters based on their spatial patterns. This is done first by developing a daily pattern by establishing a geographic similarity index based on point-to-point relationships within an individual's travel during the day. Individuals are then clustered together using a latent class cluster analysis based on the multiday geographical patterns, and exploring cluster membership in light of sociodemographics. Pas found that the sociodemographic variables of sex, income, and household status (i.e., head of household/not head of household) have high correlation to cluster membership. Day-to-day variation in trip-making attributes, such as scheduling and route choice, has also been explored with earlier datasets (see, for instance, Hatcher and Mahmassani, 1992).

More recently, researchers have worked to extend the foundation of theories built in the 1980s and 1990s. One survey that has allowed for further research on multiday behavior is the 6-week travel survey Mobidrive. Using this survey, collected in Karlsruhe and Halle/Salle, Schlich, and Axhausen, it was found that behavior is neither completely variable nor completely habitual, and the amount of variability measured depends on the method of analysis (e.g., trip-based methods vs. time-budget methods) (Schlich and Axhausen, 2003). Additionally, these researchers found that weekdays are more stable and habitual than weekends, and suggest that a measurement period of at least 2 weeks is needed to capture the variability within an individual's behavior. Research has also investigated the spatial repetitiveness of locations for discretionary activities during a one-week period (Schlich and Axhausen, 2003; Buliung et al., 2008). Buliung et al. (2008) found in a 1-week activity and travel diary in Toronto that weekday to weekend variation and day-to-day variation does exist among activity locations for several activity types and multiple travel modes. They likewise comment on the limitations of a short-timed travel survey and its usefulness. Schlich and Axhausen (2003) discuss the analysis of the 6-week Mobidrive data, finding that while there are between two and four locations that account for about 70% of locations and individual visits over a 6-week period, and about 90% of trips made by a person were to one of eight destinations, there were some instances where over 60 locations were recorded as destinations for an individual. This may or may not be a consistent finding across cultures; however, there has been no survey of comparable length and objective in the Unites States to date. Additionally, Cherchi and Cirillo use the 6-week Mobidrive survey to distinguish variability and habit in mode choice from day to day. In their work, they define two types of variation: planned and consequential variation. Planned variation is due to the daily or weekly activity patterns of an individual, whereas consequential variation is a result or consequence of either short- or long-term external changes

(Cherchi and Cirillo, 2010). Susilo and Kitamura (2004) use the Mobidrive dataset to analyze day-to-day variability in action space of urban residents. This is done by examining the second moments of activity locations (providing a metric of how the points are distributed in space). Stopher and Zhang (2010) use multiday GPS survey data collected in Australia to define 12 tour types and account for repetition in the tour types for an individual within the multiday survey period. Viti et al. (2010) combine traffic flow data (measured with inductor loops and pneumatic tubes) with survey data on behavior to examine the day-to-day and within-day variability of trip making in Ghent, Belgium. Day-to-day variability was also examined in Atlanta, Georgia, by Li et al. (2005), examining route choice, and Elango et al. (2007), who examine variability in trip attributes with respect to sociodemographic indicators.

When discussing variability in travel behavior, the role of planning and the role of habit must also be addressed. Hirsch et al. (1986) mention that an individual plans his or her activity pattern with consideration for those activities already conducted, and those that are planned for the future. More recently, several projects have focused on exploring the level of deliberation involved in the process of planning different activities and the role of habit. For instance, Mohammadian and Doherty (2005) discuss the dynamic nature of the scheduling process, which consists of preplanning, revisions, and impulsive and opportunistic decision-making. The study of scheduling and rescheduling has opened the door to a number of questions about the impromptu nature of planning and the cognitive effort involved in the process. For example, Chen et al. (2004) examine the rescheduling actions to determine whether the actual act of rescheduling is habitual or reasoned under normal disruptive circumstances (such as traffic congestion). However, although the reasoning involved with the act of planning activities has been studied, researchers have noted a lack of available information regarding the role of habitual activities on the planning process (Clarke and Doherty, 2008).

2.4 Description of the Data

The data for this research consists of a subset of records of the 2012 California Household Travel Survey. The travel survey comprised residents across the State of California, with a total sample of 42,500 households. All respondents completed a 1-day travel diary with accompanying sociodemographic information. A subset of the sample was selected to complete a GPS portion of the survey, which required the respondent to log trips using one of two types of GPS loggers. The first subsample consisted of respondents with an in-vehicle GPS logger, which was placed in the vehicle of a respondent and recorded for 7 days. A second subsample of the survey consisted of those who carried a wearable GPS data logger for a 3-day period. Additionally, the Metropolitan Transportation Commission (MTC) (covering the San Francisco Bay Area) funded an “add-on” portion of the survey, which consisted of a further sampling in this area of approximately 3,000 wearable GPS households (this add-on sample is included in the total of 42,500 households). For this research, a selection criterion was used to sample only those households within the Bay Area who were a part of the wearable GPS data collection. The California Household Travel Survey includes travel and activity details for each respondent summarizing the raw GPS data collected by the GPS loggers. In the first step of this analysis, these processed data were checked for errors or anomalies that might impact results. More information on the processing of GPS data into trips and activity locations is available in the California Household Travel Survey Final Report (NuStats Final Report, 2013). The processing of the trip and activity data provided by NuStats involved several steps, which resulted in the exclusion of five members of the original sample who were residents of the study area. First, a flag for onsite trips provided by NuStats helped eliminate trips that possibly occur in a single place. This, for instance, could be the existence of trips that were all a part of a respondent's golf game, or movement that was across a large parcel like a farm; both of these examples would not be considered trips. Four out of the five respondents excluded were omitted because all of their recorded trips were deemed to be onsite trips. The remaining respondent was omitted from the survey as an outlier due to an extremely large number of trips (134 trips in a two-day period) not attributable to any work behavior or other reasonable explanation.

The resulting sample size for the work presented in this report was 3,433 completed households, or 6,723 individuals. Individuals qualifying for the wearable GPS portion of the survey were between 16 and 75 years of age. Additional sample statistics are provided in Table 2-1 (individual level) and Table 2-2 (household level). Expansion weights provided in the processed survey data by NuStats were not utilized in this research. Trip-related statistics are provided in the next section of this chapter.

Table 2-1: Person-Level Sample Statistics (n=6,723)

Gender	
Female	3,442
Male	3,281
Age Distribution	
16-25	0.4%
26-50	10.3%
51-64	37.4%
64-75	12.2%
Missing	2.9%
Transportation Associated Indicators	
Have driver's license	6,202
Owns Transit pass	1,532

Car share program	164
Employment	
Full-time	4,664
Retired	685
Disabled	152
Home Duties	360
Student	489
Other	373
Work Location	
Fixed location	4,291
Home	308
No fixed location	42

Table 2-2: Household-Level Sample Statistics

Household Transportation Modes	
Number of vehicles	Median = 2 Min = 0 Max = 8
Number of bicycles	Median = 2 Min = 0 Max = 15
Household-Level Indicators	
Median household income	\$100,000–\$149,999
Number of members in the household	Mean = 3.04 Min = 1 Max = 8
Number of employees in the household	Mean = 1.64 Min = 0 Max = 5

Number of students in the household	Mean = .92 Min = 0 Max = 6
Persons with drivers licenses	Mean = 2.23 Min = 0 Max = 6

2.5 Analysis

In order to improve current modeling efforts, there are many aspects of day-to-day, intraperson travel variation that must be explored. A choice of destination has a variety of attributes that can be broken down and examined independently. In the research presented in the following sections, distinct attributes of a destination choice are examined. First, destinations are associated with the trip-making behavior enabling the choice and arrival at a destination. For this reason, it is pertinent to examine the changes in trip-level attributes from day to day. Changes from day to day in total trip length, total number of trips, average trip length, and standard deviation of trip length will be presented. In doing this, it is possible to understand the daily frequency of destinations chosen, and the distance for which a person travels to partake in an activity at that destination. Following this, a more sophisticated analysis using latent class cluster analysis (LCCA) is presented, whereby respondents are grouped based on similarities in change in travel attributes from day to day. These clusters are explored and described using these trip-level attributes of the destination. These clusters are then examined in a spatial context to better understand the creation of the clusters. Following this, a distance metric of activity locations is used to analyze cluster membership and changes from day to day. Lastly, destinations are analyzed using activity-type information to create a second latent class cluster model. This model, created by using changes in frequencies of different types of destinations, reveals further specifics on how people's travel varies from day to day. The latent class cluster model permits simultaneous analysis within person and across person differences in behavior.

Understanding Variation in Destination Choices Using Trip Attributes

Segmenting destination choices into trip-level attributes can uncover a large component of intraperson variation. This variation can be manifested simply in the number of destinations that an individual chooses per day (and therefore can be equated to trip frequency), but can also be manifested in the change in the distance traveled to reach those destinations, or the distribution of destinations across space.

Before conducting the analysis, it is important to understand the distribution and nature of trips recorded. During the survey period, the 6,723 respondents recorded 107,192 trips. There is quite a large distribution of distances of trip lengths that must be considered in the analysis. For instance, of all trips recorded, 1,182 trips exceeded 50 miles (approximately 1.1% of the total), which were made by a total of 626 respondents in the sample (approximately 9.3% of respondents). Of these trips over 50 miles, 2 trips were made on a Monday, 99 on a Tuesday, 237 on a Wednesday, 343 on a Thursday, 324 on a Friday, 173 on a Saturday, and 4 on a Sunday. In addition, when exploring the nature of these trips, 28 trips were flagged by NuStats as suspected to be work related (usually meaning they were reported as such in the CATI interview), and an additional 75 trips had work as an origin or destination but were not flagged as work related by NuStats. In total, 64 (approximately 10% of the respondents with trips over 50 miles) respondents had work-related—or work as origin or destination—trips that were over 50 miles. The distribution of trip lengths for those trips over 50 miles is provided in Figure 2-1 as a Cumulative Density Function.

Figure 2-1: CDF of Trips Over 50 Miles

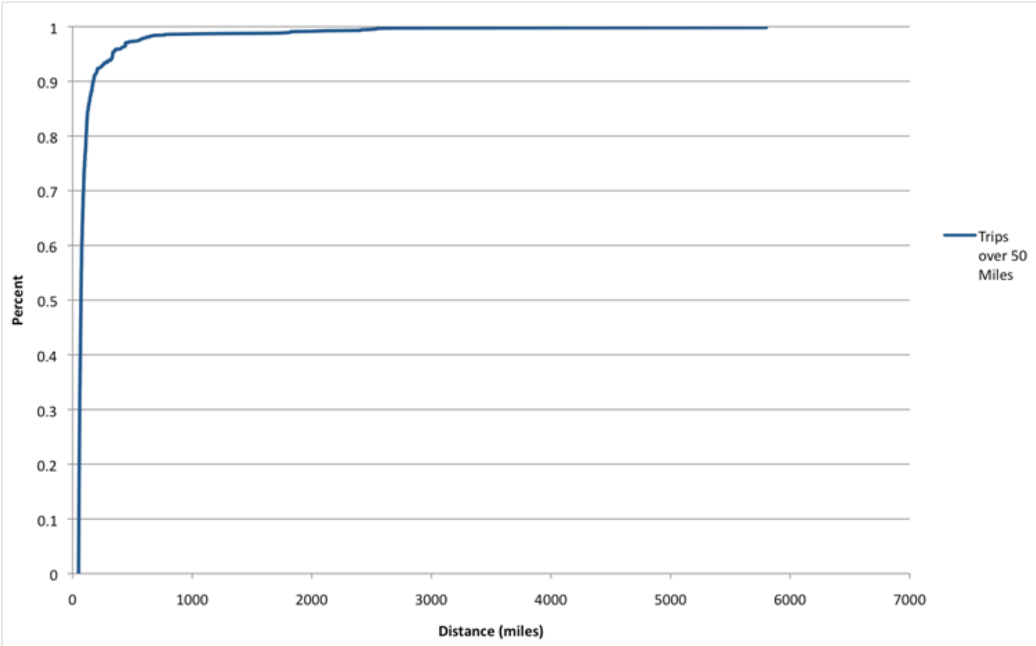
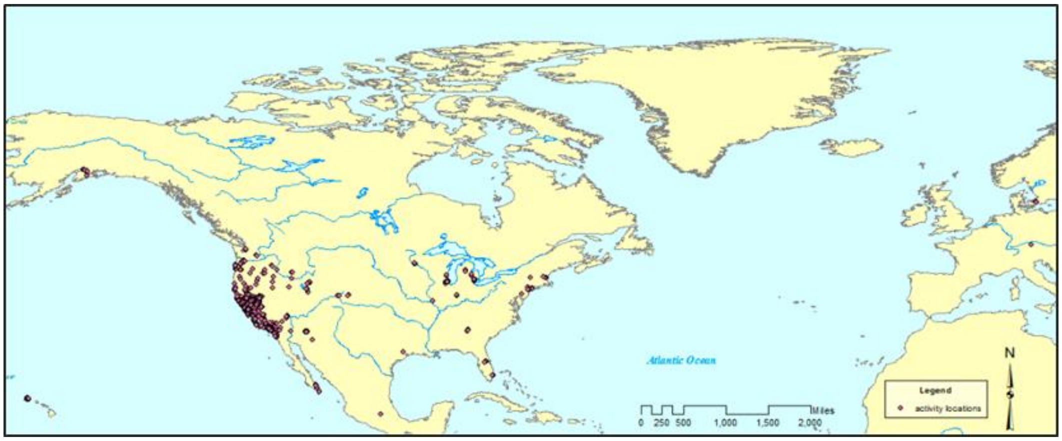


Figure 2-2 provides a map of all Bay Area residents' activity locations throughout the 3-day survey period. It is important to note that this figure and the first analysis presented in this report include all travel and activity locations. This includes long-distance trips, including Alaska, Mexico, Hawaii, and even European destinations as shown in Figure 2-2. The decision was made to retain all travel regardless of distance. A simple reason for this is to accurately display all types of multiday variation, which includes shorter-distance commutes to work as well as long-distance commutes, short-distance leisure travel, and long-distance leisure travel (and everything in-between). Although this first analysis examines all types of variation in travel, a second analysis reported later in this report specifically focuses on short-distance trips and activity locations.

Figure 2-2: Map of All Activity Locations of Survey Participants



Aggregate sample statistics of trip attributes are provided Table 2-3, Table 2-4, and Table 2-5. Table 2-3 provides summaries of attributes for all individuals for all three days of the survey, Table 2-4 provides individual trip averages for the aggregated sample, and Table 2-5 provides trip attributes reported for each day of the survey. It is important to note that this does not distinguish between weekdays and weekends.

Table 2-3: Trip Summaries for All Individuals Across all 3 Days (n=6,723)

Total Distance (Miles)	
Minimum	0.058899
Maximum	6,376.947
Average	94.41115
Median	62.36869
Total Number Of Trips	
Minimum	1
Maximum	88
Average	15.15068
Median	13
Total Duration (Minutes)	
Minimum	1.019999
Maximum	2,006.48
Average	211.282
Median	179.53
Total Modes Used	
Minimum	1

Maximum	7
Average	2.286033
Median	2

Table 2-4: Per Trip Averages for All Individuals

Per Trip Duration (Minutes)	
Average	13.94539
Per Trip Distance (Miles)	
Average	6.23148

Table 2-5: Per Day Trip Information

Total Per Day Distance (Miles)	
Average	31.47038
Median	15.9563
Total Per Day Duration (Minutes)	
Average	70.42735
Median	54.99
Total Per Day Trips	
Average	5.050226
Median	4
Number of Days with Recorded Trips	
1	887
2	1,661
3	4,175
Travel Day Included Weekend	
	1,521

To examine variation that is observed, combinations of all three days within the survey period were created for comparison. This resulted in three comparisons:

- Day 1 to Day 2
- Day 2 to Day 3
- Day 1 to Day 3

For each of these comparisons, changes across days were computed. Change was calculated as the absolute value of the difference from one day to another. Although a metric of change could have been achieved in other ways (such as taking the standard deviation) a direct comparison was chosen for interpretability. A distribution of these changes is provided in Table 2-6. As would be expected, Table 2-6 provides insight into the intraperson variation in day-to-day behavior and the interperson variation. For instance, while a majority of the sample has a change in trips of less than 20 trips from one day to the next, there is still quite a large variety in this change. For instance, 65% of the respondents in the sample have a change of between 2 and 11 trips for each day-to-day comparison. In addition, only 15% of the sample has no change in trips from one day to another for at least one combination of the 3 days of the sample, and another 15% have a change of one trip for any day-to-day comparison. Overall, 50% of the sample has a change of less than three trips for at least one of the three combinations of days. This means that the other 50% of respondents have a change in trip frequency from day to day of more than three trips. The 95th percentile of respondents shows large changes in trips (a change in between 10 and 35 trips from one day to the next).

Table 2-6: Distribution of Change Attributes for Day-to-Day Comparisons

	Change in total distance (mi.)			Change in total trips			Change in average distance (mi.)			Change in st. dev. trip distance (mi.)		
Percentiles	Day 1-2	Day 2-3	Day 3-1	Day 1-2	Day 2-3	Day 3-1	Day 1-2	Day 2-3	Day 3-1	Day 1-2	Day 2-3	Day 3-1
N	6,723	6,723	6,723	6,723	6,723	6,723	6,723	6,723	6,723	6,723	6,723	6,723
5	0.35	0.00	0.43	.00	.00	.00	0.08	0.00	0.10	0.02	0.00	0.01
10	1.03	0.02	1.35	.00	.00	.40	0.22	0.01	0.28	0.11	0.00	0.10
15	1.93	0.82	2.50	1.00	.00	1.00	0.38	0.18	0.47	0.23	0.05	0.23
20	2.91	1.88	3.70	1.00	1.00	1.00	0.54	0.36	0.68	0.37	0.18	0.40
25	3.96	3.02	4.96	1.00	1.00	1.00	0.74	0.59	0.92	0.52	0.35	0.59
30	5.26	4.44	6.34	2.00	1.00	2.00	0.95	0.81	1.17	0.69	0.53	0.80
35	6.54	6.06	8.23	2.00	2.00	2.00	1.18	1.09	1.46	0.88	0.74	1.07
40	8.28	7.84	10.17	2.00	2.00	2.00	1.50	1.42	1.81	1.11	1.01	1.34
45	10.17	9.95	12.23	2.00	2.00	3.00	1.78	1.77	2.19	1.37	1.34	1.65
50	12.26	12.36	14.70	3.00	3.00	3.00	2.12	2.17	2.62	1.69	1.68	2.07
55	14.70	14.92	17.53	3.00	3.00	4.00	2.54	2.61	3.12	2.06	2.13	2.62
60	17.64	17.97	21.03	4.00	4.00	4.00	3.02	3.19	3.72	2.57	2.68	3.21
65	20.97	22.00	24.93	4.00	4.00	5.00	3.62	3.86	4.38	3.20	3.34	3.88
70	25.20	26.51	29.47	5.00	5.00	5.00	4.40	4.66	5.24	3.90	4.17	4.70
75	30.54	32.64	35.66	5.00	5.00	6.00	5.35	5.80	6.33	4.92	5.33	5.77
80	37.43	40.33	42.74	6.00	6.00	7.00	6.72	7.31	7.72	6.33	7.02	7.30
85	46.94	51.18	52.15	7.00	7.00	7.00	8.59	9.47	9.87	8.32	9.47	9.72
90	62.47	70.50	69.12	8.00	8.00	9.00	11.76	12.95	13.04	11.76	13.02	12.97
95	101.24	112.71	105.69	10.00	11.00	11.00	18.87	20.77	20.78	18.82	21.05	20.83
100	2,763.70	6,246.41	6,297.80	34.00	34.00	35.00	923.77	2,093.69	2,093.10	1,566.28	3,203.21	3,211.60

When analyzing change in average trip distance, 25% of the sample have less than one mile change in the average distance per trip, possibly indicating high levels of similarity from day to day. However, 65% of the sample has changes in average trip distance between 1 and approximately 13 miles, indicating higher levels of variation in trip attributes. A portion of this is due to long-distance travel, as there are a number of trips that are long-distance (e.g., regional, national, or international trips) that increase the trip-distance averages. If there were many people with long-distance trips that are inflating the average trip distances, this should be reflected in high values of standard deviations. However, the distribution in the change of standard deviations of trip distances is similar to the change in average trip distance. In fact, 35% of the sample has a change in standard deviation that is less than one mile from one day to the next. Similarly, the 95th percentile of the sample includes change in distances in the thousands, indicating that many of these people have especially long-distance travel. In addition, the standard deviations for the 95th percentile are equally high, indicating that the difference between trips for a person is extreme. Changes in total distance traveled from day to day reveal similar findings. First, those with extreme long-distance trips can be identified, as the 95th percentile has changes of 2,763 (day one to day two), 6,246 (day two to day three), and 6,298 (day three compared to day one). Additionally, only the fifth percentile of individuals in the sample has changes of less than one mile for each of the 3-day combinations.

The aforementioned distributions are also provided in Figure 2-3 through Figure 2-10 as Cumulative Density Functions (CDFs). These CDFs are reported as pairs—one of the entire sample and a second showing the CDF through the 95th percentile—to minimize the impacts of extreme values on the ability to observe differences in the day comparisons through visualization.

Figure 2-3: CDF of Change in Total Trip Distance for Day Comparisons—Entire Sample

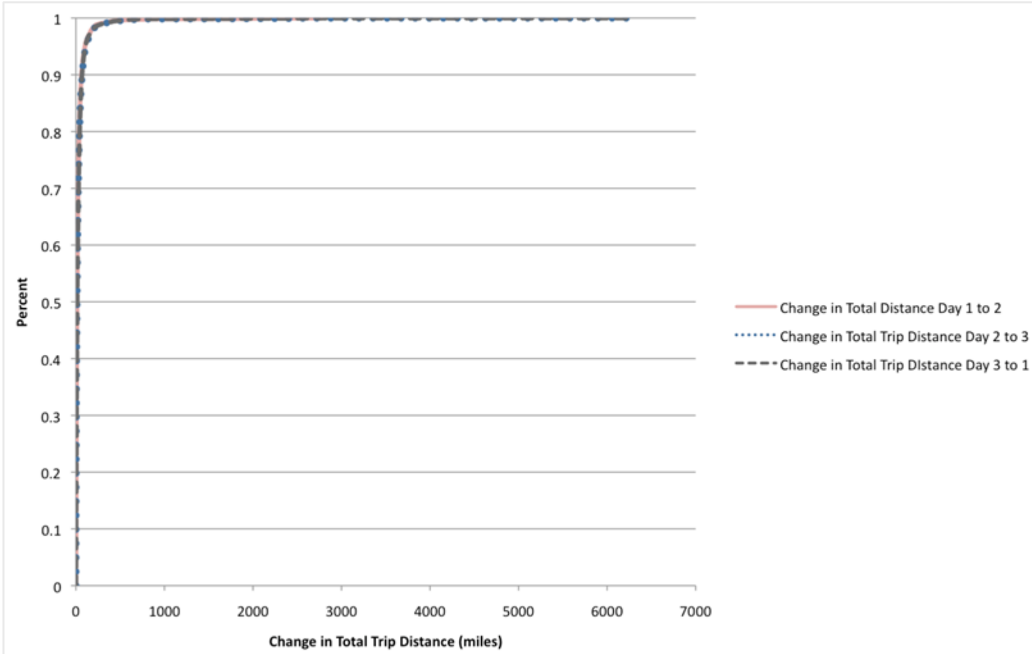


Figure 2-4: CDF of Change in Total Trip Distance for Day Comparisons—Lowest 95% of Sample

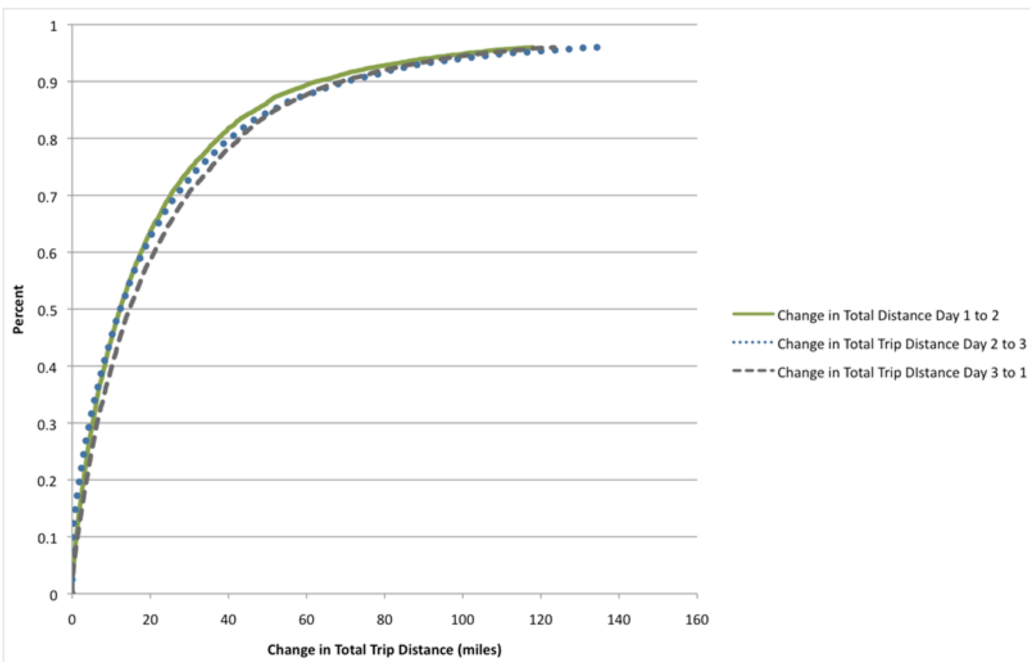


Figure 2-5: CDF of Change in Total Number of Trips for Day Comparisons—Entire Sample

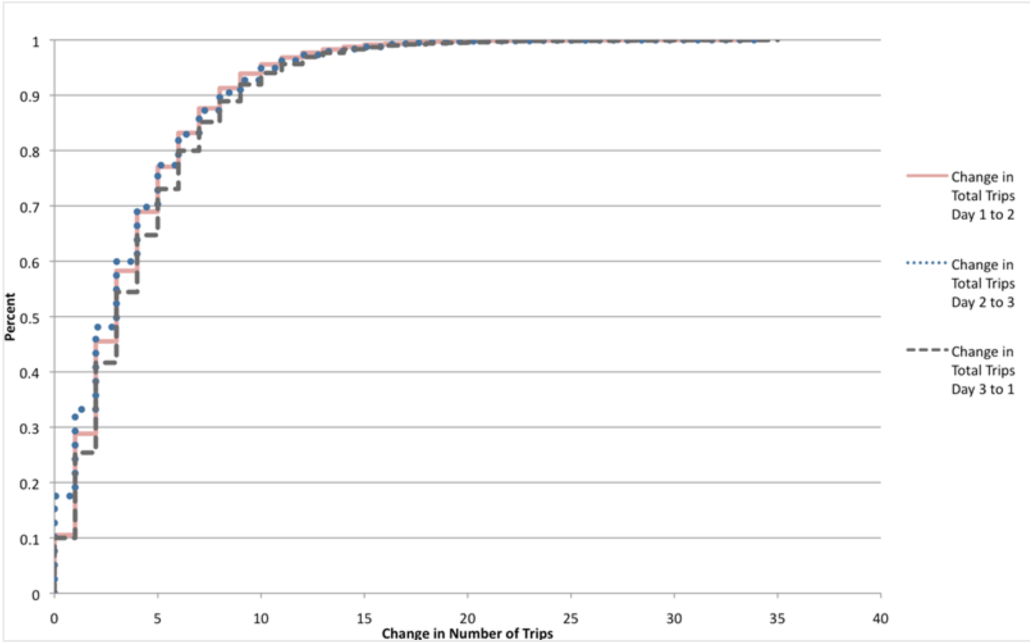


Figure 2-6: CDF of Change in Total Number of Trips for Day Comparisons—Lowest 95% of Sample

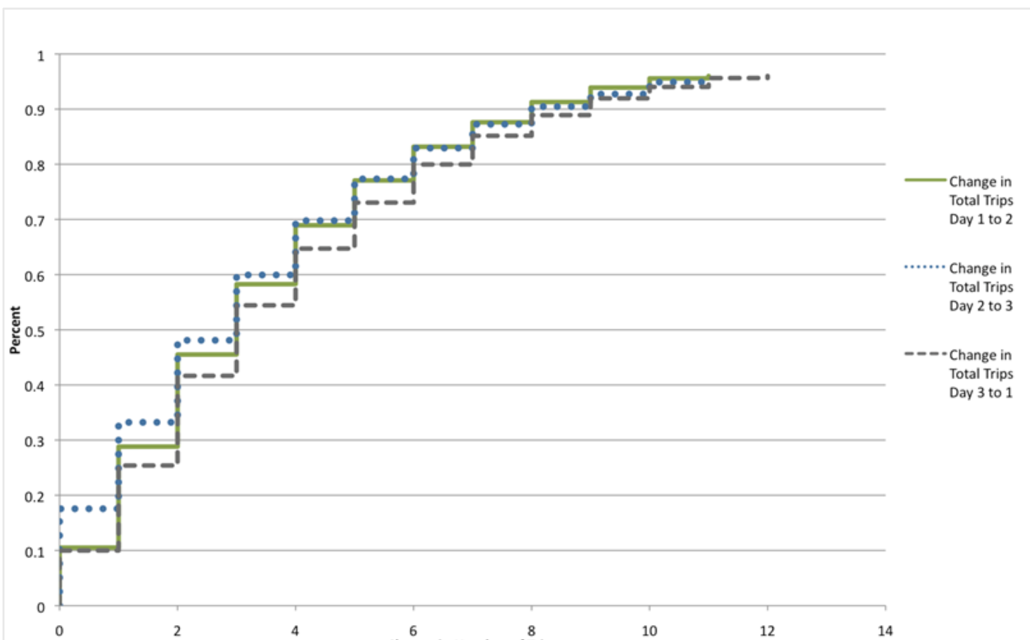


Figure 2-7: CDF of Change in Average Trip Distance for Day Comparisons—Entire Sample

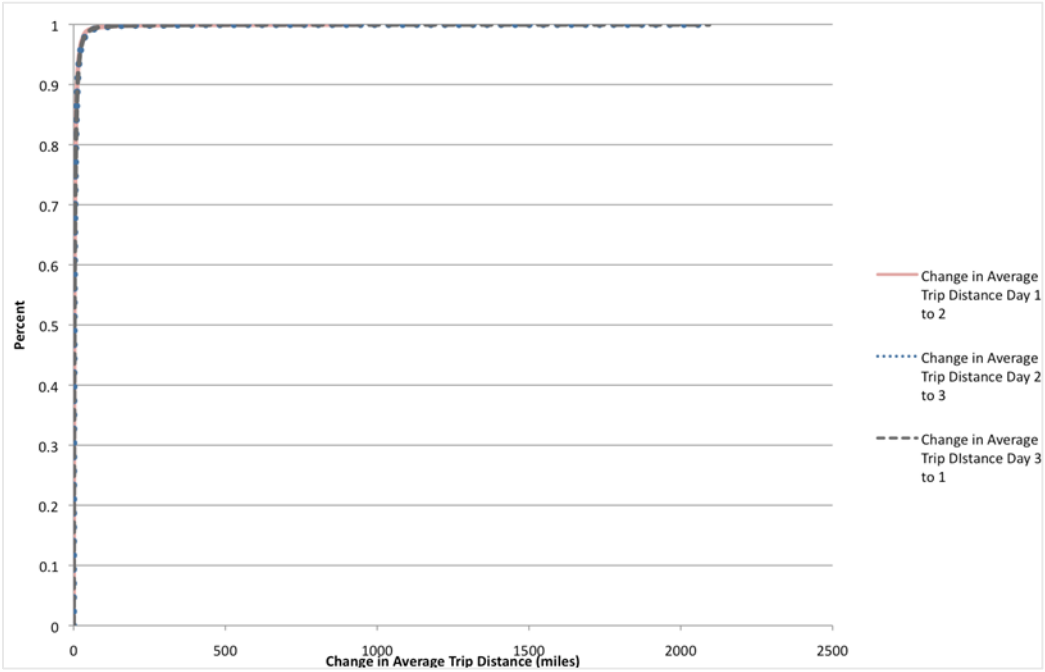


Figure 2-8: CDF of Change in Average Trip Distance for Day Comparisons—Lowest 95% of Sample

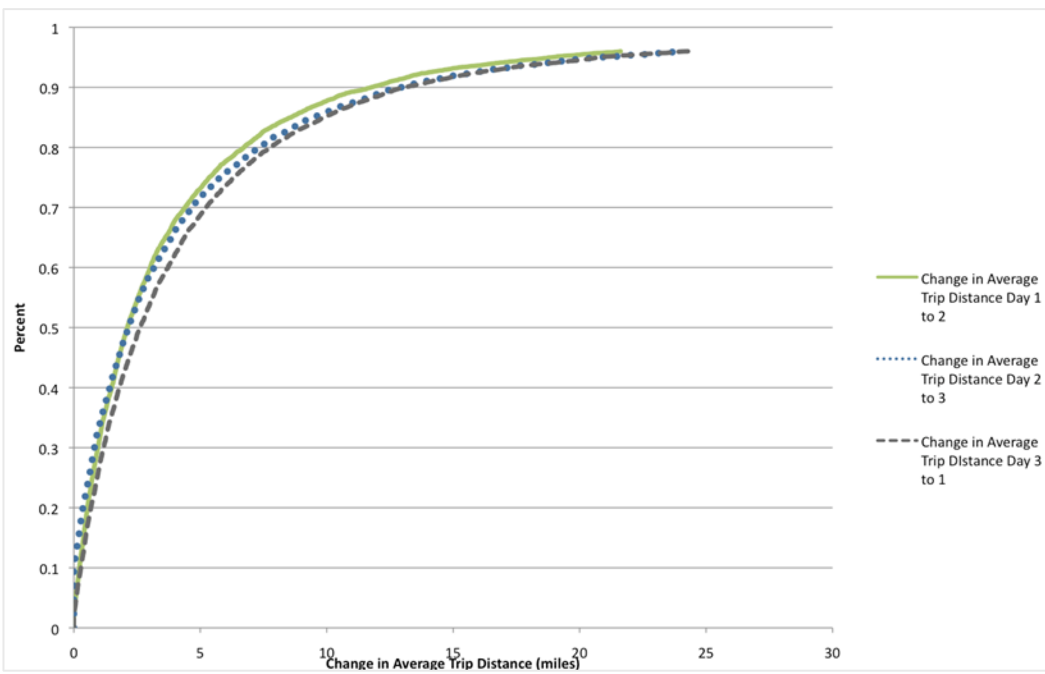


Figure 2-9: CDF of Change in Standard Deviation of Trip Distance for Day Comparisons—Entire Sample

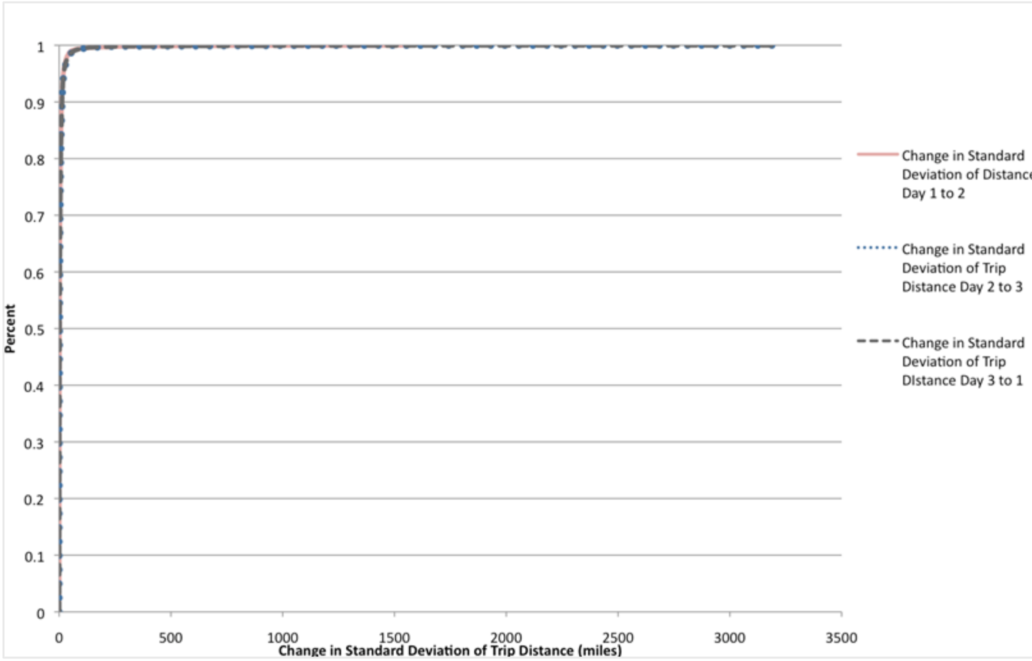
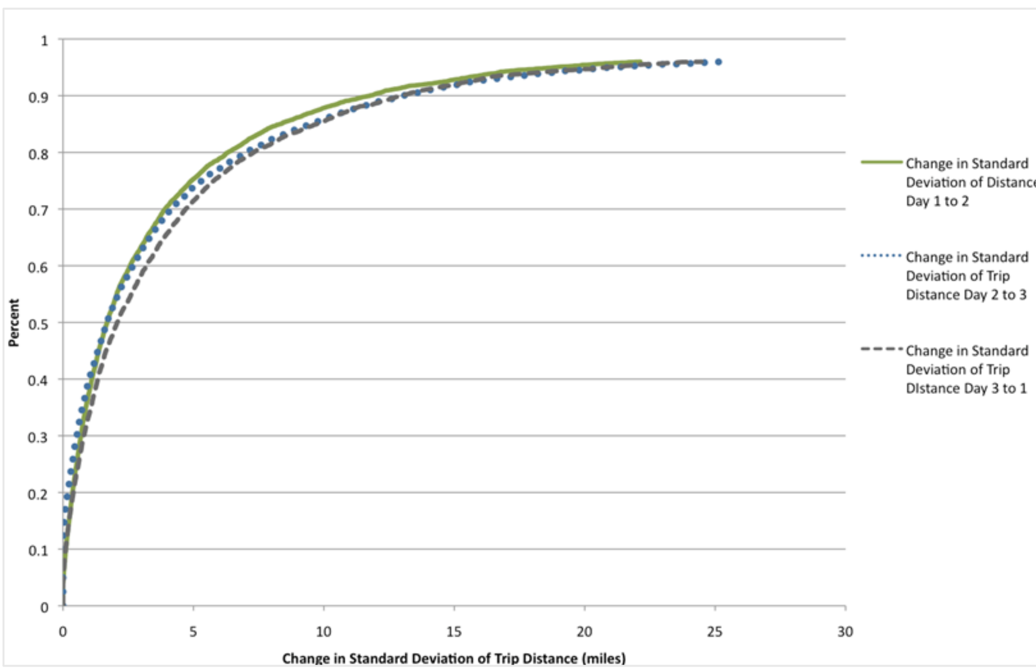


Figure 2-10: CDF of Change in Standard Deviation of Trip Distance for Day Comparisons—Lowest 95% of Sample



Although informative, a descriptive analysis of aggregate statistics provides only a limited amount of information on the variation of day-to-day travel behavior on a person level. For this reason, individual level similarities and differences in change of behavior from day to day are investigated using a LCCA. LCCA is used to cluster individuals or objects into groups or classes based on similarities. LCCA utilizes probabilistic methods in clustering. Though the most basic LCCA includes only continuous indicators, further model development has enabled models to now be estimated using both continuous and discrete variables (such as nominal or ordinal indicators). The equation used for LCCA with mixed indicator types is provided in Equation 1.

Equation 1: Equation Used for LCCA with Mixed Indicator Types

$$f(y_i \mid \theta) = \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} f_k(y_{ij} \mid \theta_{jk})$$

where

y_i is the person's response to the measured variables and $y_i|\theta$ is the distribution of y given the model parameter θ

K is the number of clusters

π_k is the prior probability of belonging to a latent class or cluster k

J is the total number of indicators

And y_{ij} is each element of y_i used to individually specify each univariate distribution

In addition to extending the model to include mixed indicator types, model development has also included the use of covariates. Covariates in this case are used as exogenous variables to predict class membership, as opposed to endogenous variables used to inform the development of clusters. Equation 2 provides the formulation of the inclusion of these covariates.

Equation 2: Formulation of the Inclusion of Covariates

$$f(y_i \mid z_i, \theta) = \sum_{k=1}^{K} \pi_{k|z_i} \prod_{j=1}^{J} f_k(y_{ij} \mid z_i, \theta_{jk})$$

where,

$\mathbf{z}_i$ is the values of the covariates for individual i, and the covariates are specified as having direct effects, avoiding the assumption that the covariates effect on the class membership only goes through the latent variable.

The development of clusters based on day-to-day variation was conducted using Latent Gold 4.5. Maximum Likelihood (ML) and Posterior Mode (PM) estimation procedures were used to estimate parameters. In addition, convergence was achieved using a two-step process: 1) Expectation Maximization (EM) was used to find a ML; and 2) Newton-Raphson (NR) methods were employed to iterate through a series of successively improved approximations (Vermunt and Magidson, 2005).

In order to arrive at an appropriate number of clusters, an iterative procedure was used. Models were estimated containing one cluster through eight clusters. Table 2-7 provides a description of indicators and covariates used in the model estimation. The model was estimated using all 6,723 respondents. Fit statistics, model parsimony, and cluster structure were analyzed and compared for each model to determine the most optimal solution. The results of this process revealed that a six-cluster model (LL = -238,776, BIC = 479,324.5, Classification error = .04) best described the latent phenomenon underlying the observed data.

Table 2-7: Variables for Latent Class Cluster Model

Variables	Description
Indicators for latent classes all changes are absolute values	
Day 1 to 2 change in total trip distance	Continuous values ranging from 0 to 2,763.70
Day 2 to 3 change in total trip distance	Continuous values ranging from 0 to 6,246.41
Day 1 to 3 change in total trip distance	Continuous values ranging from 0 to 6,297.80
Day 1 to 2 change in total trips	Count values ranging from 0 to 34
Day 2 to 3 change in total trips	Count values ranging from 0 to 34
Day 1 to 3 change in total trips	Count values ranging from 0 to 35
Day 1 to 2 change in average trip distance	Continuous values ranging from 0 to 973.77
Day 2 to 3 change in average trip distance	Continuous values ranging from 0 to 2,093.69
Day 1 to 3 change in average trip distance	Continuous values ranging from 0 to 2,093.10
Day 1 to 2 change in total distance	Continuous values ranging from 0 to 1,566.28
Day 2 to 3 change in total distance	Continuous values ranging from 0 to 3,203.21
Day 1 to 3 change in total distance	Continuous values ranging from 0 to 3,211.60
Covariates for class membership prediction	
Number of days where no travel was recorded	Values of 0, 1, and 2 days
Indicator for female	Binary indicator, 1 if female 0 if male
Indicator for age group 26 through 50	Binary indicator, 1 if within age group
Indicator for age group 51 through 64	Binary indicator, 1 if within age group
Indicator for age group 65 and older	Binary indicator, 1 if within age group
Employed	Binary indicator, 1 if employed
Retired	Binary indicator, 1 if retired
Indicator for no fixed work location	Binary indicator, 1 if no fixed location for work

Variables	Description
Day one of survey was a work day	Binary indicator if day 1 was a work day
Day two of survey was a work day	Binary indicator if day 2 was a work day
Day three of survey was a work day	Binary indicator if day 3 was a work day
Indicator for middle income household	Binary indicator if income is between $35,000-$99,999
Indicator for high-income household	Binary indicator if income is $100,000 or more
Number of members in the household	Count variable ranging from 1 to 8

The results of the six-cluster model are provided in a series of tables. Table 2-8 provides the profile of each cluster, reporting mean values for each cluster for all indicators and covariates. Table 2-9 provides the probability means of the model, indicating the likelihood for respondents with values or value ranges to be members of each cluster. Table 2-10 provides the coefficient values and significance for covariates of the model.

Table 2-8: Profile of Six Cluster Models

			Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
	Cluster Size		0.30	0.25	0.24	0.11	0.08	0.02
Indicators	Change in total dist. day 1-2 Mean		14.58	27.52	4.52	79.66	34.55	429.11
	Change in total dist. day 2-3 Mean		16.90	35.99	4.30	88.15	0.49	543.82
	Change in total dist. day 1-3 Mean		16.17	33.75	4.81	82.07	34.61	497.66
	Change in total trips day 1-2 Mean		3.93	3.49	2.45	4.25	6.62	4.37
	Change in total trips day 2-3 Mean		4.67	4.28	2.23	4.59	0.17	4.21
	Change in total trips day 1-3 Mean		4.41	3.97	2.63	4.76	6.57	4.94
	Change in avg. trip dist. day 1-2		2.06	5.35	0.90	15.40	5.78	82.63
	Change in avg. trip dist. day 2-3		2.54	6.92	0.86	17.17	0.13	108.66
	Change in avg. trip dist. day 1-3		2.45	6.60	0.96	17.02	5.79	99.12
	Chg. in st. dev. trip dist. day 1-2		1.89	5.14	0.67	15.32	5.21	147.87
	Chg. in st. dev. trip dist. day 2-3		2.24	6.63	0.66	17.40	0.11	184.35
	Chg. in st. dev. trip dist. day 1-3		2.16	6.34	0.71	16.56	5.21	173.12
Covariates	Days with no travel	0	0.69	0.66	0.64	0.62	0.13	0.67
		1	0.27	0.30	0.19	0.31	0.08	0.27
		2	0.04	0.03	0.17	0.07	0.78	0.07
		Mean	0.35	0.37	0.52	0.45	1.65	0.40
	Female	0	0.46	0.51	0.46	0.54	0.51	0.56
		1	0.54	0.49	0.54	0.46	0.49	0.44
	Age 26- 50	0	0.61	0.63	0.67	0.62	0.63	0.61
		1	0.39	0.37	0.33	0.38	0.37	0.39
	Age 51- 64	0	0.64	0.58	0.66	0.59	0.65	0.60
		1	0.36	0.42	0.34	0.41	0.35	0.40
	Age 65 and older	0	0.86	0.88	0.83	0.89	0.87	0.87
		1	0.14	0.12	0.17	0.11	0.13	0.13
	Employed	0	0.32	0.23	0.38	0.28	0.32	0.24
		1	0.68	0.77	0.62	0.72	0.68	0.76
	Retired	0	0.89	0.92	0.89	0.88	0.90	0.93
		1	0.11	0.08	0.11	0.12	0.10	0.07
	No fixed work location	0	1.00	0.99	0.99	1.00	0.99	0.96
		1	0.00	0.01	0.01	0.00	0.01	0.04

		Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
Travel day one was a work day	0	0.43	0.36	0.49	0.41	0.43	0.34
	1	0.57	0.64	0.51	0.59	0.57	0.66
Travel day two was a work day	0	0.45	0.38	0.55	0.48	0.87	0.40
	1	0.55	0.62	0.45	0.52	0.13	0.60
Travel day three was a work day	0	0.62	0.65	0.62	0.70	0.88	0.59
	1	0.38	0.35	0.38	0.30	0.12	0.41
Income $35,000-$99,999	0	0.65	0.68	0.62	0.67	0.63	0.70
	1	0.35	0.32	0.38	0.33	0.37	0.30
Income 100,000 or higher	0	0.52	0.46	0.57	0.45	0.51	0.42
	1	0.48	0.54	0.43	0.55	0.49	0.58
Number of household members	1	0.09	0.09	0.12	0.06	0.08	0.09
	2	0.31	0.35	0.32	0.34	0.28	0.39
	3	0.22	0.22	0.20	0.25	0.21	0.24
	4	0.25	0.22	0.23	0.24	0.28	0.20
	5 to 8	0.13	0.12	0.13	0.11	0.15	0.09
	Mean	3.10	2.98	3.01	3.06	3.20	2.83

Table 2-9: Cluster Probability Means

			Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
		Overall	0.30	0.25	0.24	0.11	0.08	0.02
Indicators	Change in total distance day 1 to 2	0 - 2.912	0.24	0.17	0.53	0.04	0.00	0.01
		2.915 - 8.274	0.29	0.16	0.48	0.05	0.02	0.00
		8.283 - 17.64	0.43	0.21	0.21	0.05	0.09	0.01
		17.65 - 37.43	0.45	0.34	0.00	0.06	0.14	0.01
		37.43 - 2764	0.07	0.38	0.00	0.36	0.14	0.05
	Change in total distance day 2 to 3	0 - 1.872	0.10	0.04	0.45	0.05	0.36	0.01
		1.881 - 7.838	0.27	0.13	0.53	0.03	0.03	0.00
		7.846 - 17.96	0.52	0.20	0.24	0.04	0.00	0.00
		17.99 - 40.32	0.53	0.40	0.00	0.06	0.00	0.00
		40.34 - 6246	0.06	0.49	0.00	0.38	0.00	0.06
	Change in total distance day 3 to 1	0 - 3.693	0.22	0.13	0.61	0.04	0.00	0.00
		3.702 - 10.17	0.33	0.13	0.46	0.04	0.04	0.01
		10.18 - 21.02	0.48	0.19	0.16	0.05	0.11	0.00
		21.05 - 42.73	0.41	0.38	0.00	0.08	0.13	0.00
		42.78 - 6298	0.05	0.42	0.00	0.36	0.11	0.06
	Change in total trips day 1 to 2	0 - 1	0.27	0.27	0.33	0.10	0.01	0.01
		2	0.27	0.23	0.33	0.10	0.05	0.02
		3	0.29	0.28	0.26	0.11	0.06	0.01
		4 to 5	0.33	0.25	0.21	0.11	0.08	0.02
		6 to 34	0.32	0.23	0.10	0.14	0.19	0.02
	Change in total trips day 2 to 3	0 - 0	0.11	0.09	0.32	0.07	0.40	0.01
		1	0.28	0.23	0.35	0.09	0.03	0.02
		2 to 3	0.28	0.28	0.29	0.12	0.01	0.01
		4 to 5	0.35	0.31	0.21	0.11	0.00	0.01
		6 to 34	0.43	0.30	0.09	0.15	0.00	0.02
	Change in total trips day 3 to 1	0 - 1	0.27	0.25	0.35	0.10	0.02	0.01
		2	0.29	0.23	0.31	0.10	0.05	0.01
		3 to 4	0.28	0.27	0.26	0.10	0.07	0.02
		5 to 6	0.29	0.26	0.20	0.11	0.11	0.02
		7 to 35	0.36	0.23	0.08	0.14	0.17	0.02
	Change in average trip distance day 1 to 2	0 - 0.536	0.31	0.13	0.51	0.04	0.01	0.01
		0.536 - 1.498	0.35	0.13	0.46	0.03	0.02	0.01
		1.499 - 3.021	0.45	0.19	0.23	0.05	0.07	0.00
		3.021 - 6.721	0.35	0.38	0.02	0.06	0.18	0.01
		6.728 - 923.8	0.02	0.43	0.00	0.38	0.12	0.05
	Change in average trip	0 - 0.361	0.13	0.03	0.45	0.04	0.34	0.01

			Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
	distance day 2 to 3	0.361 - 1.420	0.33	0.10	0.50	0.02	0.05	0.00
		1.421 - 3.194	0.55	0.14	0.26	0.04	0.01	0.01
		3.196 - 7.314	0.46	0.47	0.02	0.05	0.00	0.00
		7.317 - 2094	0.02	0.52	0.00	0.40	0.00	0.06
	Change in average trip distance day 3 to 1	0 - 0.675	0.28	0.12	0.57	0.03	0.01	0.00
		0.675 - 1.808	0.36	0.10	0.47	0.04	0.03	0.00
		1.808 - 3.724	0.50	0.17	0.18	0.05	0.10	0.00
		3.725 - 7.711	0.34	0.41	0.01	0.06	0.18	0.01
		7.737 - 2093	0.01	0.46	0.00	0.38	0.09	0.06
	Change in standard deviation of trip distances day 1 to 2	0 - 0.365	0.26	0.14	0.52	0.05	0.03	0.01
		0.366 - 1.108	0.34	0.13	0.44	0.05	0.03	0.01
		1.109 - 2.572	0.44	0.19	0.25	0.05	0.06	0.01
		2.574 - 6.330	0.43	0.34	0.01	0.05	0.16	0.00
		6.330 - 1566	0.02	0.45	0.00	0.37	0.12	0.05
	Change in standard deviation of trip distances day 2 to 3	0 - 0.180	0.12	0.06	0.41	0.05	0.35	0.01
		0.180 - 1.005	0.30	0.11	0.51	0.04	0.03	0.00
		1.007 - 2.677	0.51	0.15	0.28	0.04	0.01	0.00
		2.677 - 7.018	0.54	0.38	0.02	0.05	0.00	0.00
		7.026 - 3203	0.01	0.55	0.00	0.38	0.00	0.06
	Change in standard deviation of trip distances day 3 to 1	0 - 0.396	0.25	0.13	0.54	0.06	0.02	0.00
		0.396 - 1.335	0.33	0.11	0.47	0.04	0.04	0.00
		1.336 - 3.211	0.51	0.14	0.21	0.05	0.08	0.00
		3.213 - 7.302	0.39	0.39	0.01	0.06	0.16	0.00
		7.315 - 3212	0.00	0.49	0.00	0.35	0.09	0.06
Covariates	Days with no travel	0	0.33	0.27	0.25	0.11	0.02	0.02
		1	0.33	0.31	0.18	0.14	0.03	0.02
		2	0.08	0.06	0.31	0.06	0.47	0.01
	Female	0	0.28	0.26	0.23	0.13	0.08	0.02
		1	0.31	0.24	0.26	0.10	0.08	0.01
	Age 26 to 50	0	0.29	0.25	0.26	0.11	0.08	0.02
		1	0.31	0.25	0.22	0.11	0.08	0.02
	Age 51 to 64	0	0.31	0.23	0.26	0.11	0.08	0.02
		1	0.28	0.28	0.22	0.12	0.07	0.02
	Age 65 and older	0	0.30	0.26	0.24	0.12	0.08	0.02
		1	0.30	0.22	0.30	0.09	0.07	0.01
	Employed full time	0	0.31	0.19	0.30	0.10	0.08	0.01
		1	0.29	0.28	0.22	0.12	0.08	0.02
	Retired	0	0.30	0.26	0.24	0.11	0.08	0.02
		1	0.31	0.21	0.26	0.13	0.08	0.01
	No fixed work location	0	0.30	0.25	0.24	0.11	0.08	0.02
		1	0.21	0.22	0.28	0.05	0.15	0.10
	Day 1 was a workday	0	0.30	0.21	0.28	0.11	0.08	0.01
		1	0.29	0.28	0.22	0.11	0.08	0.02
	Day 2 was a workday	0	0.27	0.19	0.27	0.11	0.14	0.01
		1	0.32	0.31	0.22	0.11	0.02	0.02
	Day 3 was a workday	0	0.28	0.25	0.23	0.12	0.11	0.01
		1	0.33	0.26	0.27	0.10	0.03	0.02
	Income $35,000-$99,999	0	0.30	0.26	0.23	0.11	0.08	0.02
		1	0.30	0.23	0.26	0.11	0.09	0.01
	Income 100,000 or higher	0	0.31	0.23	0.27	0.10	0.08	0.01
		1	0.29	0.27	0.22	0.12	0.08	0.02
	Number of household members	1	0.29	0.24	0.32	0.07	0.07	0.01
		2	0.28	0.27	0.24	0.12	0.07	0.02
		3	0.30	0.25	0.23	0.13	0.08	0.02
		4	0.31	0.23	0.23	0.11	0.09	0.01
		5 to 8	0.30	0.24	0.26	0.10	0.09	0.01

Table 2-10: Coefficients and Significance for Covariates

Covariates		Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Wald	p-value
Days with no travel		-0.542	-0.506	-0.094	-0.357	1.841	-0.344	314.072	0.000
Female	0	-0.084	-0.013	-0.086	0.082	-0.011	0.111	20.574	0.001
	1	0.084	0.013	0.086	-0.082	0.011	-0.111		
Age 26 to 50	0	-0.027	-0.030	0.184	-0.011	0.033	-0.150	17.701	0.003
	1	0.027	0.030	-0.184	0.011	-0.033	0.150		
Age 51 to 64	0	0.001	-0.094	0.187	-0.041	0.062	-0.115	20.174	0.001
	1	-0.001	0.094	-0.187	0.041	-0.062	0.115		
Age 65 and older	0	-0.047	-0.047	0.028	0.163	0.072	-0.170	5.056	0.410
	1	0.047	0.047	-0.028	-0.163	-0.072	0.170		
Employed full time	0	0.039	-0.238	0.208	-0.290	0.158	0.122	53.430	0.000
	1	-0.039	0.238	-0.208	0.290	-0.158	-0.122		
Retired	0	-0.009	-0.065	0.131	-0.277	-0.039	0.259	19.707	0.001
	1	0.009	0.065	-0.131	0.277	0.039	-0.259		
No fixed work location	0	0.276	0.236	0.005	0.336	-0.047	-0.806	14.710	0.012
	1	-0.276	-0.236	-0.005	-0.336	0.047	0.806		
Day 1 was a workday	0	0.175	0.147	0.068	0.088	-0.452	-0.026	41.498	0.000
	1	-0.175	-0.147	-0.068	-0.088	0.452	0.026		
Day 2 was a workday	0	-0.170	-0.192	0.028	0.070	0.329	-0.065	42.443	0.000
	1	0.170	0.192	-0.028	-0.070	-0.329	0.065		
Day 3 was a workday	0	0.023	0.229	-0.252	0.213	-0.234	0.021	117.455	0.000
	1	-0.023	-0.229	0.252	-0.213	0.234	-0.021		
Income $35,000-$99,999	0	0.082	0.054	0.073	-0.080	-0.098	-0.031	8.534	0.130
	1	-0.082	-0.054	-0.073	0.080	0.098	0.031		
Income 100,000 or higher	0	0.144	0.022	0.192	-0.110	-0.088	-0.160	31.487	0.000
	1	-0.144	-0.022	-0.192	0.110	0.088	0.160		
Number of household members		0.059	0.018	-0.024	0.052	0.018	-0.123	11.411	0.044

By interpreting the preceding tables, a description of cluster composition (understanding the impact of indicators that were used to form clusters) and cluster membership (understanding the role of covariates in cluster membership) is achieved. The six clusters that result from the LCCA reveal some striking differences in variation across the members of the sample. These differences can be seen in a qualitative sense in Table 2-11, which presents a simple graduated color scheme representing the mean value of each indicator by cluster. Light colors correspond to low values for cluster means, and dark values correspond to high values. Comparing across indicators allows for the comparison of cluster means. For instance, it is apparent that cluster five stands out in the mean value of change for total distance from day two to day three. Comparing across days for each type of change allows for the comparison of change within each cluster. Color should only be interpreted across clusters and day comparisons for the same indicator, not across indicator types (i.e., total distance or total trips). Numerical values for the cluster means for each attribute have been included for reference.

Table 2-11: Quantitative and Qualitative Comparison of Mean Values for Cluster Indicators

Indicator	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Change in total dist. day 1-2 Mean	14.58	27.52	4.52	79.66	34.55	429.11
Change in total dist. day 2-3 Mean	16.90	35.99	4.30	88.15	0.49	543.82
Change in total dist. day 1-3 Mean	16.17	33.75	4.81	82.07	34.61	497.66
Change in total trips day 1-2 Mean	3.93	3.49	2.45	4.25	6.62	4.37
Change in total trips day 2-3 Mean	4.67	4.28	2.23	4.59	0.17	4.21
Change in total trips day 1-3 Mean	4.41	3.97	2.63	4.76	6.57	4.94
Change in avg. trip dist. day 1-2 Mean	2.06	5.35	0.90	15.40	5.78	82.63
Change in avg. trip dist. day 2-3 Mean	2.54	6.92	0.86	17.17	0.13	108.66
Change in avg. trip dist. day 1-3 Mean	2.45	6.60	0.96	17.02	5.79	99.12
Chg. in st. dev. trip dist. day 1-2 Mean	1.89	5.14	0.67	15.32	5.21	147.87
Chg. in st. dev. trip dist. day 2-3 Mean	2.24	6.63	0.66	17.40	0.11	184.35
Chg. in st. dev. trip dist. day 1-3 Mean	2.16	6.34	0.71	16.56	5.21	173.12

COLOR SCALE low — high

Cluster 1: Mid-Variation, Local Trip Makers

Cluster composition. Cluster 1 comprises people who have changes in total distance that range from the middle to higher end of respondents (between approximately 8 and 38 miles). There are still some members of cluster 1 who have lower changes in total distance, although it is less likely. The average change in trips is 3.93 for day 1 to day 2, 4.67 for day 2 to 3 and 4.41 for day 3, compared to day 1. These changes vary from day comparison to day comparison, indicating a small level of variability within this cluster in the change. There is a change from day to day of slightly less than 4 trips from day 1 to day 2, and an approximately 4.5 trip change for day 2 to 3 or day 1 to 3. In other words, while there is variation in trip frequency for all days, there is slightly more for some days than others. Additionally, when examining the probability means, it is notable that there is an approximately even probability for all change categories (although more so for the change from day 1 and day 2 than from day 2 to day 3). The cluster mean for change in trips from day to day falls within the middle of cluster means when comparing across clusters. Similarly, the changes in standard deviations from day to day are also in the middle of the range for changes across all clusters. This indicates that although there is some change in the distribution of distance of trips within a day, the change is neither the most extreme nor the least extreme.

Cluster members. Individuals belonging to cluster 1 are less likely to have days in which there was no travel. However, some individuals have one day that was not a travel day in this cluster. Across all members of this cluster, the average number of days for no travel is 35 days. Females are slightly more likely to belong to cluster 1, although the coefficient indicates that this is not large. Those without a fixed location for work are less likely to belong to this cluster. Those who work on day 1 of the survey are less likely to be members of cluster 1, whereas working on day two has a positive impact on membership. Though many of the other covariates are significant in the model, the impact that they have on membership in cluster 1 is small.

Cluster 2: High-Variation, Longer-Distance Local Trip Makers (Local Venturers)

Cluster composition. Cluster 2 is composed of members who have higher changes in total distance than cluster 1. In fact, the means for change in distance are twice as much as cluster 1 (27.52 for day 1 to 2, 35.99 for day 2 to 3, and 33.75 for day 3 to 1). Although these members have higher changes for total distance, the change in number of trips from day to day is lower than that of cluster 1. This indicates that members of cluster 2 have a larger change in trip distances from day to day, which is reflected in the change in average distance indicators. When comparing the mean changes in average distance, it is evident that cluster 2 has neither the lowest change in average distance nor the highest. However, considering the existence of longer-distance travel, it is possible that cluster 2 contains members that travel longer distances locally, rather than any regional or longer-distance travel. In addition, the changes in standard deviations of travel distances are also within the middle of the ranges of changes for all clusters.

Cluster members. Cluster 2 members have a higher tendency to travel on all 3 days. The mean for cluster 2 for the number of days in which no travel was recorded is .37 days. Cluster 2 has a higher frequency of those who are employed full time and had work on day 2 of the travel day. However, those who did not have work on day 1 and day 3 are more likely to be members of cluster 2. Additionally, those who have no fixed work location are less likely to be members of cluster 2.

Cluster 3: Low-Variation, Local Trip Makers

Cluster composition. Cluster 3 members have small changes in total distance from day to day. Additionally, cluster 3 members also have a smaller change in number of trips from day to day (cluster means of 2.45, 2.23, and 2.63 for days 1 to 2, 2 to 3, and 1 to 2, respectively). Cluster 3 has the lowest numbers across all clusters in the change across days in average trip distance. All three comparison day pairs have a change in average distances of less than one mile. This indicates that the trips added or subtracted from travel across days are likely similar in distance to the other trips. Cluster 3 also has the lowest mean for the change in standard deviation of trips, indicating that there is less variability across days in the length distribution of trips within a day.

Cluster members. Females are slightly more likely to be members of cluster 3 compared to other clusters. Cluster 3 also has some of the highest age related coefficients across all clusters. Respondents within the age of 26 to 50 and 51 to 64 are less likely to be members of cluster 3. Those respondents who are retired are also less likely to be members of cluster 3. Additionally, those who are employed are less likely to be members of cluster 3. For those members who work, working on day 3 has a positive effect on membership in cluster 3, and work on day 1 and 2 have only a slight negative impact.

Cluster 4: High-Variation, Mid-Distance Trip Makers

Cluster composition. Cluster 4 has a high change in total distance from day to day (mean values of 79.66, 88.15, and 82.07 for days 1 to 2, 2 to 3, and 1 to 3). Most members of this cluster are within the highest category for change in total distance from day to day, but are likely not the individuals in the sample with the longest-distance travel (international travel). These individuals are perhaps those with national or longer regional travel episodes. This is also supported by the fact that the mean changes in total trips for day to day in cluster 4 are higher than clusters 1, 2, 3, and similar to cluster 6. The changes in average trip distance across days for members of cluster 4 are also higher than all clusters except for cluster 6. Lastly, members of cluster 4 have large changes in standard deviations for trip distance from day to day. This further supports the hypothesis that cluster 4 likely contains those with mid-length, long-distance travel. However, this hypothesis needs to be further explored to confirm trip attributes.

Cluster members. Cluster 4 members, like clusters 1 and 2, are more likely to have traveled on all 3 days, or at least 2 out of 3 days. Cluster 4 members are also more likely to be employed full time or retired. Although a full-time employment status is possibly different day in and day out from a retired person, there are some striking similarities that could explain this result. For instance, if cluster 4 members are indeed regional long-distance travelers, this could be a result of retired pleasure trips or business trips.

Cluster 5: Extreme-Variation, One- or Two-Day Nontravelers

Cluster composition. The trends that are observed in cluster 5 are unique. Cluster 5 members show extreme variability across days. For instance, the cluster means for change in total distance vary from some of the higher values (in the mid-30s) to the lowest (0.49 miles). Similarly, this trend is found in the change of total number of trips (ranging from a mean change of 6.62 trips for day 1 to 2, to a mean change of 0.17 trips for day 2 to 3 across all cluster members). The changes in day to day for average distance per trip and standard deviation are similar to the findings for total distance and total trips. When comparing the larger change values to other clusters, cluster 5 compares most closely with cluster 2.

Cluster members. Cluster 5 is the cluster with the highest likelihood of membership for respondents who have days without recorded travel. The cluster mean for days of nontravel is 1.84, reflecting the high percentage of individuals who have one or 2 days during which there is no travel. Those in the sample who are employed full time are less likely to belong to cluster 5. For those who do work and belong to cluster 5, work on day one and day three are likely to be in cluster 5. However, work on the second day of the survey has a negative effect on cluster 5 membership.

Cluster 6: High-Variation, Long-Distance Trip Makers

Cluster composition. Cluster 6 is the least populated cluster. Though there are fewer members of this cluster, there are striking differences among these people. The changes in average distance from day-to-day within this cluster are higher than any other cluster. The mean change in total distance from day to day for members of this cluster are 429.11, 543.82, and 497.66 for day comparisons of day 1 to 2, 2 to 3, and 1 to 3. Most likely, this is a result of the changes due to long-distance travel. The changes in total trips from day to day are on the higher end of the values across clusters, but are not exceptionally high, indicating that the changes in total distance are due to changes from longer travel distances rather than higher trip frequencies. This is reflected in the mean values for the change in average trip distance, as these values range from 82.63 and 108.66 miles per trip. This cluster also has high changes in standard deviation of trip distance, which would also be the result of long-distance trips accompanying short-distance trips within a day.

Cluster members. Like clusters 1, 2, and 4, having days where no travel is recorded has a negative effect on cluster membership for cluster 6. In addition to this, individuals between the ages of 26 and 64 have a higher likelihood of being members of cluster 6. Additionally, females are less likely to be members of cluster 6. Those who are employed and retired are also less likely to be members of cluster 6. However, those who do not have a fixed workplace are more likely to be members of cluster 6.

Through the analysis of each cluster, additional perspective is gained on how individuals vary from day to day, and explore latent reasons for similarity among respondents in their variation. However, further attributes of the clusters can be examined to gain greater understanding of the differences and similarities in the clusters that have been developed. The distribution of total trips for all 3 days, plotted against total distance for all 3 days, is provided in Figure 2-11. There are several notable findings from analyzing this distribution. First, it is clear that members of cluster 6 (in red and diamond shaped) have the highest total distance values; this confirms the hypothesis that these members make up a majority of the long-distance travelers. Cluster 4 members (in purple and square shaped) appear to be the second highest in total trip distance, likely confirming the hypothesis that these individuals are the mid-length, long-distance travelers. While both cluster 4 and cluster 6 members have longer distances traveled for the 3 days, these members have comparatively smaller numbers for total cumulative trips across the three-day period. On the other hand, members of clusters 1—mid-variation local travelers (in grey and oval shaped) and 3—low-variation local travelers (in blue and triangle shaped) have a much higher distribution of number of trips, while in both clusters the total distance traveled for all 3 days is comparatively lower than other clusters. Cluster 5 members—extreme variation with one or 2 nontravel days (in yellow and rectangle shaped) have both low values for total distance traveled and total number of trips, while cluster 2—local longer-distance trip makers (in green and circle shaped) seem to be more distributed across both total distance and total number of trips. The low values observed for cluster 5 members are likely influenced by the fact that this cluster has the highest number of individuals who did not record travel for one or 2 days of the survey period. In order to see the patterns of distribution of the clusters, the axes were rescaled. This zoomed-in graph is shown in Figure 2-12. This rescaling removed the extreme values (mostly from cluster 6 for total distance, and cluster 1 for total trips) out of the figure to focus more closely on the lower values. This figure confirms the conclusions mentioned previously for the trip distance and frequencies for clusters 1, 2, 3, 4, and 5.

Figure 2-11: Total Distance vs. Total Number of Trips for All Days

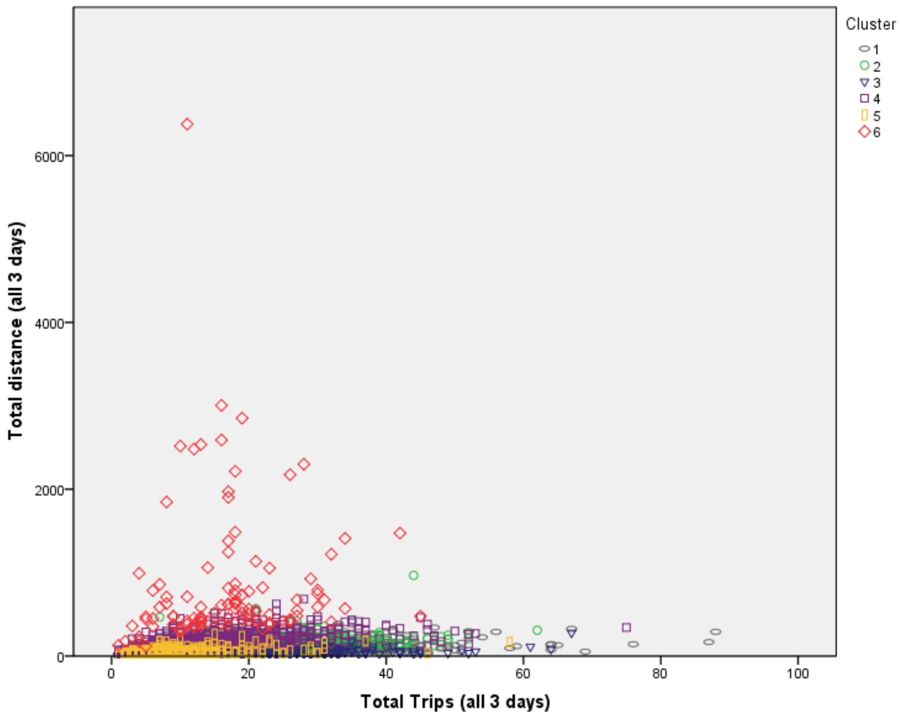
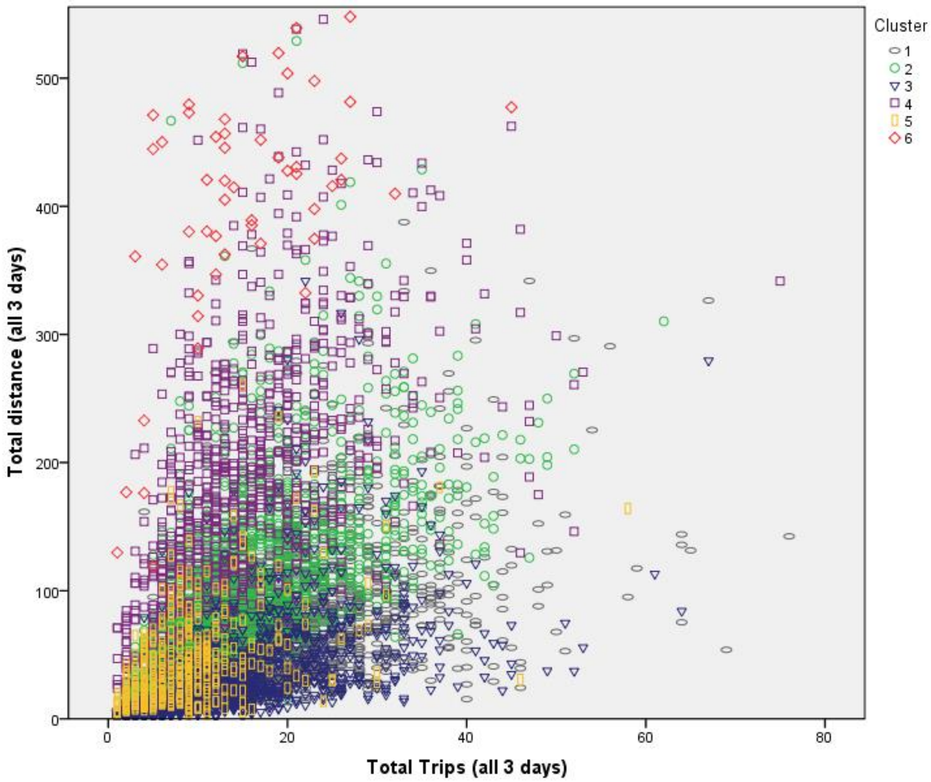
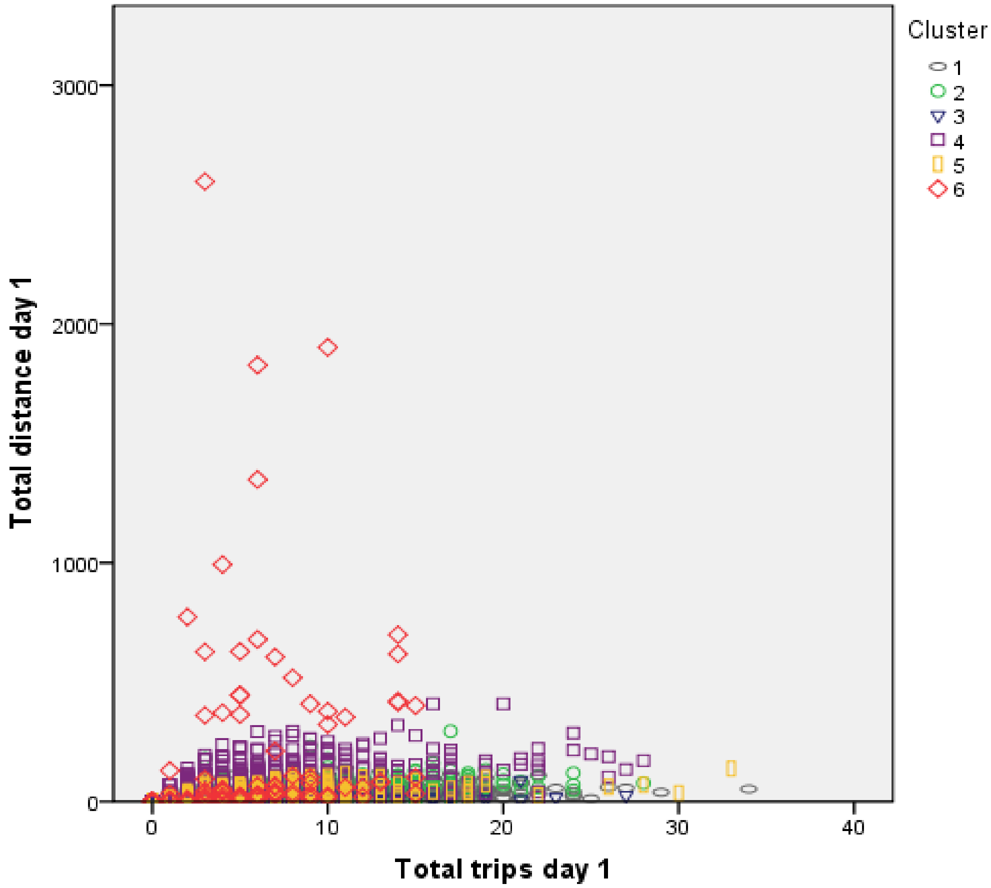


Figure 2-12: Total Distance vs. Total Trips for All Days (rescaled axis)



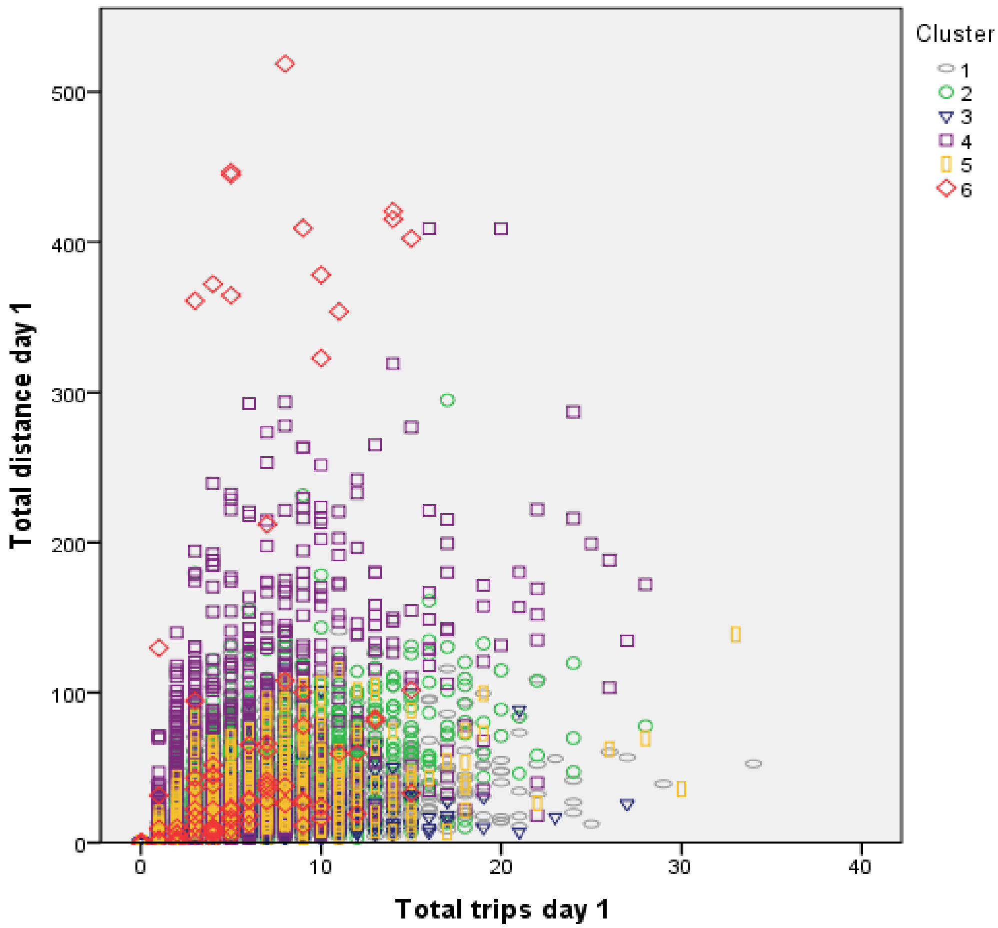
In addition to the preceding figures, similar figures have been created for each day of data collection. Figure 2-13 shows the total distance traveled for day 1 on the y-axis, plotted against total number of trips for day 1. Similar conclusions can be drawn from day 1 statistics as from the figures showing all 3 days cumulatively. Cluster 6—the long-distance trip makers (in red and diamond shaped) is again the cluster with the highest travel distances, although this cluster also appears to have many members with low total distance values. Cluster 5—the nontravelers (in yellow and rectangle shaped) has a majority of the lowest values for total distance, but is more evenly distributed (similar to the other clusters) in number of trips. Cluster 4 (in purple and square shaped) has the mid-range distance values when comparing across clusters, and also seems to be more distributed across total number of trips. Cluster 3 (in blue and triangle shaped) members are difficult to identify in this figure, likely due to their low values and being "buried" under other data points from other clusters.

Figure 2-13: Total Distance vs. Total Number of Trips for Day 1



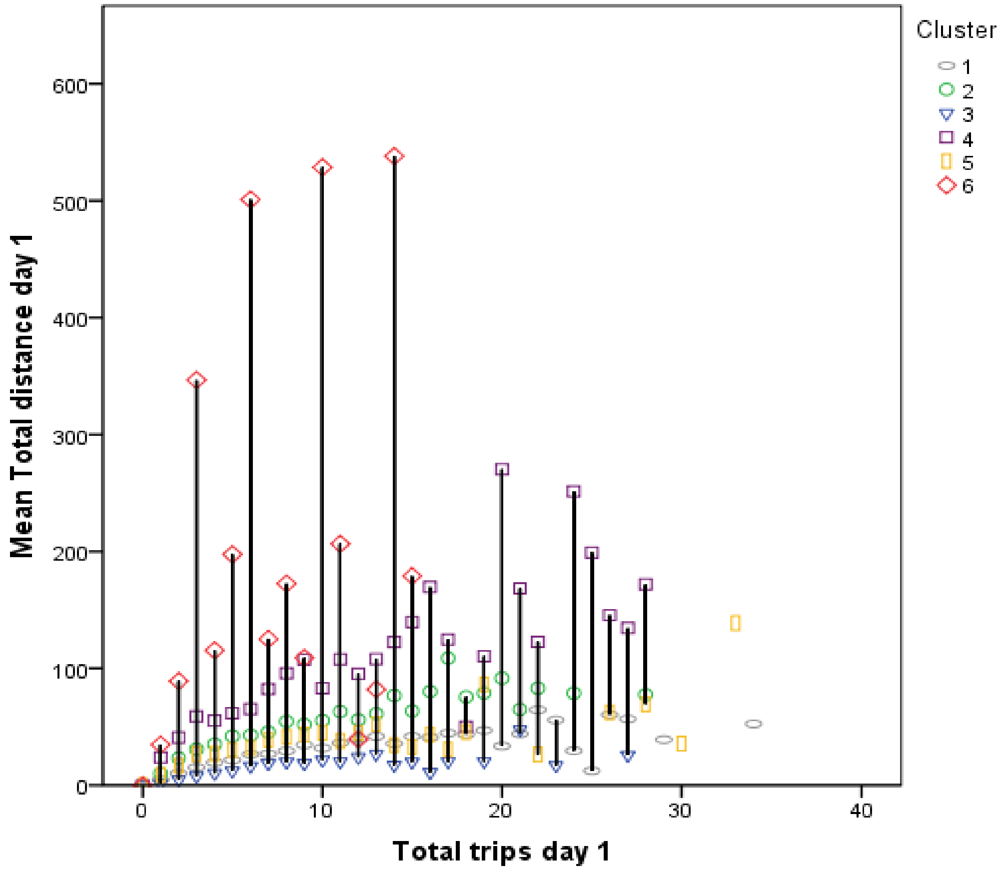
For this reason, Figure 2-14 provides this distribution, but this time with a rescaled axis in order to observe differences more closely. It is again noticeable that cluster 4 members have higher total distances and lower total trips than other clusters. However, there is less of a noticeable difference in the number of total trips across clusters. Although the presence of clusters 1 (in grey and oval shaped), 2 (in green and circle shaped), and 3 (in blue and triangle shaped) are noticed, they are again perhaps buried under the numerous other data points and it is difficult therefore to distinguish.

Figure 2-14: Total Distance by Total Number of Trips for Day 1 (Rescaled Axis)



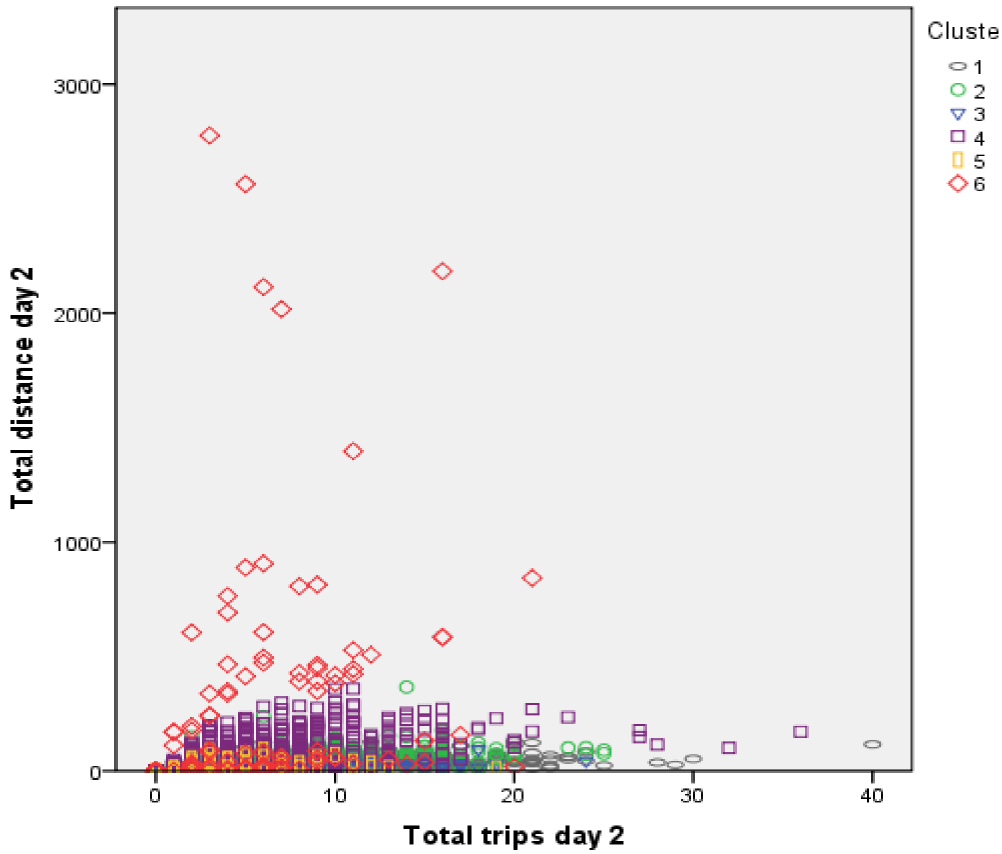
To examine the underlying trends further, Figure 2-15 reports the mean total distance for all respondents categorized by number of trips. Additionally, the respondents are further grouped into cluster by color. Using this figure, a much larger average for total distance among cluster 6 members for the lower numbers for total trips is observable. As the total number of trips increases for day 1, cluster 4 members begin to have the highest mean in total distance. Cluster 3 has a consistent representation across nearly all numbers for total trips, and is consistently the lowest across all clusters for mean total distance.

Figure 2-15: Mean Total Distance vs. Total Number of Trips for Day 1



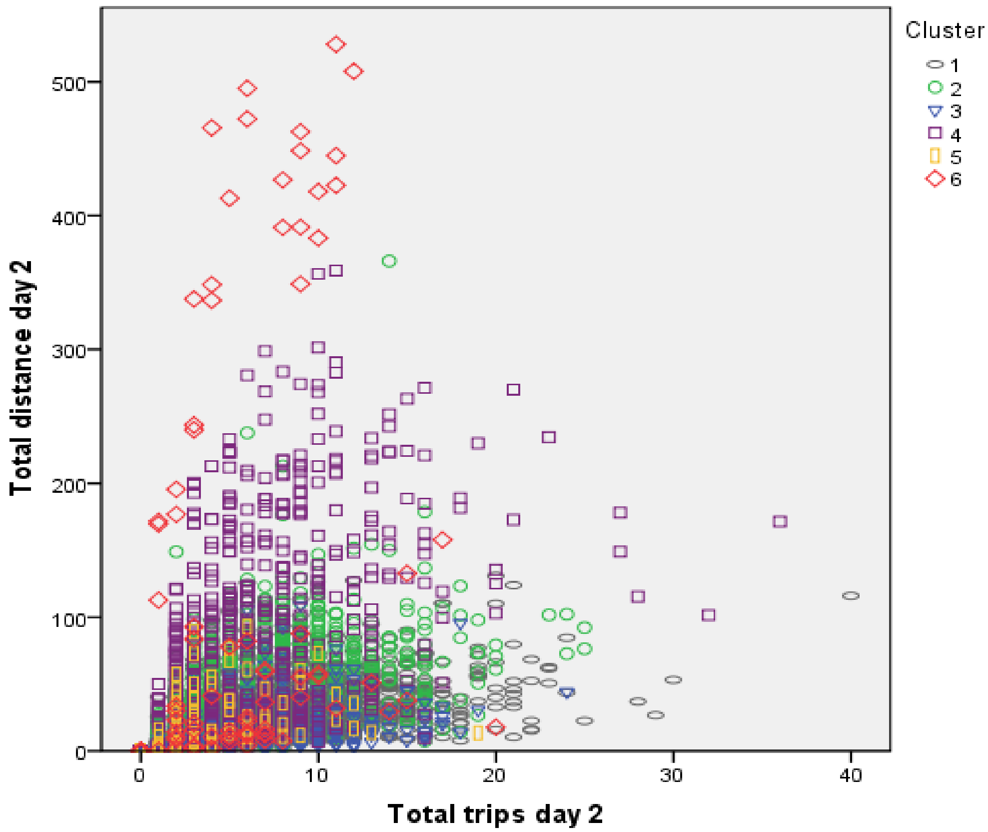
The analysis of travel attributes by cluster membership for day 1 (as presented previously) was repeated for day 2 and day 3. The results of this analysis are provided in Figure 2-16 and Figure 2-17 (for day 2) and Figure 2-19 and Figure 2-20 (for day 3). Similar patterns can be observed in the total distance versus total trips for each cluster on day 2. Cluster 6 again shows the highest total distance values across all clusters. Additionally, cluster 4 is noticeable for the mid-range distance values across a majority of the values for total trips. Strikingly different, however, is the much lower presence of distributed values for cluster 5 (in yellow and rectangle shaped) across total distance and total number of trips. This result is explained by the fact that cluster 5 has a large number of days in which there was no travel recorded for each of its members (mean of 1.84 days).

Figure 2-16: Total Distance by Total Trips for Day 2



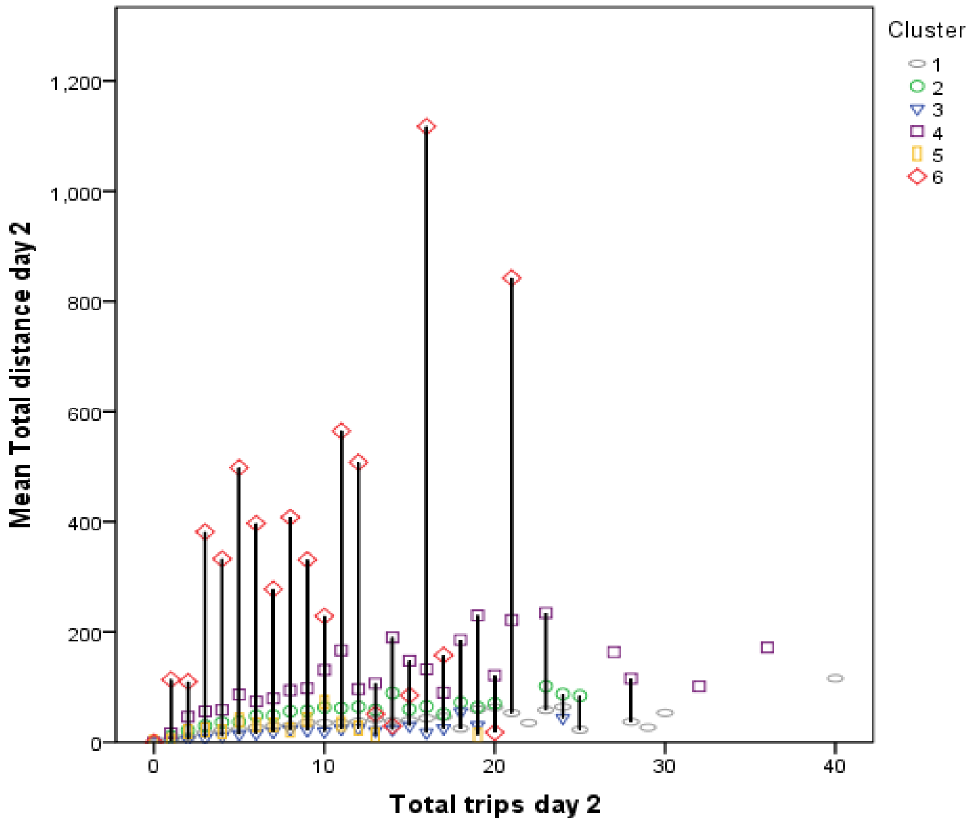
By rescaling the axes, some of the distributions within clusters of total distance and total trips are better understood. Figure 2-16 provides the total distance by total trips displayed by cluster membership for day 2. Clusters 1 and 4 have the highest values for total number of trips across all clusters. It is clear that cluster 2 (in green and circle shaped) and cluster 3 (in blue and triangle shaped) have mid-range values for total distance across the five clusters with lower total distance values. Additionally, cluster 5 (in yellow and rectangle shaped) has much higher densities within the range of one to five trips (and, as previously mentioned, a much higher likelihood for members to have no travel).

Figure 2-17: Total Distance by Total Trips for Day 2 (Rescaled Axis)



Again, an examination of the mean total distance for day 2 by the total number of trips reported for each cluster reveals some of these trends more clearly (as seen in Figure 2-18). The y-axis on Figure 2-18 is higher than for day 1 due to the higher mean total distance for cluster 6. Cluster 3 shows the lowest mean total distances, and cluster 6 shows the highest. Cluster 1 has some of the highest numbers for total trips recorded, although lower mean total distances for the members of the cluster with these values.

Figure 2-18: Mean Total Distance by Total Number of Trips for Day 2



Day 3 observations are similar to day 2 observations. Cluster 6 has the highest values (and range in values) for total distance traveled in the day, as seen in Figure 2-19. However, day 3 has the highest total distance—more than twice that of day 1 or day 2, at 6,000 miles. Because of this, the patterns in the rest of these data are difficult to decipher, as the axis is too large in scale. In order to observe the other remaining clusters, the axis was again rescaled.

Figure 2-19: Total Distance by Total Trips for Day 3

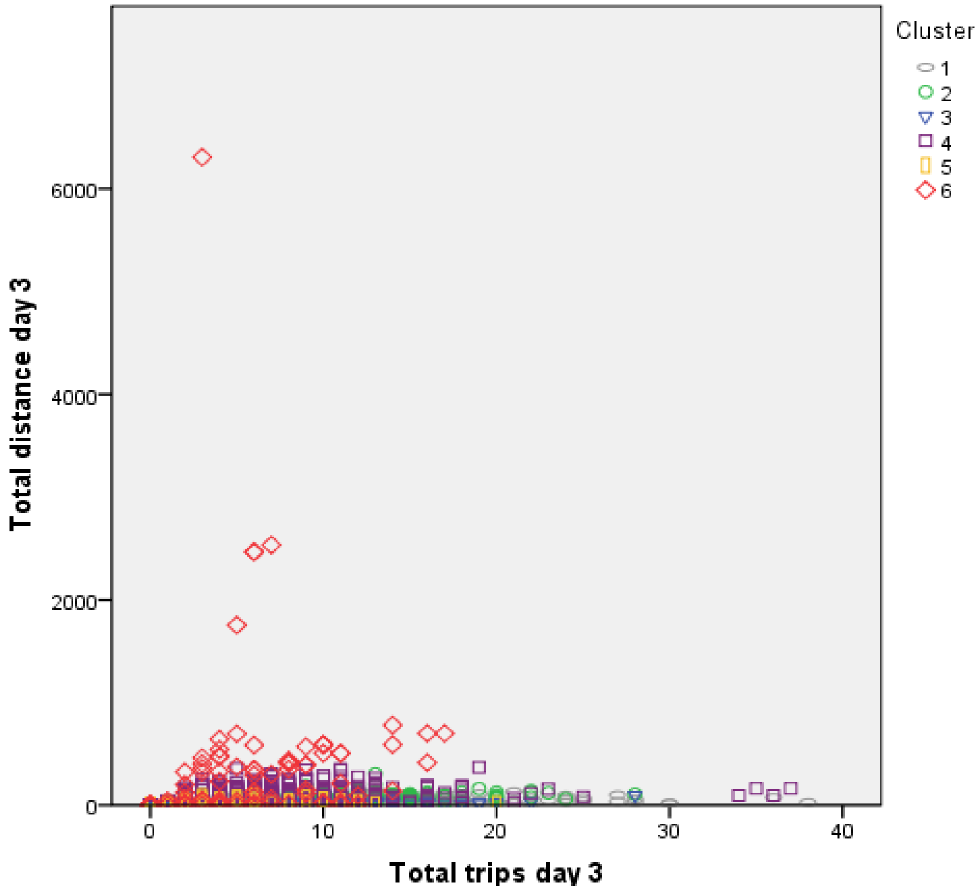


Figure 2-20 provides the distribution of total distance by total number of trips for day 3 with a rescaled y-axis. Rescaling unveils similar patterns to day 2. Although there might be slightly more obvious groupings of cluster number 5, there is still a low visual frequency of these cluster members compared to day 1. Day 3 is similar to day 2 observations in most other respects.

Figure 2-20: Total Distance by Total Trips for Day 3 (Rescaled Axis)

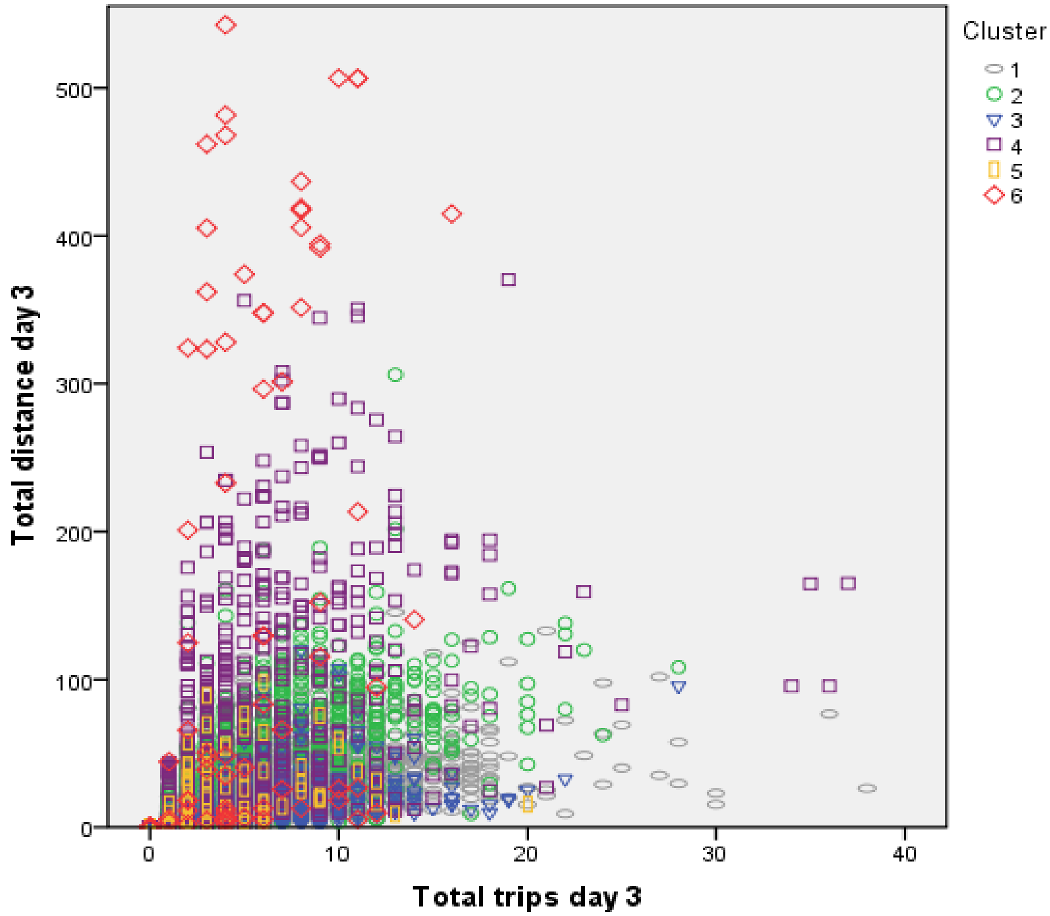
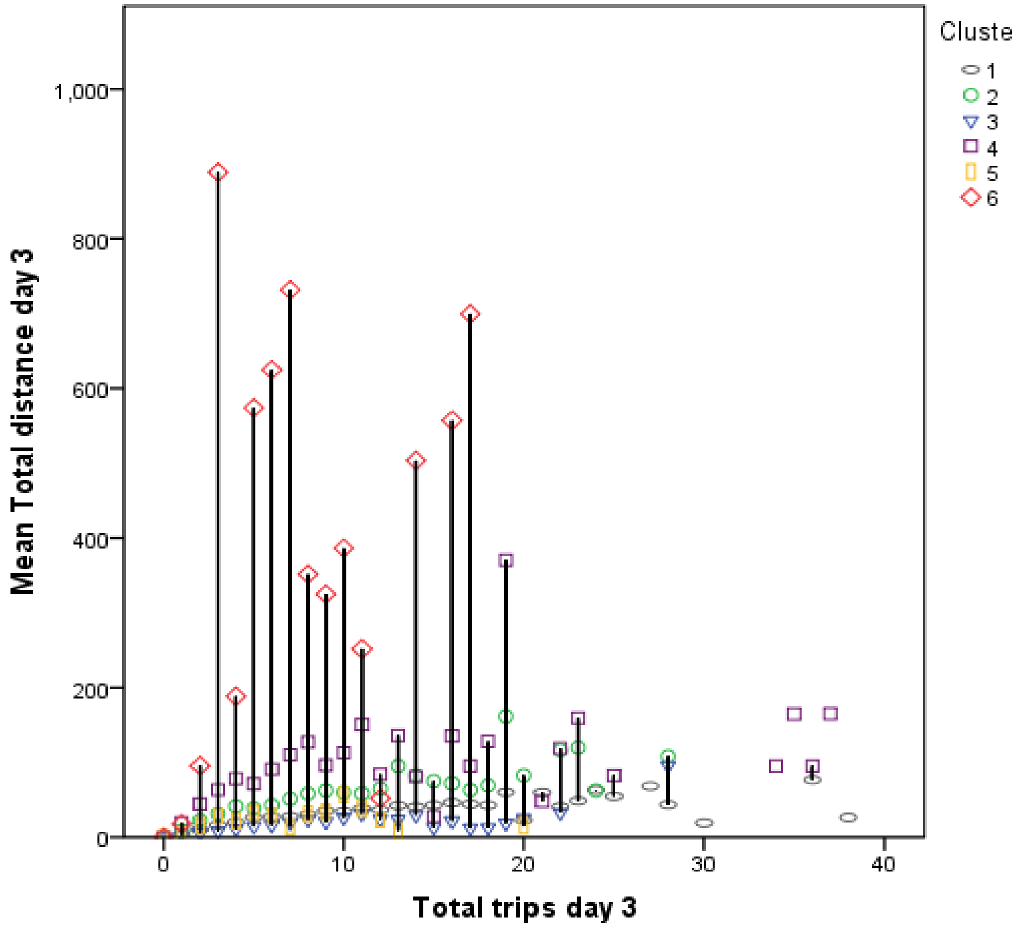


Figure 2-21 provides a more simplified version of the previous two figures; it does this by presenting the mean total distances by number of trips for day 3 on an aggregate level. As would be expected due to the aforementioned similarities, this figure is similar to day 2, with the highest distances belonging to cluster 6, and the highest numbers of total trips belonging to cluster 1 and 4 members. There are a few notable differences, however. First, there is a higher frequency of mean total distances above 600 miles within cluster 6 members, and higher frequencies of 30 or more trips.

Figure 2-21: Mean Total Distance by Number of Trips for Day 3



To further examine differences in respondents in light of cluster membership, trip attributes by cluster are provided in Table 2-12. As seen in Table 2-12, clusters 3 and 5 have the lowest mean total distances, and cluster 5 has the lowest total number of trips across all days. The mean total distance for cluster 3, by day, is consistently low across all 3 days, showing little change in the total distance traveled. However, cluster 5 shows a much different story for mean total distance for each day. Cluster 5 members have much lower total travel distances for day 2 and day 3 when compared to day 1.

Table 2-12: Travel Descriptive Statistics, by Cluster

		Cluster 1			Cluster 2			Cluster 3			Cluster 4			Cluster 5			Cluster 6		
		1998			1679			1638			768			536			104		
		Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Total Dist		6.42	387.65	67.48	15.57	967.16	110.58	0.06	341.72	32.32	23.78	682.20	196.71	3.21	260.82	40.73	118.85	6,376.95	850.10
Total Trips		1.00	88.00	17.84	2.00	62.00	15.75	1.00	67.00	12.97	1.00	75.00	16.13	1.00	58.00	8.15	1.00	45.00	16.98
Ave Dist		0.39	40.39	4.07	0.78	66.68	7.77	0.06	21.83	2.31	2.81	70.97	14.79	0.29	25.32	5.79	10.61	579.72	62.86
Sd Dist		0.02	24.59	3.77	0.00	58.20	8.54	0.00	19.70	1.75	0.05	72.89	20.55	0.06	21.52	5.15	33.61	1,737.60	168.65
Total dist	1	0.00	140.97	22.87	0.00	294.79	38.23	0.00	116.20	11.14	0.00	409.15	68.22	0.00	138.69	32.57	0.00	2,596.48	200.45
	2	0.00	130.97	23.92	0.00	366.20	38.24	0.00	108.65	10.82	0.00	359.04	68.02	0.00	93.06	4.10	0.00	2,775.10	309.19
	3	0.00	145.43	20.68	0.00	306.17	34.11	0.00	119.64	10.36	0.00	370.35	60.47	0.00	97.85	4.05	0.00	6,307.05	340.45
Avg distance	1	0.00	40.44	3.83	0.00	59.87	7.54	0.00	21.95	2.21	0.00	71.08	13.16	0.00	27.94	5.71	0.00	865.49	39.84
	2	0.00	40.36	3.99	0.00	74.45	7.46	0.00	22.04	2.11	0.00	66.82	12.26	0.00	29.72	1.21	0.00	925.03	57.85
	3	0.00	40.39	3.56	0.00	69.14	6.58	0.00	21.48	2.04	0.00	87.82	12.49	0.00	29.73	1.20	0.00	2,102.35	67.39
Std trip dist	1	0.00	26.56	3.09	0.00	71.51	6.24	0.00	21.21	1.55	0.00	79.35	11.13	0.00	20.56	4.94	0.00	1,459.86	69.99
	2	0.00	28.12	3.14	0.00	75.00	6.11	0.00	20.79	1.49	0.00	77.22	11.64	0.00	23.78	0.64	0.00	1,566.89	94.66
	3	0.00	25.92	2.78	0.00	64.82	5.61	0.00	21.43	1.45	0.00	86.94	11.11	0.00	23.77	0.64	0.00	3,211.60	111.92
Total trips	1	0.00	34.00	6.17	0.00	28.00	5.54	0.00	27.00	4.74	0.00	28.00	6.10	0.00	33.00	6.63	0.00	15.00	5.60
	2	0.00	40.00	6.29	0.00	25.00	5.51	0.00	24.00	4.27	0.00	36.00	5.46	0.00	19.00	0.75	0.00	21.00	6.04
	3	0.00	38.00	5.38	0.00	28.00	4.70	0.00	28.00	3.96	0.00	37.00	4.57	0.00	20.00	0.77	0.00	17.00	5.35

The difference between clusters is illustrated in Figure 2-22 through Figure 2-25, which show the distribution of travel distances for day 1 and day 2. Due to the extreme differences between day 1 and day 2 for cluster 5, the axes are not on the same scale across all histograms. There is a striking difference in day 3 and day 5 cluster members in the 2 days. Cluster 3 members are spread across a range in total distance of less than 60 miles similarly across the 2 days. However, frequency charts of cluster 5 show a massive shift to much lower total distances traveled from day 1 to day 2. As discussed previously, this is due to the larger number of people who did not record travel for 1 or 2 days in the survey period.

Figure 2-22: Frequencies for Total Distance Traveled on Day 1 for Cluster 3

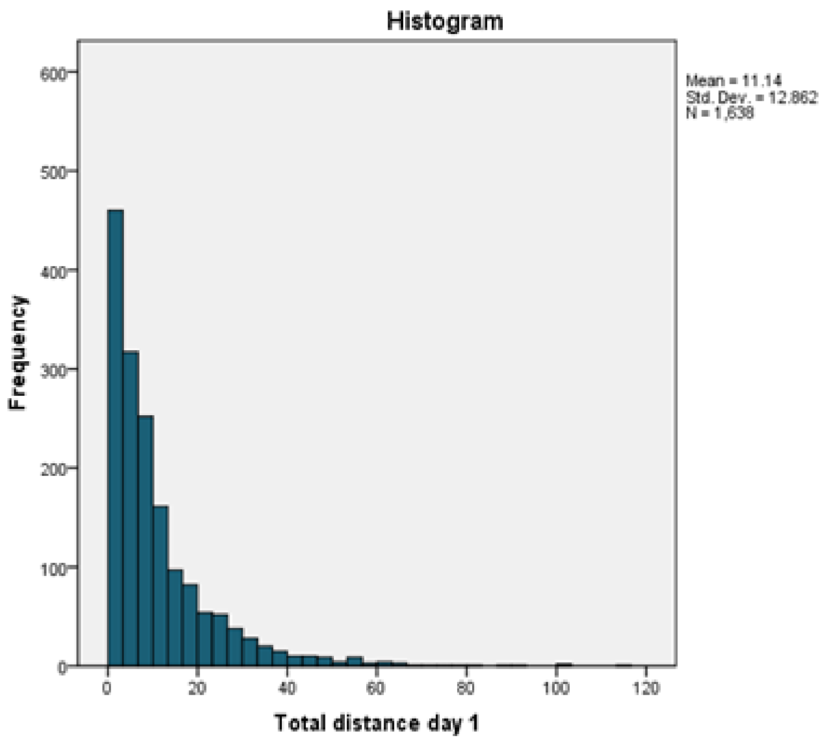


Figure 2-23: Frequencies for Total Distance Traveled on Day 2 for Cluster 3

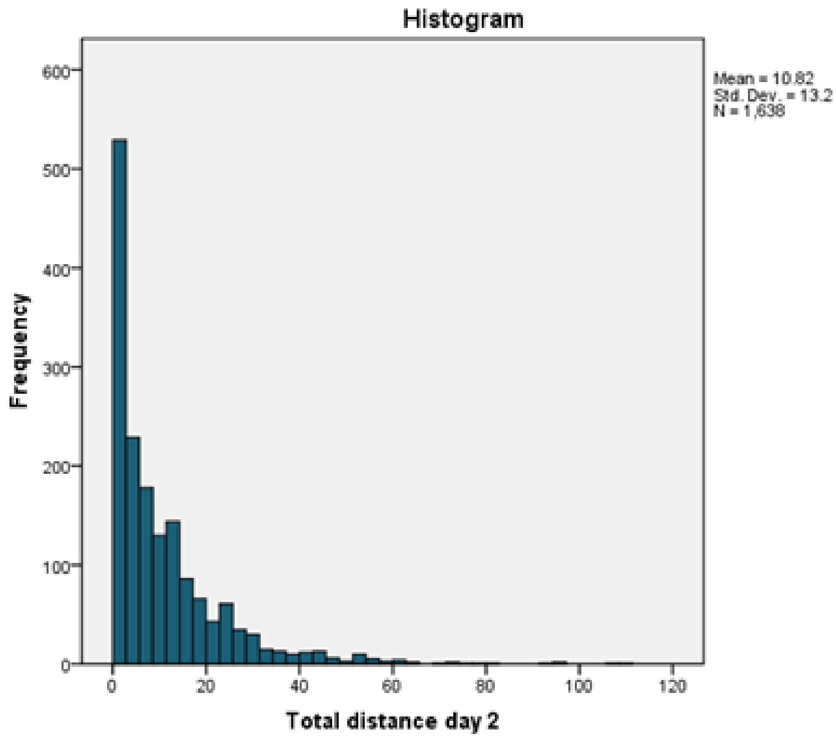


Figure 2-24: Frequencies for Total Distance Traveled on Day 1 for Cluster 5

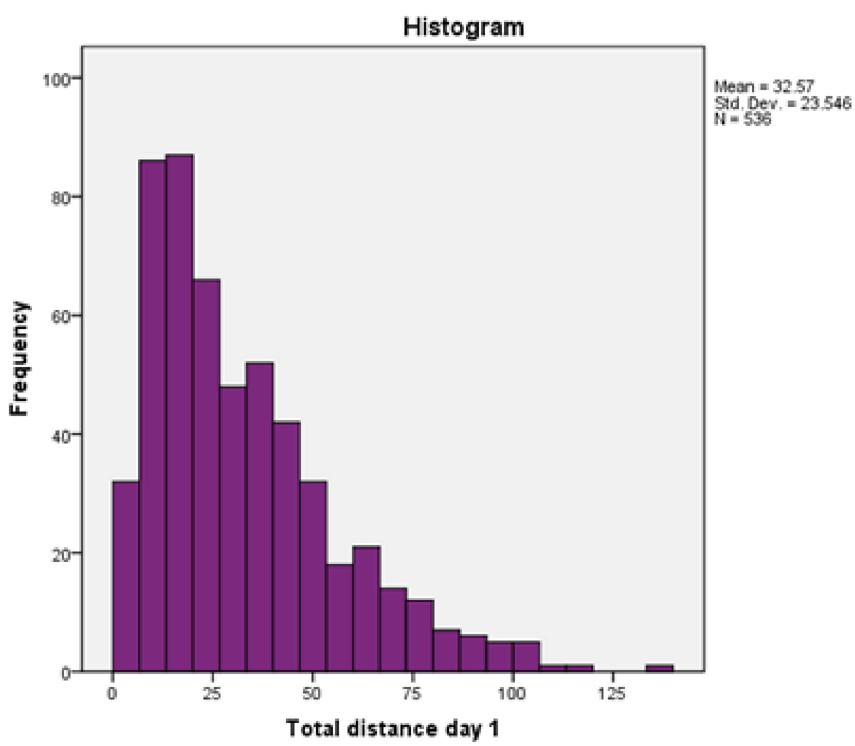
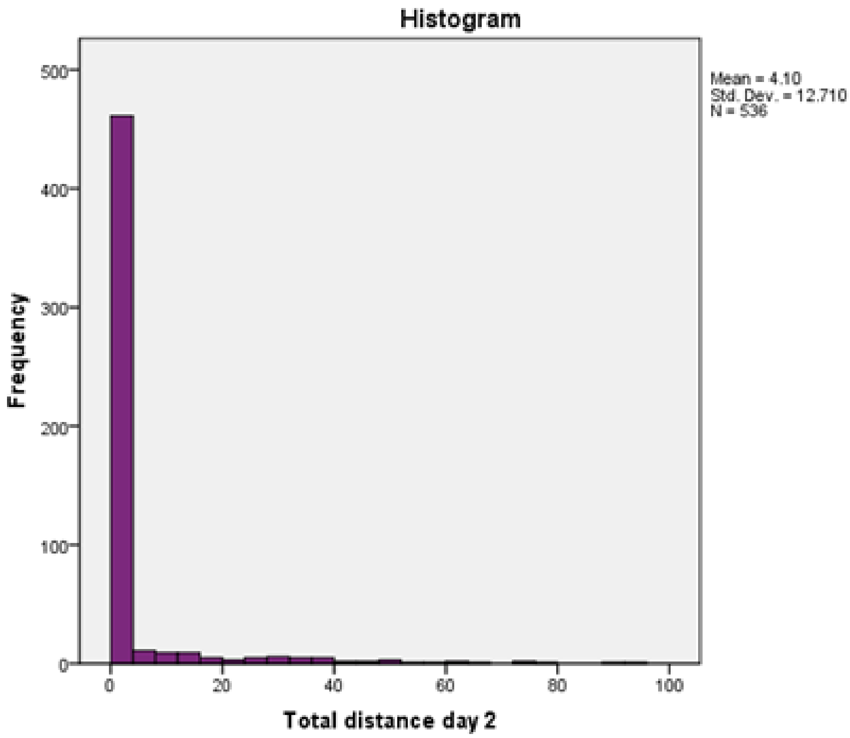


Figure 2-25: Frequencies for Total Distance Traveled on Day 2 for Cluster 5



In addition, distributions for average distances and standard deviation of trip distances were examined. Figure 2-26 through Figure 2-31 provide the average distances for each respondent (referenced on the left side y-axis) and standard deviation of trip distances for each respondent (referenced by the right axis) for day 1 (graph 1A and 1B), day 2 (graph 2A and 2B), and day 3 (graph 3A and 3B). In order to further explore the distribution of average trip distances and standard deviation of trip distances across clusters, the axes were rescaled to focus on clusters 1 through 5 (graphs 1B, 2B, and 3B). Respondents are categorized by their cluster membership (distributed on the x-axis). The impacts of high travel distances are again observed in cluster 6. This is evident in two ways. First, cluster 6 has an individual in each of the 3 days who has the highest average distance (865 miles, 965 miles, and 2,102 miles for days 1, 2, and 3, respectively) and highest standard deviation of trip distance (approximately 1,460 miles, 1,567 miles, and 3,212 miles for days 1, 2, and 3, respectively). Additionally, the distribution of the remaining cluster 6 members is much wider than any other cluster. This is most notable in day 2, although it is observable in day 1 and day 3. The highest value for any other cluster for average distance for day 1 was 71 miles (cluster number 4), and 79 miles for standard deviation of trips (also cluster 4), day 2 highest average was 74 miles (cluster 2), and 77 for standard deviation (cluster 4), and day 3 average was 87 miles for average distance and 86 miles for standard deviation (both cluster 4). When the axes are rescaled, it is notable that cluster 3 has the lowest average distances and standard deviations of distances (although cluster 5 is a close second, and even lower in maximum standard deviation for day 1). Cluster 3 and 5 members have much smaller distributions. Cluster 4 has much higher values, and its members are much more distributed across the values. In addition, cluster 4 has higher maximum values for both average distance and standard deviation of trips for day 3 compared to days 1 and 2, although the mean of these values across all cluster members remains similar. Interestingly, with a rescaled axis, an even more exaggerated distribution of cluster number 6 is observed, with many members having smaller values, and an additional cohort of cluster 6 having higher values, with some noticeable gaps in between.

Figure 2-26: Average Distance and Standard Deviation of Trip Distance by Cluster Membership for Day 1A

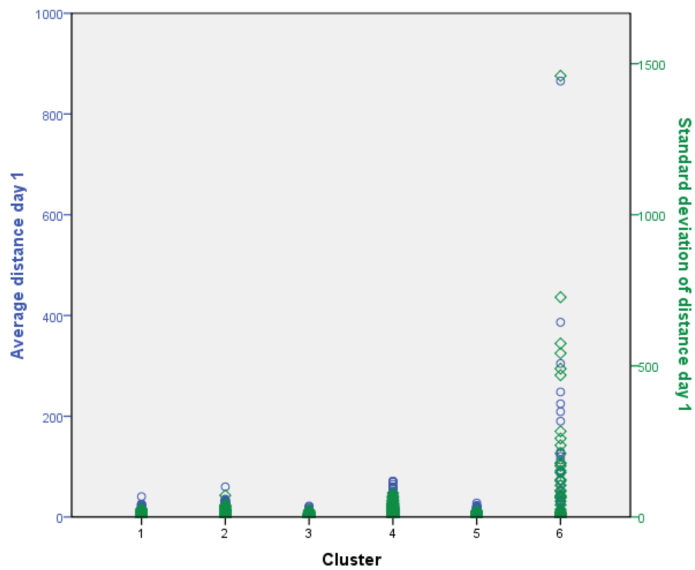


Figure 2-27: Average Distance and Standard Deviation of Trip Distance by Cluster Membership for Day 1B

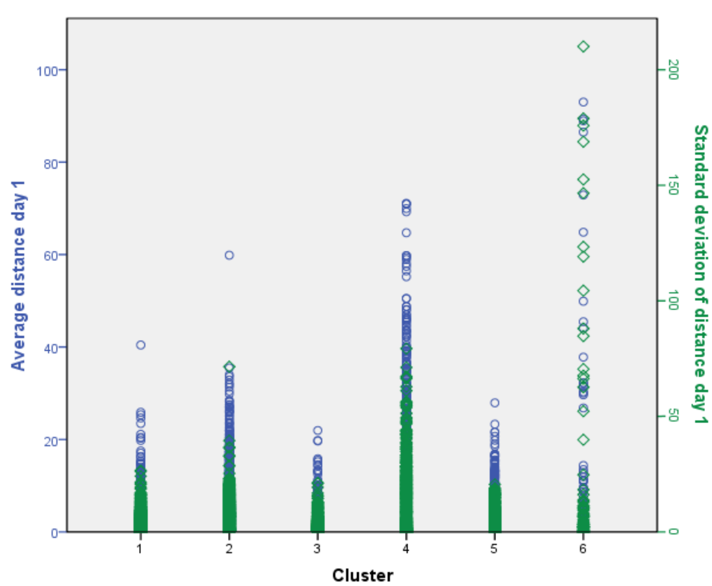


Figure 2-28: Average Distance and Standard Deviation of Trip Distance by Cluster Membership for Day 2A

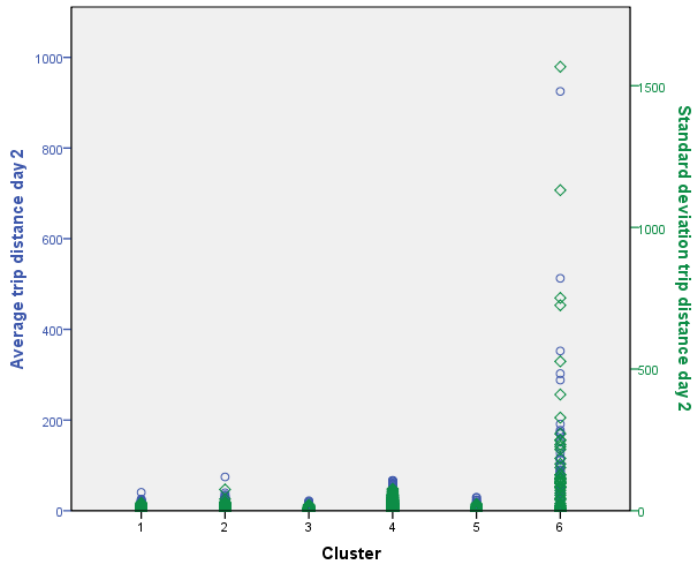


Figure 2-29: Average Distance and Standard Deviation of Trip Distance by Cluster Membership for Day 2B

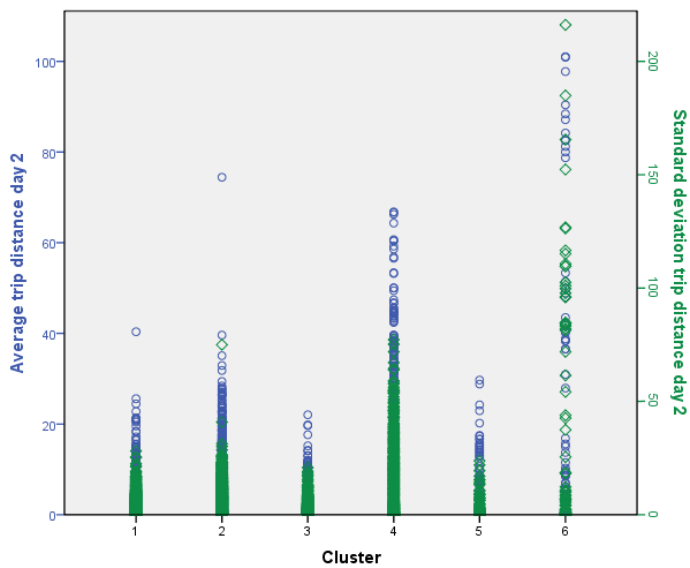


Figure 2-30: Average Distance and Standard Deviation of Trip Distance by Cluster Membership for Day 3A

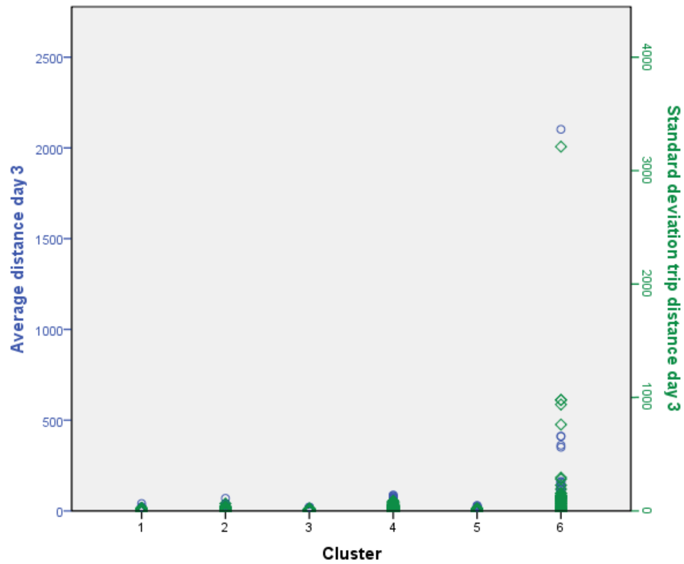
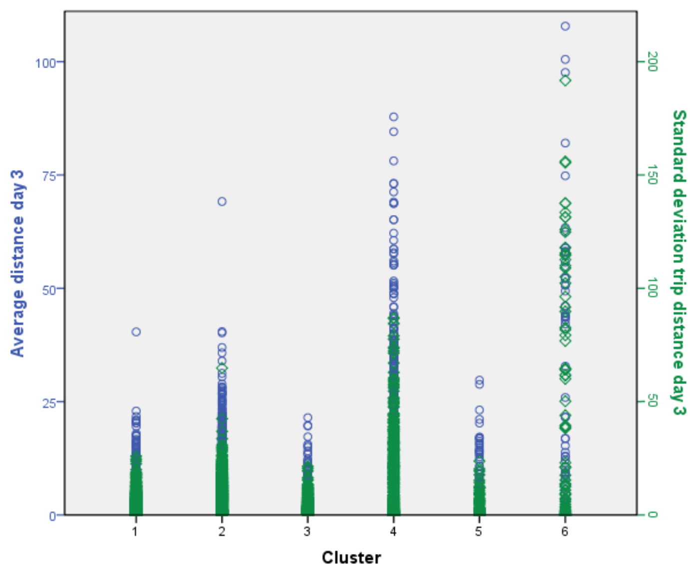


Figure 2-31: Average Distance and Standard Deviation of Trip Distance by Cluster Membership for Day 3B



Lastly, to finish the analysis of cluster membership and differences across clusters, many of the conclusions regarding travel distances and destinations should be validated. Figure 2-32 through Figure 2-37 provide maps of activity locations, but this time separately reported by cluster. Though many of the activity locations are focused in the Bay Area, some activities (especially in certain clusters) occur on a much larger geographic scale. For this reason, maps are provided of a larger geographic area to visualize the cluster membership of the long-distance travelers in the survey. In addition to these maps, a zoomed-in version of each is provided.

These second maps are focused on the activity locations located in the Bay Area for each cluster.

Though the activity locations in cluster 1—the mid-variation local trip makers—are mainly clustered in the Bay Area, there are still a few individuals who conducted activities in places that required long-distance travel. It is important to note that it is still possible that an individual did not record his or her long-distance travel, as this might have been a part of a travel day before the survey period, but many of these long-distance trips are included in the measurements. Cluster 2 members —the local long-distance travelers—similarly have activity locations clustered in the Bay Area, but the cluster of points stretches to a larger area within the Bay Area. Cluster 2 also has some long-distance travelers; however, these long-distance trips likely took place before the travel day, or distributed across many of the travel days, as the maximum recorded total distance for any day is around 300 miles. Cluster 3—the low-variation local trip makers—is similar to cluster 1 and has a much tighter geographic footprint. Cluster 4—the mid-range, long-distance trip makers—is one of the wider spreading clusters, and has activity location points that cover a large portion of the State of California. This confirms hypotheses made earlier in the discussion that cluster 4 has many regional travelers. Cluster 5—the partial nontravelers—is similar to cluster 1 and 3 and is comprised mostly of local travelers; it differs in that these members have one or 2 days with no travel recorded. Cluster 6—the long-distance trip makers—have activity locations across a much larger area, including Mexico, Alaska, and Europe (although the European points are not shown on the map). Cluster 6 has cluster members who traveled across many states, including the eastern states; however, many of the activity locations are still located on the western portion of the United States.

Figure 2-32: Map of Activity Locations for Cluster 1

Figure 2-33: Map of Activity Locations for Cluster 2

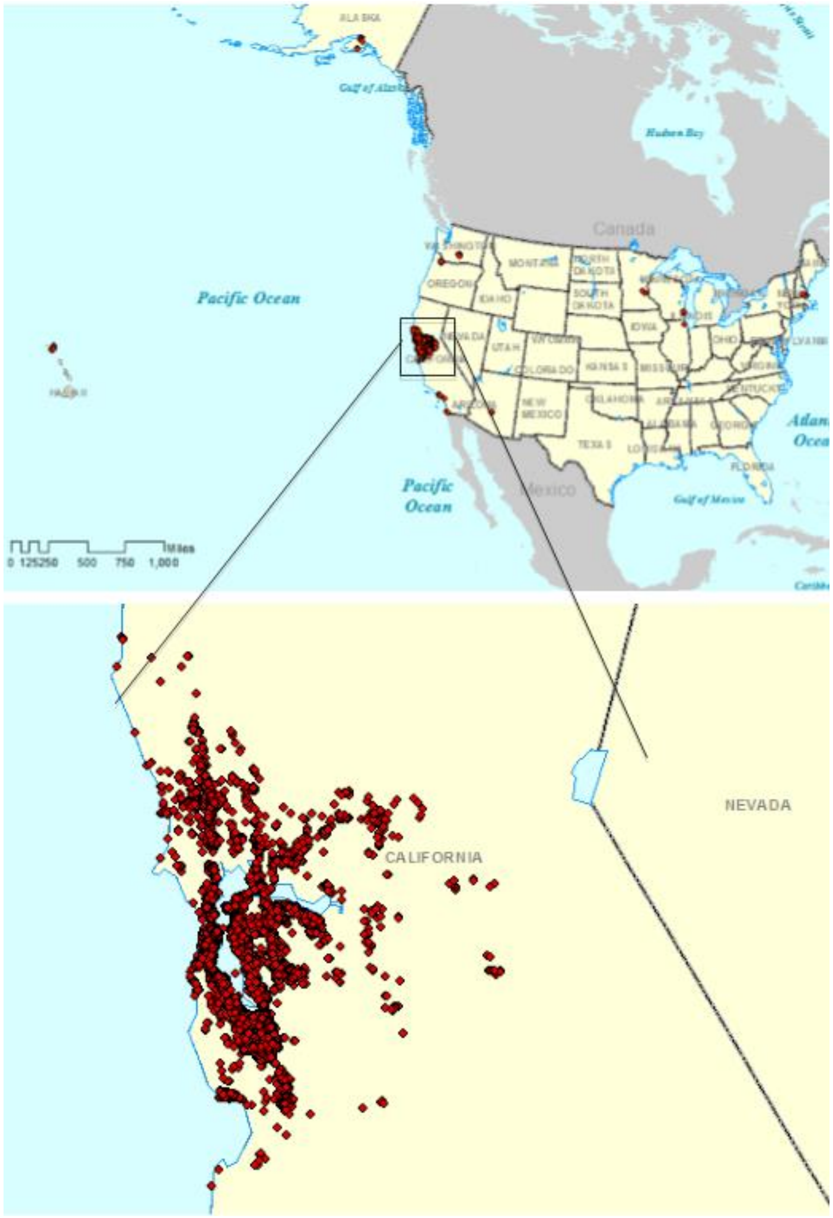


Figure 2-34: Map of Activity Locations for Cluster 3

Figure 2-35: Map of Activity Locations for Cluster 4

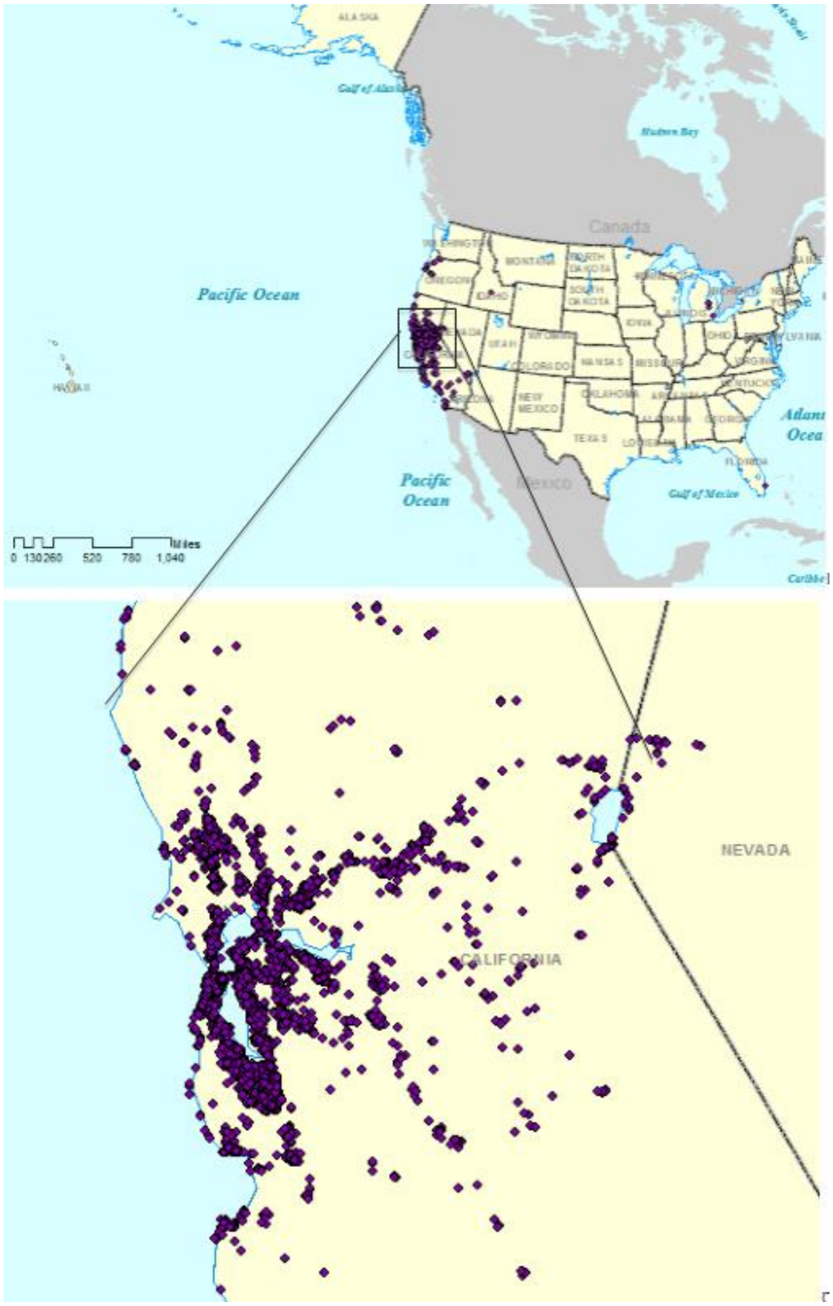


Figure 2-36: Map of Activity Locations for Cluster 5

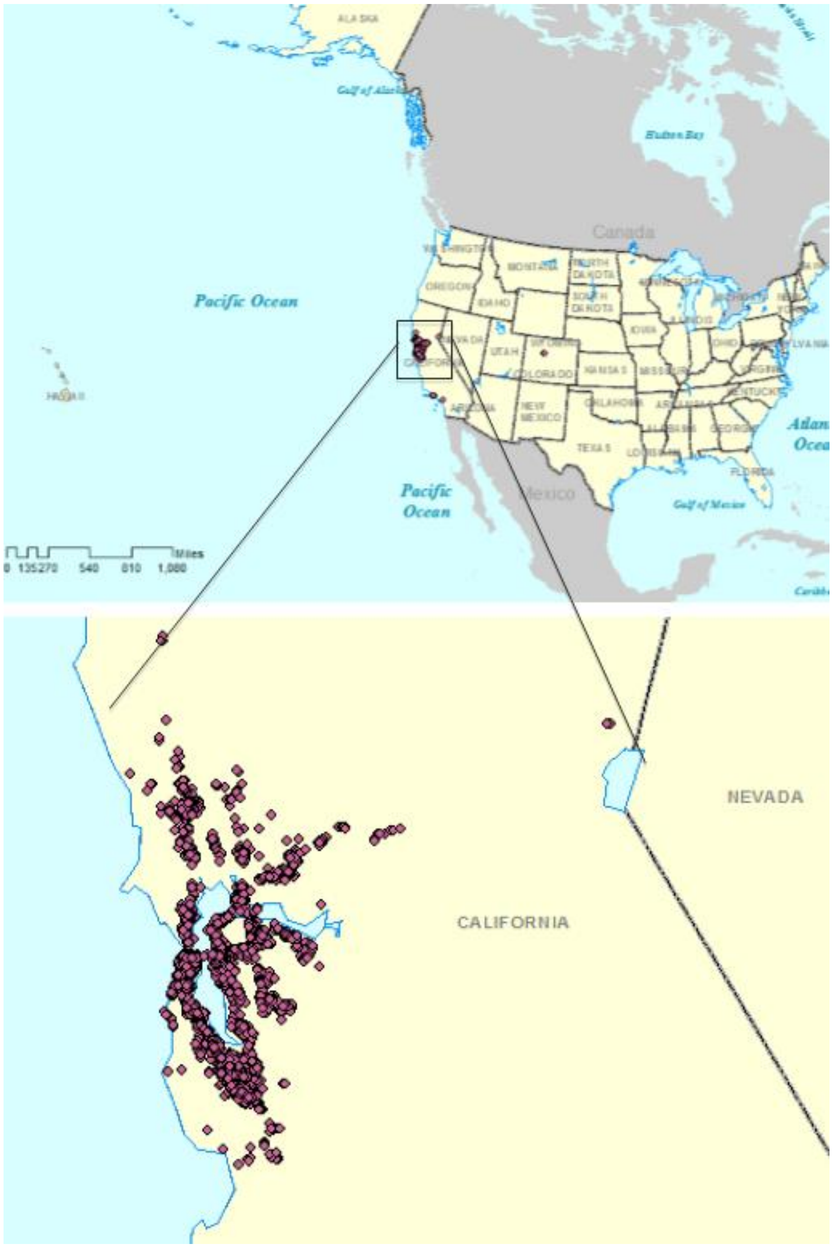


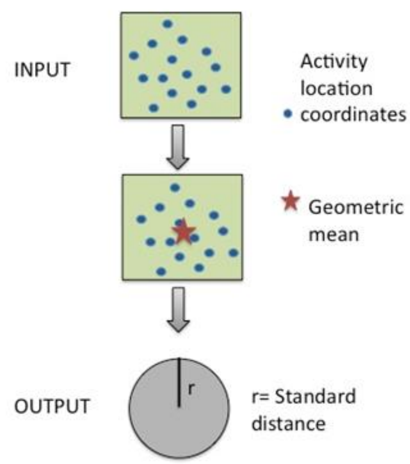
Figure 2-37: Map of Activity Locations for Cluster 6

It is important when analyzing the geographic nature of the clusters to keep in mind that cluster membership is not based on how far from home one is, but rather how widely the geographic extent of travel varies from day to day (clusters were formed using variables of change). To gain further insight into this aspect of cluster membership, an additional geographic analysis was conducted.

Examining Cluster Membership Using Point Clustering Methods

In effort to examine the geographic point clustering of an individual's activity locations, point data was analyzed using ArcGIS 10. An iterative model was developed utilizing the standard-distance tool in the spatial statistics toolbox. The standard-distance tool uses all points recorded by an individual for the desired unit of analysis (across a day or across the survey duration, in this case) to calculate a measure of geographic clustering. A geometric mean center point is calculated for each set of points, and the distance of each point to that mean is calculated. The final output is a value indicating the dispersion of the points, which is then used as the radius of a circle that encompasses the point locations. Standard distance can essentially be equated to the geographic equivalent to a standard deviation. While standard deviation measures the distribution around the statistical mean of a set of data values, the standard distance measures the distribution of geographic points around a geometric mean. Figure 2-38 provides a schematic for this process.

Figure 2-38: Schematic for Standard Distance Calculation



The results of the standard-distance computation are reported in Table 2-13. There are still high standard distances, which are the result of long-distance travel within the survey period. In addition to summary statistics, standard distances and changes in standard distances across days are further explored.

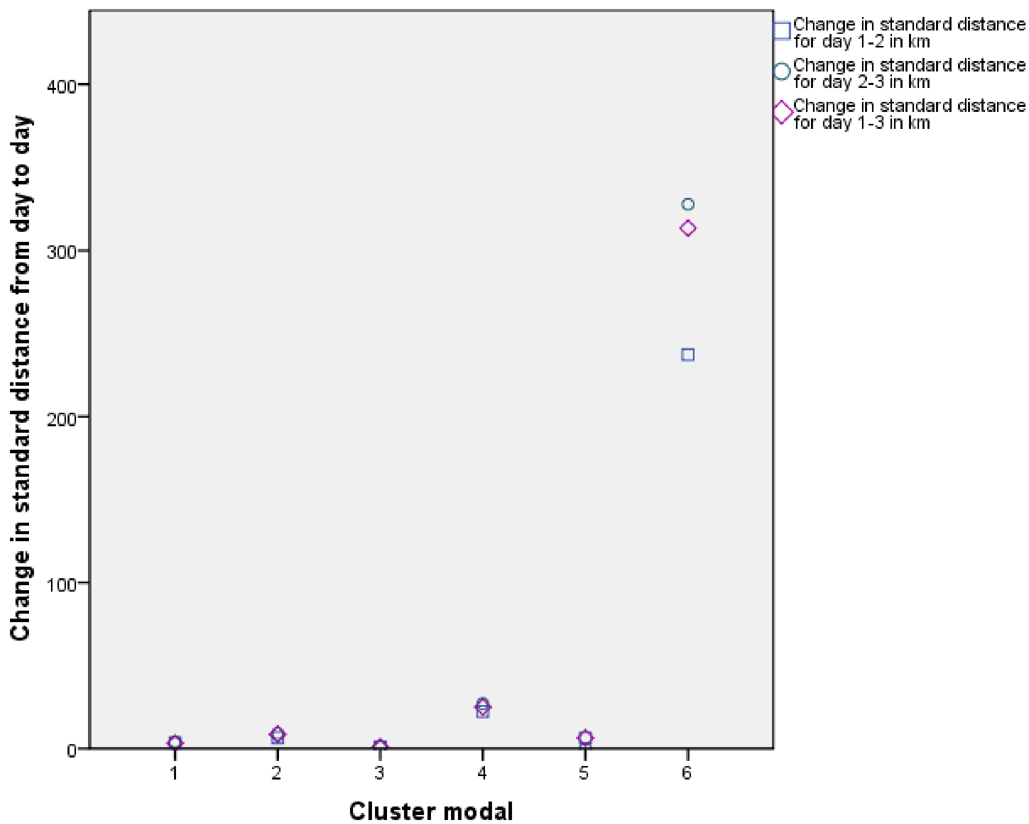
Table 2-13: Summary Statistics for Standard Distance (in km)

	Min	Max	Mean	Std. Deviation
Standard distance day 1 in kilometers	0.00	1,964.36	8.25	40.86
Standard distance day 2 in kilometers	0.00	2,089.17	8.61	51.26
Standard distance day 3 in kilometers	0.00	4,673.35	9.11	81.60

	Min	Max	Mean	Std. Deviation
Change in standard distance for day 1-2 in km	0.00	2,088.92	9.75	63.57
Change in standard distance for day 2-3 in km	0.00	4,667.61	11.95	94.80
Change in standard distance for day 1-3 in km	0.00	4,673.35	11.65	88.78

Figure 2-39 presents the average change in standard distance for each day by cluster numbers. Similar findings are apparent in the analysis of change in spatial distribution of activities. Again, cluster 6 has the largest values for change and is the most spread out. Cluster 4 also has some significant changes. Cluster 3 has the lowest change values, followed by cluster 1 (with the exception of one member who traveled to Florida during the survey period). Cluster 5, the other cluster with mostly local travel, also has low values for standard distances. There is a clear difference in the mean standard distance for change between day 1 and 3.

Figure 2-39: Mean Change in Standard Distance from Day to Day, by Cluster



These findings are further explored in Figure 2-40 through Figure 2-45, which show the change in standard distances from day to day. The first of each day-to-day comparison shows the full comparison, and the second shows a rescaled axis (only up to 200 km) to show the differences among the clusters with primarily local travel.

These figures illustrate trends that have been consistent in previous discussion. However, it is interesting to note from this analysis that—with the exception of cluster 5—all other clusters have outlier members for at least one of the day comparisons. Although the degree to which these outliers differ, there is a noticeable difference between the majority of the members of each cluster and the few members with higher standard-distance changes.

Figure 2-40: Change in Standard Distance for Day 1 to 2, by Cluster

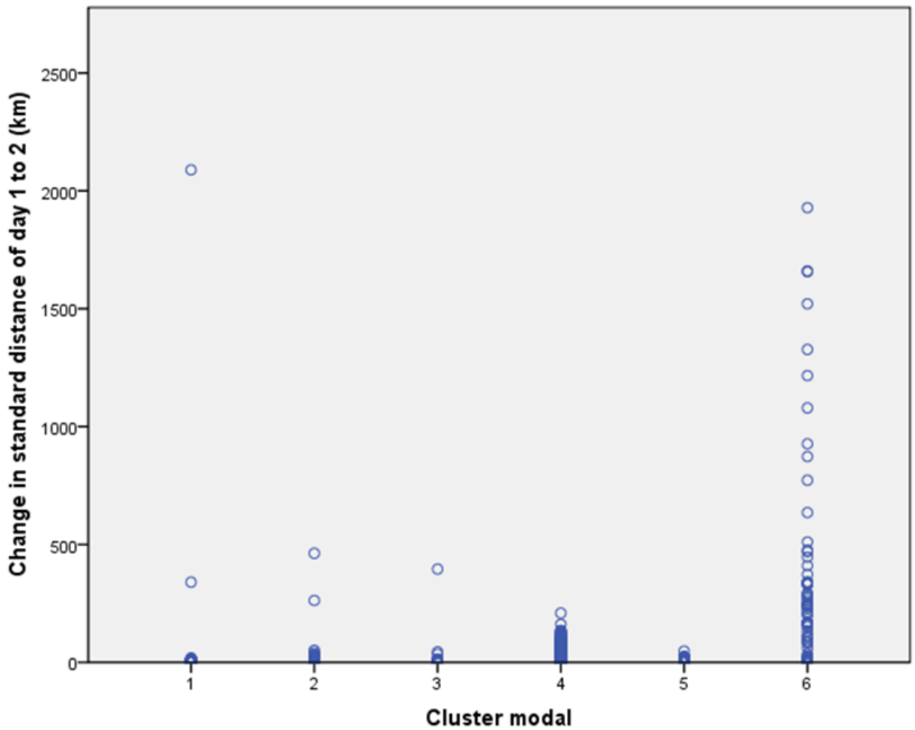


Figure 2-41: Change in Standard Distance for Day 1 to 2, by Cluster (with Rescaled Axis)

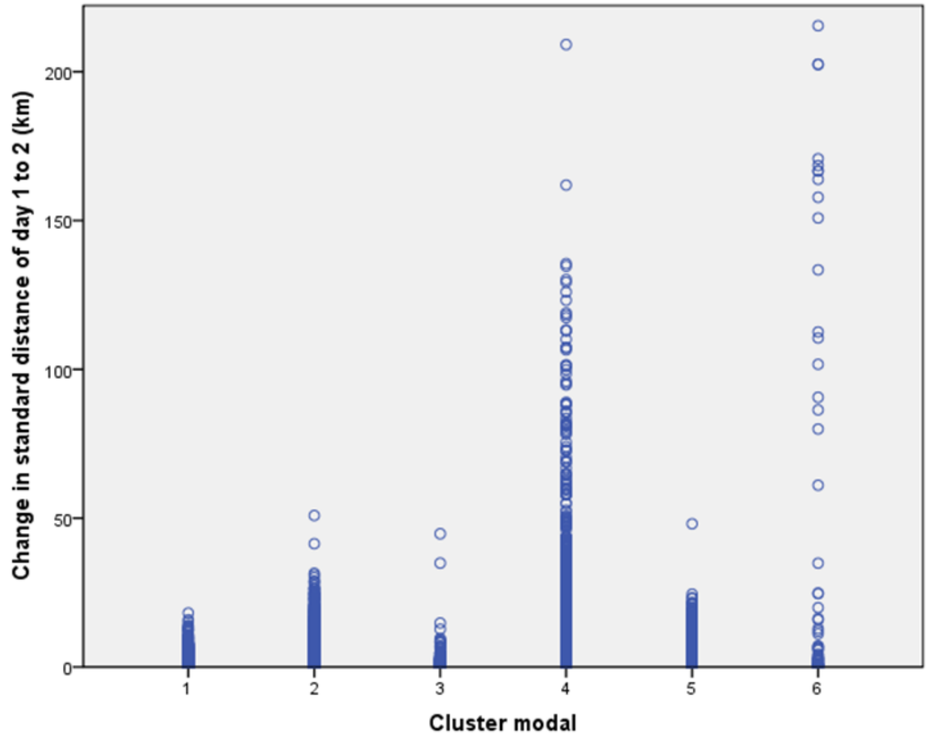


Figure 2-42: Change in Standard Distance for Day 2 to 3, by Cluster

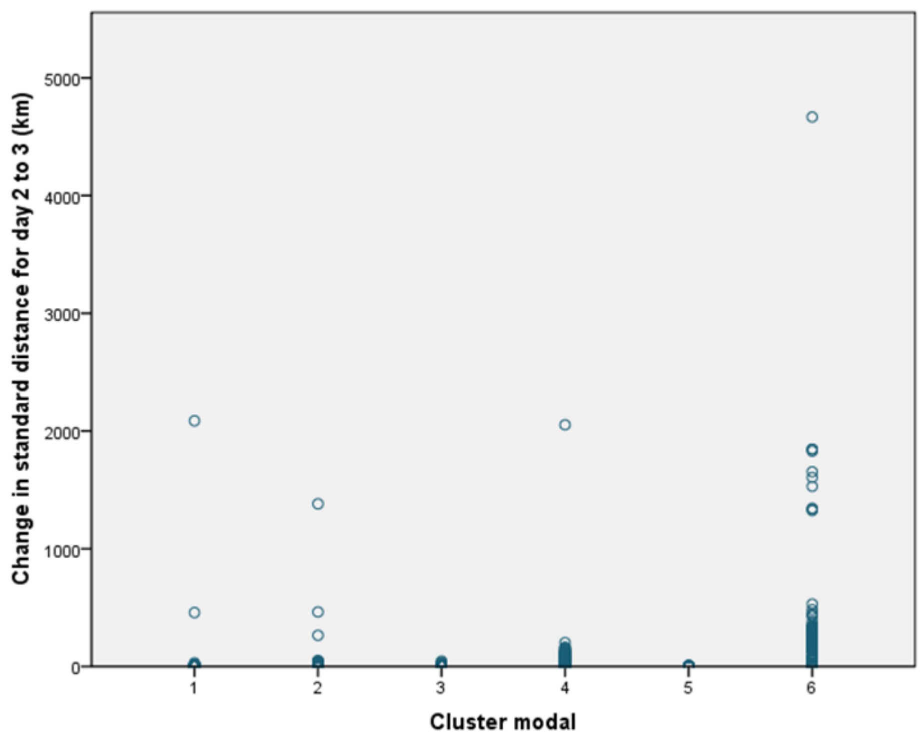


Figure 2-43: Change in Standard Distance for Day 2 to 3, by Cluster (with Rescaled Axis)

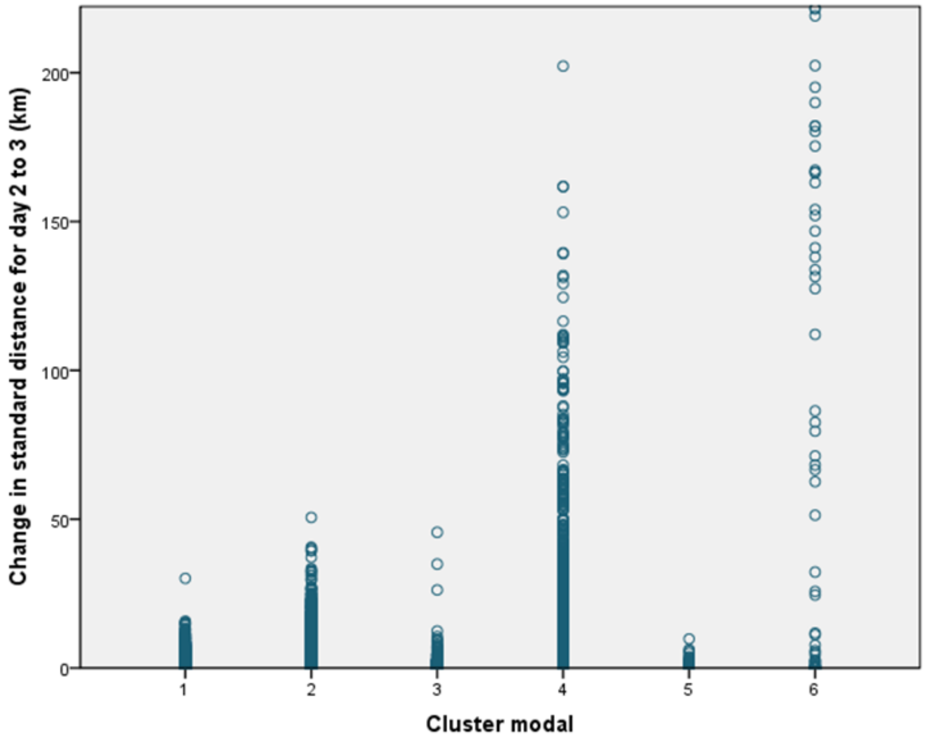


Figure 2-44: Change in Standard Distance for Day 1 to 3, by Cluster

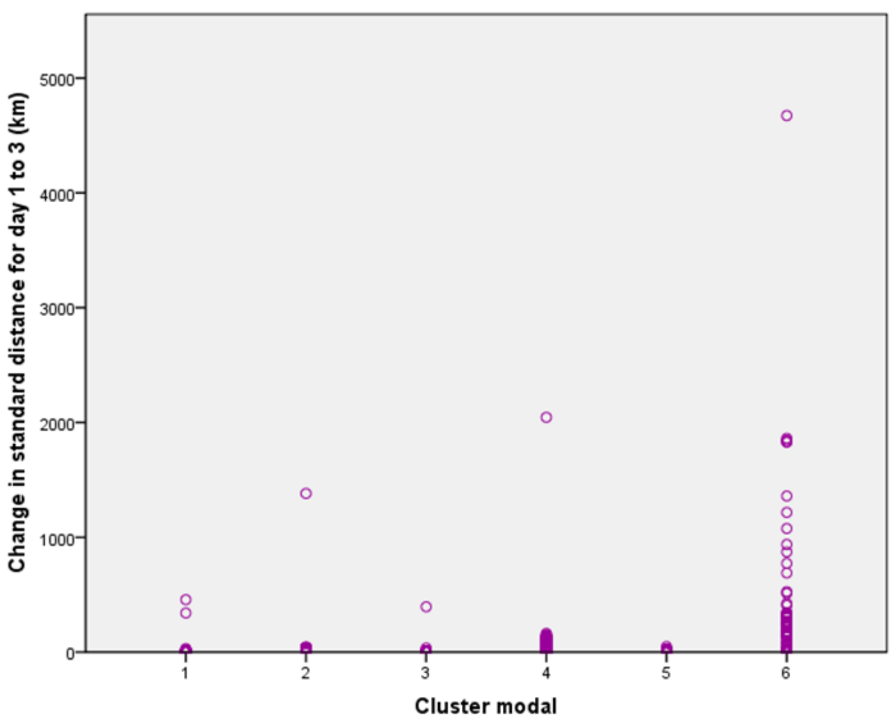
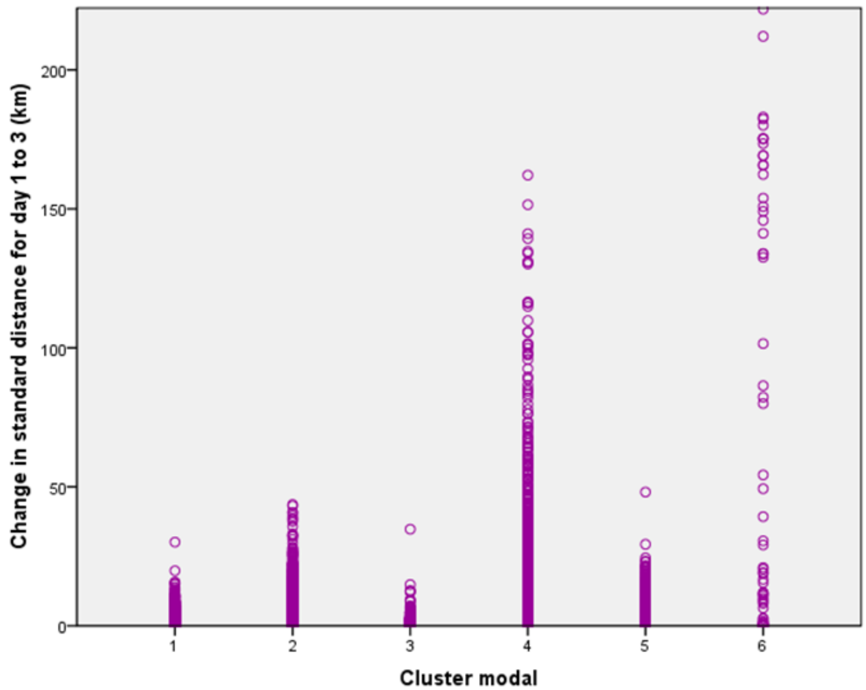


Figure 2-45: Change in Standard Distance for Day 1 to 3, by Cluster (with Rescaled Axis)



In addition to the preceding figures, Table 2-14 provides descriptive statistics of standard distance, by cluster.

Table 2-14: Descriptive Statistics of Standard Distance for Clusters 1 Through 6

	Cluster 1			Cluster 2			Cluster 3			Cluster 4			Cluster 5			Cluster 6		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Standard distance day 1 in kilometers	0.00	342.74	4.42	0.00	62.52	8.24	0.00	396.37	2.51	0.00	139.74	16.47	0.00	48.10	6.61	0.00	1,964.36	120.51
Standard distance day 2 in kilometers	0.00	2,089.17	5.33	0.00	473.33	8.57	0.00	45.60	2.17	0.00	210.13	17.33	0.00	27.10	1.01	0.00	1,664.83	148.78
Standard distance day 3 in kilometers	0.00	457.88	4.00	0.00	1,384.93	7.92	0.00	27.64	2.07	0.00	2,074.38	17.92	0.00	32.51	1.01	0.00	4,673.35	214.13
Change in standard distance for day 1-2 in km	0.00	2,088.92	3.62	0.00	462.44	6.61	0.00	395.78	1.16	0.00	209.12	22.42	0.00	48.10	6.47	0.00	1,928.73	237.20
Change in standard distance for day 2-3 in km	0.00	2,087.31	4.19	0.00	1,381.48	9.23	0.00	45.60	0.91	0.00	2,052.63	27.11	0.00	9.77	0.16	0.00	4,667.61	327.76
Change in standard distance for day 1-3 in km	0.00	457.37	3.23	0.00	1,381.98	8.64	0.00	395.03	1.18	0.00	2,045.38	25.18	0.00	48.10	6.54	0.00	4,673.35	313.41

Although differences and similarities from cluster to cluster in travel attributes can be identified at a glance, it is important to confirm these observations using statistical methods. For this reason, a one-way ANOVA was used. By using an ANOVA, the null hypothesis that there is no significant difference in attributes of standard distance and change in standard distance across members of different clusters can be rejected. As presented in Table 2-15, results of the one-way ANOVA allow for the definitive conclusion that this is certainly not the case, as indicated by the high F values with highly significant results for each day and change for raw standard distance values and cluster membership.

Table 2-15: Results of a One-Way ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Change in standard distance for day 1-2 in km	Between Groups	5,721,942.949	5	1,144,388.590	358.495	.000
	Within Groups	21,442,045.468	6,717	3,192.206		
	Total	27,163,988.417	6,722			
Change in standard distance for day 2-3 in km	Between Groups	10,956,135.850	5	2,191,227.170	297.640	.000
	Within Groups	49,450,533.421	6,717	7,361.997		
	Total	60,406,669.271	6,722			
Change in standard distance for day 1-3 in km	Between Groups	9,961,286.913	5	1,992,257.383	311.048	.000
	Within Groups	43,022,273.171	6,717	6,404.983		
	Total	52,983,560.084	6,722			
Standard distance day 1 in kilometers	Between Groups	1,447,242.980	5	289,448.596	198.856	.000
	Within Groups	9,777,061.234	6,717	1,455.570		
	Total	11,224,304.214	6,722			
Standard distance day 2 in kilometers	Between Groups	2,221,899.455	5	444,379.891	193.320	.000
	Within Groups	15,440,237.621	6,717	2,298.681		
	Total	17,662,137.077	6,722			
Standard distance day 3 in kilometers	Between Groups	4,601,991.960	5	920,398.392	153.963	.000
	Within Groups	40,154,634.568	6,717	5,978.061		
	Total	44,756,626.529	6,722			

Day-to-Day Variation in Activity Type at Destinations

In addition to the dispersion of activities across space, it is also important to examine the type of activities one conducts from day to day. To do this, GPS coordinates were used to enumerate unique destinations and to infer location type for activities. Google Places API was used to search for place information for each point location in the survey. Google Places API is a service that returns place information with the input of latitude and longitude coordinates. Google Places API returns attributes for each place, including the place type and the place name. The service allows users to select a search radius around coordinates to obtain place information. Any place within that search radius is returned, with each place's attribute values reported in an array.

For this research, place attributes were desired for all places that were not associated with a respondent's home, work, or school location. Within the survey, there were 107,911 activity locations of interest. Of the 107,911 destination points, 27,367 points were identified as home, 9,811 point were identified as work destinations, and 4,523 points were identified as school destinations using reported home, work, and school location within the survey. The remaining 66,210 destinations were investigated using the Google Places API. The Google Places API was run on these points with a search radius of 40 meters. The search radius was determined as an optimal radius after comparing possible radius lengths and the results they provided. It is important to determine the most optimal radius; one that is too small may not provide place information because there is nothing close enough to be included. Additionally, a radius that is too large may return too many places, and conclusions regarding activity type may be difficult to reach. After the information was retrieved from Google Places API, place information was parsed to provide data on activity types.

There are several aspects of the place type and activity type that must be addressed before further discussing the findings. First, of the 66,210 destination points, 22,658 were identified as having no place type other than "route" (which only signifies that the point is close to a road). After a qualitative analysis of these destination locations, it can be reasonably concluded that many of these destinations are in residential neighborhoods, where activities were likely conducted at another person's house. The importance of the search radius criterion is illustrated here, as a search radius that is set too small might incorrectly select points that should not be included in this category. In addition to this result, another result confounds the practicality of using GPS coordinates to obtain activity-type information. The resulting activity parsing revealed that many of the destination coordinates are associated with many establishments. In fact, of the 43,552 remaining destinations, zero had one place associated with it, 726 had two places, 7,975 had three places, 6,248 had four place, 4,536 had five place, and 24,067 had six or more places associated with it. After a manual analysis of destinations that are associated with large numbers of location types using Google Maps, several reasons for these associations can be surmised. First, there are instances where it is obvious that a respondent left the GPS logger in his or her vehicle. Because of this, it is difficult to attribute the activity to a specific type, unless all establishments were of the same type (e.g., all establishments nearby were restaurants). Second, some of the GPS coordinates are located on streets that are populated with several establishments along the street. Third, GPS coordinates are also located in large shopping malls or centers, where a person could have conducted a range of activity types. Figure 2-46 through Figure 2-49 provide examples of these types of situations in addition to an example of a residential location destination with no place information.

Figure 2-46: Example Situation #1

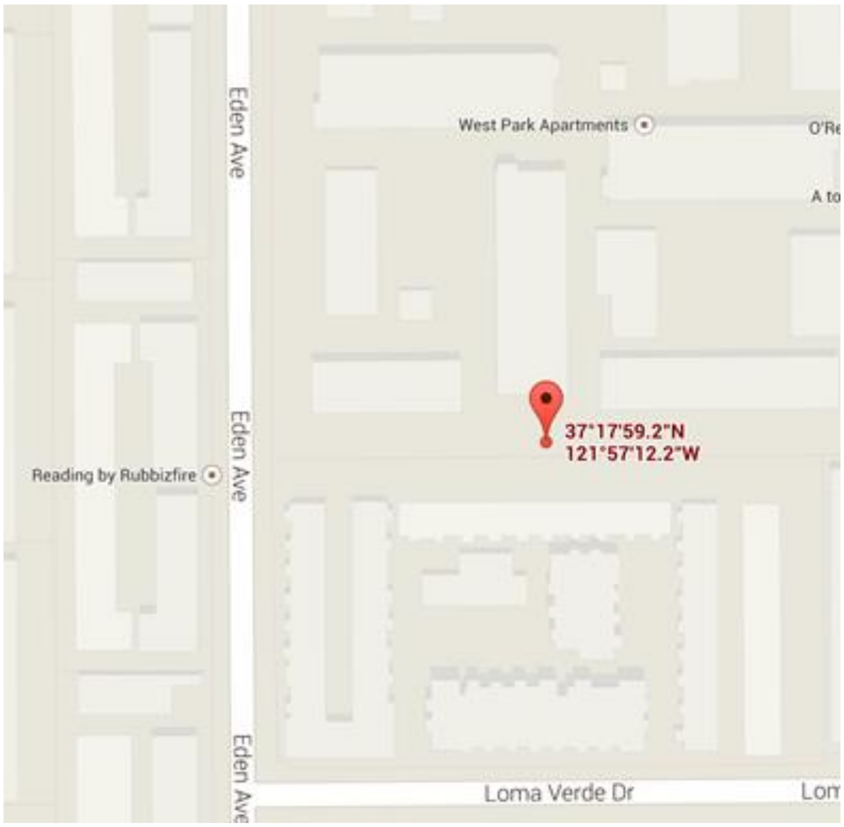


Figure 2-46 represents no associated place type and activity type.

Figure 2-47: Example Situation #2

Figure 2-47 represents a GPS coordinate that is within a clearly defined parking space, with many establishment types nearby.

Figure 2-48: Example Situation #3

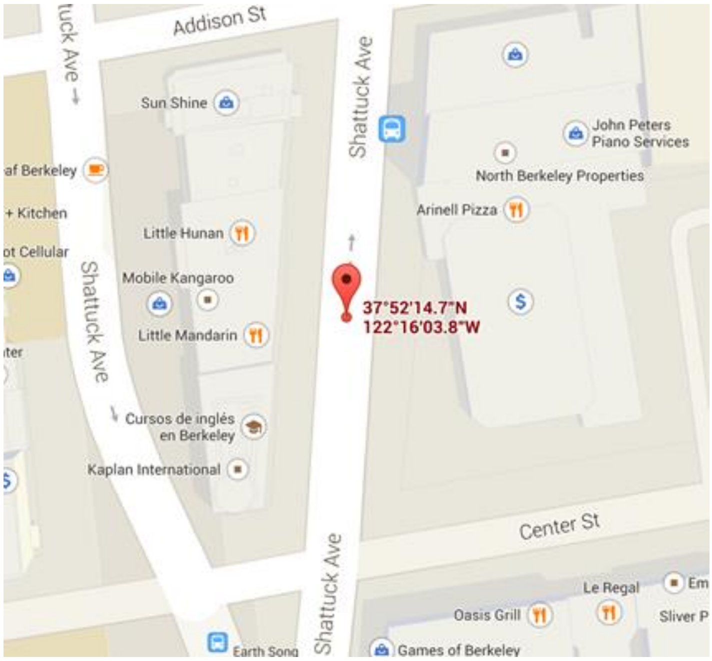


Figure 2-48 represents coordinates located on the street, with many nearby activity types (e.g., eating out, personal or household maintenance, personal business, shopping, etc.).

Figure 2-49: Example Situation #4

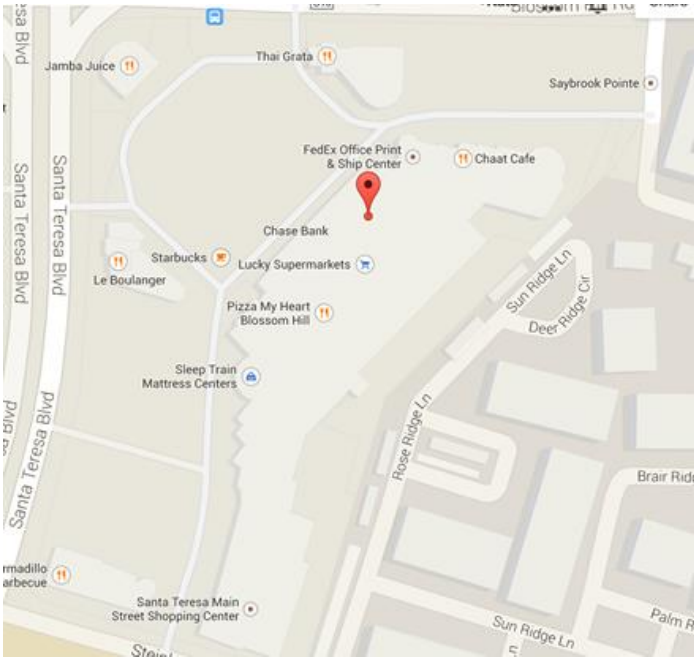


Figure 2-49 represents an instance where a respondent was inside of a building, which has many possible activity locations and types.

For this reason, destinations were only binned into five categories: home, work, school, residential other (for instance a friend's house), and other.

Using the standard-distance measures, a subset of respondents was selected for a second LCCA. This cluster analysis only included individuals with a standard distance across all 3 days that was equal to or less than 300 kilometers. This to eliminated any respondents who engaged in the longest-distance travel. In this way, smaller changes in behavior from day to day are not diluted by such drastic changes in behavior. The resulting sample size for this cluster model was 6,609 respondents. For this model, each of the five activity types were considered, and change from day to day was again calculated for each person. A description of these variables is provided in Table 2-16. Covariates were again used in the model specification to examine socioeconomic traits that are significant in explaining cluster membership; these are also included in this table.

Table 2-16: Description of Variables for Latent Class Cluster Model

Variables	Description
Indicators for latent classes all changes are absolute values	
Change in home destinations (for day one to two, day two to three and day three to one)	Absolute value (example number of destinations labeled "home" for day one- number of destinations labeled "home" for day two)
Change in work destinations (for day one to two, day two to three and day three to one)	Absolute value (example number of destinations labeled "home" for day one- number of destinations labeled "home" for day two)
Change in school destinations (for day one to two, day two to three and day three to one)	Absolute value (example number of destinations labeled "home" for day one- number of destinations labeled "home" for day two)
Change in residential other destinations (for day one to two, day two to three and day three to one)	Absolute value (example number of destinations labeled "home" for day one- number of destinations labeled "home" for day two)
Change in other destinations (for day one to two, day two to three and day three to one)	Absolute value (example number of destinations labeled "home" for day one- number of destinations labeled "home" for day two)
Covariates for class membership prediction	
Day one had no travel recorded	Binary indicator, 1 if no travel
Day two had no travel recorded	Binary indicator, 1 if no travel
Day three had no travel recorded	Binary indicator, 1 if no travel
Indicator for female	Binary indicator 1 if female 0 if male

Variables	Description
Indicator for age group 51 through 64	Binary indicator 1 if within age group
Indicator for age group 65 and older	Binary indicator, 1 if within age group
Employed	Binary indicator, 1 if employed
Retired	Binary indicator, 1 if retired
Home duties	Binary indicator, 1 if full-time home duties
Student	Binary indicator, 1 if student
Day one of survey was a work day	Binary indicator if day 1 was a work day
Day two of survey was a work day	Binary indicator if day 2 was a work day
Day three of survey was a work day	Binary indicator if day 3 was a work day
Indicator for high-income household	Binary indicator, income is $100,000 or more
Number of members in the household	Count variable ranging from 1 to 8
Number of employed persons in the household	Count variable ranging from 1 to 6

Table 2-17 provides descriptive statistics for each of the change indicators used to specify the model. The lowest means across the sample, in change from day to day, are for work destinations and school destinations. The highest change is for the "other" destination category, which is attributable to the size of this category.

Table 2-17: Descriptive Statistics for Change in Activity Types (n=6609)

	Minimum	Maximum	Mean	Std. Deviation
Change day 1 to 2 Home	.00	19.00	1.05	1.14
Change day 2 to 3 Home	.00	10.00	0.99	1.12
Change day 1 to 3 Home	.00	19.00	1.14	1.19
Change day 1 to 2 Work	.00	18.00	0.47	1.04
Change day 2 to 3 Work	.00	9.00	0.38	0.83
Change day 1 to 3 Work	.00	18.00	0.52	1.09
Change day 1 to 2 School	.00	12.00	0.23	0.76
Change day 2 to 3 School	.00	12.00	0.21	0.75
Change day 1 to 3 School	.00	12.00	0.25	0.83
Change day 1 to 2 Residential	.00	35.00	1.23	1.78
Change day 2 to 3 Residential	.00	34.00	1.19	1.84

	Minimum	Maximum	Mean	Std. Deviation
Change day 1 to 3 Residential	.00	35.00	1.33	2.02
Change day 1 to 2 Other	.00	27.00	2.13	2.38
Change day 2 to 3 Other	.00	27.00	2.06	2.43
Change day 1 to 3 Other	.00	27.00	2.22	2.51

The LCCA model involved an iterative procedure of estimating models with one through eight clusters. Review of the fit statistics for the models, model parsimony, and cluster composition indicates that a six-cluster model is the best model for representing the underlying latent phenomena within the data (-Log Likelihood = -119,645.95, BIC = 240,831.24, Classification error = 0.05). Although this model has the same number of clusters as the previous model, the similarities in cluster numbers are only linked because of the use of similar data. The results of the six-cluster model are provided in Table 2-18 (profile), Table 2-19 (probability means), and Table 2-20 (covariate significance and coefficients).

Table 2-18: LCCA Profile

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
	Cluster Size	0.34	0.24	0.14	0.13	0.08	0.06
Indicators	Change in home destinations day 1 to 2 Mean	0.93	0.77	1.34	1.26	1.51	1.13
	Change in home destinations day 2 to 3 Mean	0.95	0.85	1.50	1.34	0.00	1.15
	Change in home destinations day 1 to 3 Mean	0.98	0.85	1.47	1.46	1.51	1.27
	Change in work destinations day 1 to 2 Mean	0.00	1.27	0.51	0.17	0.57	0.30
	Change in work destinations day 2 to 3 Mean	0.00	1.15	0.43	0.16	0.00	0.29
	Change in work destinations day 1 to 3 Mean	0.00	1.43	0.60	0.18	0.57	0.35
	Change in school destinations day 1 to 2 Mean	0.00	0.04	0.06	1.52	0.14	0.12
	Change in school destinations day 2 to 3 Mean	0.00	0.05	0.07	1.43	0.00	0.10
	Change in school destinations day 1 to 3 Mean	0.00	0.04	0.07	1.68	0.14	0.12
	Change in other residential destinations day 1 to 2 Mean	0.85	0.75	1.83	1.11	1.10	4.38
	Change in other residential destinations day 2 to 3 Mean	0.90	0.78	1.94	1.12	0.00	4.42
	Change in other residential destinations day 1 to 3 Mean	0.92	0.75	2.00	1.11	1.10	5.21
	Change in other destinations day 1 to 2 Mean	1.69	1.50	4.78	1.87	2.70	0.61
	Change in other destinations day 2 to 3 Mean	1.80	1.55	5.45	1.95	0.00	0.60
	Change in other destinations day 1 to 3 Mean	1.68	1.48	5.35	2.03	2.70	0.61

			Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Covariates	Day one had no travel	0	0.86	0.98	0.93	0.97	1.00	0.95
		1	0.14	0.02	0.07	0.03	0.00	0.05
	Day two had no travel	0	0.89	0.94	0.90	0.91	0.00	0.88
		1	0.11	0.06	0.10	0.09	1.00	0.12
	Day three had no travel	0	0.83	0.82	0.78	0.75	0.00	0.78
		1	0.17	0.18	0.22	0.25	1.00	0.22
	Gender is female	0	0.47	0.53	0.50	0.40	0.49	0.55
		1	0.53	0.47	0.50	0.60	0.51	0.45
	Age 51 to 64	0	0.58	0.56	0.62	0.86	0.67	0.64
		1	0.42	0.44	0.38	0.14	0.33	0.36
	Age is 65 or older	0	0.81	0.91	0.89	0.95	0.85	0.81
		1	0.19	0.09	0.11	0.05	0.15	0.19
	Employed full time	0	0.44	0.03	0.26	0.47	0.36	0.34
		1	0.56	0.97	0.74	0.53	0.64	0.66
	Retired	0	0.81	0.99	0.91	0.97	0.90	0.86
		1	0.19	0.01	0.09	0.03	0.10	0.14
	Full-time home duties	0	0.93	0.99	0.96	0.87	0.94	0.96
		1	0.07	0.01	0.04	0.13	0.06	0.04
	Student status	0	0.93	0.99	0.95	0.76	0.93	0.93
		1	0.07	0.01	0.05	0.24	0.07	0.07
	Survey day one was a workday	0	0.59	0.10	0.41	0.58	0.46	0.45
		1	0.41	0.90	0.59	0.42	0.54	0.55
	Survey day two was a work day	0	0.56	0.16	0.47	0.60	1.00	0.53
		1	0.44	0.84	0.53	0.40	0.00	0.47
	Survey day three was a workday	0	0.65	0.47	0.70	0.74	1.00	0.67
		1	0.35	0.53	0.30	0.26	0.00	0.33
	Income higher than $100,000 /year	0	0.56	0.42	0.56	0.43	0.53	0.56
		1	0.44	0.58	0.44	0.57	0.47	0.44
	Household size	1	0.10	0.11	0.11	0.00	0.10	0.10
		2	0.40	0.37	0.31	0.04	0.28	0.38
		3	0.22	0.22	0.23	0.23	0.21	0.19
		4	0.19	0.21	0.22	0.43	0.26	0.21
		5 to 8	0.09	0.09	0.12	0.29	0.15	0.12
		Mean	2.82	2.84	2.98	4.12	3.17	2.96
	Number of employees in the household	0-1	0.14	0.00	0.07	0.04	0.11	0.15
		2	0.40	0.32	0.35	0.37	0.32	0.31
		3	0.35	0.53	0.42	0.50	0.40	0.43
		4 to 6	0.10	0.15	0.16	0.09	0.17	0.11
		Mean	1.44	1.88	1.72	1.66	1.67	1.52

Table 2-19: Probability Means for Six-Cluster Model

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
	Overall	0.34	0.24	0.14	0.13	0.08	0.06
Indicators							
Change in home destinations day 1 to 2	*0*	0.38	0.30	0.12	0.11	0.03	0.05
	1	0.35	0.25	0.13	0.11	0.10	0.06
	2 to 19	0.29	0.14	0.21	0.16	0.13	0.07
Change in home destinations day 2 to 3	*0*	0.31	0.24	0.09	0.09	0.21	0.05
	1	0.40	0.28	0.14	0.12	0.00	0.06
	2 to 10	0.31	0.18	0.24	0.19	0.00	0.08
Change in home destinations day 1 to 3	*0*	0.40	0.30	0.11	0.10	0.04	0.06
	1	0.34	0.25	0.13	0.12	0.09	0.06
	2 to 19	0.28	0.16	0.20	0.16	0.12	0.07
Change in work destinations day 1 to 2	*0*	0.48	0.09	0.14	0.15	0.08	0.07
	1 to 18	0.00	0.63	0.16	0.06	0.10	0.05
Change in work destinations day 2 to 3	*0*	0.45	0.09	0.14	0.14	0.11	0.06
	1 to 9	0.00	0.72	0.16	0.07	0.00	0.05
Change in work destinations day 1 to 3	*0*	0.50	0.07	0.13	0.15	0.08	0.06
	1	0.00	0.62	0.16	0.08	0.08	0.07
	2 to 18	0.00	0.64	0.18	0.03	0.11	0.03
Change in school destinations day 1 to 2	*0*	0.39	0.27	0.16	0.03	0.09	0.06
	1 to 12	0.00	0.07	0.06	0.76	0.06	0.05
Change in school destinations day 2 to 3	*0*	0.39	0.26	0.15	0.04	0.09	0.06
	1 to 12	0.00	0.09	0.07	0.79	0.00	0.05
Change in school destinations day 1 to 3	*0*	0.40	0.27	0.16	0.03	0.09	0.06
	1 to 12	0.00	0.07	0.06	0.76	0.06	0.04
Change in other residential destinations day 1 to 2	*0*	0.38	0.30	0.09	0.12	0.09	0.01
	1	0.38	0.26	0.13	0.13	0.08	0.03
	2	0.36	0.20	0.19	0.13	0.07	0.05
	3 to 35	0.16	0.09	0.27	0.11	0.08	0.28

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Change in other residential destinations day 2 to 3	*0*	0.34	0.27	0.08	0.11	0.19	0.01
	1	0.42	0.27	0.14	0.14	0.00	0.03
	2	0.37	0.24	0.18	0.15	0.00	0.06
	3 to 34	0.20	0.10	0.31	0.11	0.00	0.28
Change in other residential destinations day 1 to 3	*0*	0.38	0.30	0.09	0.13	0.09	0.01
	1	0.38	0.26	0.12	0.14	0.08	0.02
	2	0.36	0.23	0.18	0.12	0.07	0.04
	3 to 35	0.19	0.07	0.28	0.10	0.08	0.28
Change in other destinations day 1 to 2	*0*	0.37	0.26	0.04	0.12	0.06	0.16
	1	0.35	0.32	0.08	0.14	0.08	0.04
	2	0.41	0.26	0.10	0.12	0.08	0.04
	3	0.38	0.23	0.14	0.13	0.09	0.03
	4 to 27	0.23	0.13	0.41	0.10	0.13	0.01
Change in other destinations day 2 to 3	*0*	0.25	0.21	0.02	0.11	0.27	0.13
	1	0.43	0.34	0.05	0.14	0.00	0.04
	2 to 3	0.45	0.27	0.10	0.15	0.00	0.04
	4 to 27	0.26	0.13	0.49	0.12	0.00	0.01
Change in other destinations day 1 to 3	*0*	0.36	0.28	0.04	0.11	0.06	0.16
	1	0.39	0.30	0.05	0.14	0.07	0.04
	2	0.39	0.28	0.08	0.14	0.08	0.04
	3	0.38	0.23	0.12	0.16	0.09	0.03
	4 to 27	0.22	0.12	0.43	0.11	0.12	0.01
Covariates							
Day one had no travel	*0*	0.32	0.26	0.14	0.13	0.09	0.06
	1	0.69	0.08	0.15	0.05	0.00	0.04
Day two had no travel	*0*	0.37	0.28	0.16	0.14	0.00	0.06
	1	0.23	0.08	0.08	0.06	0.50	0.05
Day three had no travel	*0*	0.39	0.27	0.15	0.13	0.00	0.06
	1	0.22	0.16	0.12	0.12	0.32	0.05

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Gender is female	*0*	0.33	0.26	0.15	0.10	0.08	0.07
	1	0.35	0.22	0.14	0.15	0.08	0.05
Age 51 to 64	*0*	0.32	0.22	0.14	0.17	0.09	0.06
	1	0.39	0.29	0.15	0.05	0.07	0.06
Age is 65 or older	*0*	0.32	0.26	0.15	0.14	0.08	0.06
	1	0.49	0.16	0.12	0.05	0.09	0.09
Employed full time	*0*	0.49	0.03	0.12	0.19	0.10	0.07
	1	0.28	0.34	0.15	0.10	0.08	0.06
Retired	*0*	0.31	0.27	0.15	0.13	0.08	0.06
	1	0.65	0.02	0.13	0.04	0.08	0.08
Full-time home duties	*0*	0.34	0.25	0.15	0.11	0.08	0.06
	1	0.42	0.02	0.10	0.31	0.10	0.04
Student status	*0*	0.35	0.26	0.15	0.10	0.08	0.06
	1	0.33	0.03	0.09	0.41	0.08	0.06
Survey day one was a workday	*0*	0.48	0.06	0.14	0.17	0.09	0.06
	1	0.25	0.38	0.15	0.09	0.08	0.06
Survey day two was a work day	*0*	0.39	0.08	0.14	0.15	0.17	0.07
	1	0.30	0.40	0.15	0.10	0.00	0.06
Survey day three was a workday	*0*	0.34	0.17	0.15	0.14	0.13	0.06
	1	0.35	0.38	0.13	0.09	0.00	0.06
Income higher than $100,000 /year	*0*	0.38	0.20	0.16	0.11	0.09	0.07
	1	0.31	0.29	0.13	0.14	0.08	0.05
Household size	*1*	0.38	0.29	0.18	0.00	0.09	0.06
	2	0.43	0.27	0.14	0.01	0.07	0.07
	3	0.34	0.25	0.15	0.13	0.08	0.05
	4	0.27	0.22	0.13	0.23	0.09	0.05
	5 to 8	0.25	0.17	0.14	0.29	0.10	0.06
Number of employees in the household	*0-1*	0.59	0.00	0.13	0.06	0.11	0.11
	2	0.39	0.21	0.14	0.13	0.07	0.05

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
	3	0.28	0.30	0.14	0.14	0.08	0.06
	4 to 6	0.28	0.29	0.18	0.09	0.11	0.05

Table 2-20: Covariate Coefficients and Significance

Covariates		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Wald	p-value
Day one had no travel	0	-0.40	-0.09	-0.14	0.60	-0.05	0.08	73.91	0.00
	1	0.40	0.09	0.14	-0.60	0.05	-0.08		
Day two had no travel	0	0.97	1.05	1.04	1.12	-5.13	0.95	12.58	0.03
	1	-0.97	-1.05	-1.04	-1.12	5.13	-0.95		
Day three had no travel	0	1.02	1.02	1.02	0.86	-4.86	0.92	16.23	0.01
	1	-1.02	-1.02	-1.02	-0.86	4.86	-0.92		
Gender is female	0	0.00	0.04	0.04	-0.18	-0.01	0.12	24.25	0.00
	1	0.00	-0.04	-0.04	0.18	0.01	-0.12		
Age 51 to 6	0	-0.19	-0.08	0.00	0.32	0.03	-0.08	62.48	0.00
	1	0.19	0.08	0.00	-0.32	-0.03	0.08		
Age is 65 or older	0	-0.06	0.08	0.20	0.41	-0.51	-0.13	34.36	0.00
	1	0.06	-0.08	-0.20	-0.41	0.51	0.13		
Employed full time	0	0.14	-0.64	-0.22	-0.37	1.29	-0.19	45.58	0.00
	1	-0.14	0.64	0.22	0.37	-1.29	0.19		
Retired	0	-0.34	-0.04	-0.26	-0.14	1.02	-0.25	7.00	0.22
	1	0.34	0.04	0.26	0.14	-1.02	0.25		
Home duties	0	-0.13	0.02	0.01	-0.35	0.40	0.05	8.51	0.13
	1	0.13	-0.02	-0.01	0.35	-0.40	-0.05		
Student status	0	-0.20	-0.17	0.00	-0.76	1.39	-0.27	43.56	0.00
	1	0.20	0.17	0.00	0.76	-1.39	0.27		
Survey day one was a workday	0	0.36	-0.36	0.07	0.31	-0.40	0.02	85.03	0.00
	1	-0.36	0.36	-0.07	-0.31	0.40	-0.02		
Survey day two was a work day	0	-0.10	-0.15	0.12	0.09	-0.19	0.22	21.04	0.00
	1	0.10	0.15	-0.12	-0.09	0.19	-0.22		

Covariates		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Wald	p-value
Survey day three was a workday	0	-0.17	0.08	0.28	0.01	-0.28	0.07	62.27	0.00
	1	0.17	-0.08	-0.28	-0.01	0.28	-0.07		
Income higher than $100,000 /year	0	0.05	-0.06	0.14	-0.15	-0.08	0.10	38.59	0.00
	1	-0.05	0.06	-0.14	0.15	0.08	-0.10		
Household size		-0.14	-0.19	-0.13	0.51	-0.03	-0.02	280.73	0.00
Number of employees in the household		0.02	0.17	0.21	-0.30	0.00	-0.09	53.37	0.00

The preceding tables illustrate notable differences between clusters in the types of change observed and important sociodemographic indicators. In addition to these tables, Table 2-21 provides a qualitative rendering of the mean values of cluster members for each indicator (as reported in the top portion of Table 2-21) for easier interpretation and comparison across clusters and days. As in the similar figure presented for the first cluster analysis, comparisons of colors should not be made across indicator types, as each indicator type has its own scaling for color rendering. Values of cluster means for each attribute are again included for reference in the figure.

Table 2-21: Qualitative Comparison of Mean Values for Cluster Indicators

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Change in home destinations day 1 to 2 Mean	0.93	0.77	1.34	1.26	1.51	1.13
Change in home destinations day 2 to 3 Mean	0.95	0.85	1.50	1.34	0.00	1.15
Change in home destinations day 1 to 3 Mean	0.98	0.85	1.47	1.46	1.51	1.27
Change in work destinations day 1 to 2 Mean	0.00	1.27	0.51	0.17	0.57	0.30
Change in work destinations day 2 to 3 Mean	0.00	1.15	0.43	0.16	0.00	0.29
Change in work destinations day 1 to 3 Mean	0.00	1.43	0.60	0.18	0.57	0.35
Change in school destinations day 1 to 2 Mean	0.00	0.04	0.06	1.52	0.14	0.12
Change in school destinations day 2 to 3 Mean	0.00	0.05	0.07	1.43	0.00	0.10
Change in school destinations day 1 to 3 Mean	0.00	0.04	0.07	1.68	0.14	0.12
Change in other residential destinations day 1 to 2 Mean	0.85	0.75	1.83	1.11	1.10	4.38
Change in other residential destinations day 2 to 3 Mean	0.90	0.78	1.94	1.12	0.00	4.42
Change in other residential destinations day 1 to 3 Mean	0.92	0.75	2.00	1.11	1.10	5.21
Change in other destinations day 1 to 2 Mean	1.69	1.50	4.78	1.87	2.70	0.61
Change in other destinations day 2 to 3 Mean	1.80	1.55	5.45	1.95	0.00	0.60
Change in other destinations day 1 to 3 Mean	1.68	1.48	5.35	2.03	2.70	0.61

COLOR SCALE low high

Cluster 1: The Low-Variation, Errand Runners

Cluster 1 members have, on average, less than one trip change from day to day for every activity type except for "other." For this reason, it is thought that many of these changes in destinations are a result of the addition or subtraction of errands, eating meals out, etc. The members of cluster 1 also have, on average, zero change in the number of work and school destinations from day to day. Upon further investigation, a large part of cluster 1 membership does not have work trips (approximately 2086 of the 2334) but the other members do. These respondents do not change their work destination patterns across days, however. Because of the high presence of people with zero work trips over the three-day survey period, it makes sense that people who are employed have a lower probability of belonging to cluster 1. Additionally, students and those who are 51 to 64 or 65 and older have a higher likelihood of being members of cluster 1.

Cluster 2: The Workers with One Nonwork Day

Cluster 2 members have the lowest change in home as a destination across all clusters. Additionally, they also have low changes in school destinations. Cluster 2 members do have higher changes from day to day in the frequency of work as a destination and the other category destinations. It is thought that these respondents most likely replace work trips with other discretionary trips on the day that was included in the survey in which they are off from work. Cluster 2 members are less likely to have a day where there is no travel recorded, and have a higher likelihood of belonging if they are employed or are students. Also interesting is the positive effect of number of employees per household. A higher number for employees per household is correlated positively to cluster 2 membership. Cluster membership is also positively impacted by the presence of people who had workdays on day 1 and day 2 of the survey period.

Cluster 3: The Post-Work Activities

Cluster 3 members have the highest change in "other" destinations (means of 4.78 to 5.45 compared to the next highest of a mean of 2.70 trip difference from day to day). These members are also second highest across all clusters for the change in the frequency of residential-related destinations. This cluster has low changes in work and school destinations, however. For this reason, it is thought that these respondents conduct activities after work that are likely to change from one day to the next. Respondents who recorded no travel (for day 2 and 3 especially) are less likely to belong to cluster 3. Those who are employed and who have higher numbers of employees in the house are more likely to belong to cluster 3. Respondents with high income and those with larger households are less likely to be members of cluster 3.

Cluster 4: The Active Students and Young Professionals

Cluster 4 members have a consistent change in frequency of destinations from day to day for almost all activity types. With the exception of work, cluster 4 members have a change of between 1.11 and 2.03 across all days and activity types. Cluster 4 members are more likely to be students. For these reasons, it is believed that this cluster is comprised mainly of students who make many different trips from day to day due to a highly flexible schedule, and younger professionals who work all 3 days and conduct activities after work. Additionally, workers are also more likely to belong to cluster 4. Cluster 4 members are less likely to be older (age 51 and older). Females have a higher likelihood of belonging to cluster 4. Additionally, cluster 4 members tend to have a higher income and household size. In fact, cluster 4 has the highest mean household size across all six clusters (4.12 persons).

Cluster 5: The Periodic Nontravelers

Cluster 5 members are unique in the change behavior observed. On average, cluster 5 members have no change in destinations for any activity type from day 2 to day 3. This cluster, like the previous cluster analysis on change in trip attributes, mainly consists of individuals who did not record travel on day 2 and day 3 of the survey. They also have high changes in the number of "other" activities for day 1 to 2 when compared to other clusters. Those who are employed or are students are much less likely to belong to cluster 5, while those who are 65 and older are more likely to belong to cluster 5.

Cluster 6: Residential Visitors

Cluster 6 members have the highest mean change for frequency of residential destinations across all clusters. These individuals also have low changes for "other," work, and school destinations. Cluster membership is positively correlated with being a student or being employed full time. Being a female from a high-income household, larger household size, or having work days during the survey period are negatively correlated with cluster 6 membership. This cluster could possibly be composed of both students who might go to friends' houses often and people who work in the service sector making house calls.

2.6 Discussion

The development of clusters reveals different groups of people based on variability types. The use of attributes of change in travel behavior shows that there are people who have small variation in their day-to-day behavior and people who have large variations. Those who have large variations have multiple reasons for these variations, including long-distance international, national, or regional travel. In addition to these extremes, there are also those who have moderate levels of change in their behavior from day to day. The latent clusters that underlie behavioral similarities of change can be further explained with sociodemographic variables as correlates. Additionally, travel behavior data for each day, rather than change in behavior, can also be used to enhance the understanding of clusters.

Through all of these analyses, there are several notable findings. In the first cluster analysis, which was based solely on change in trip attributes across the three-day survey period, it was notable that cluster 1 members tend to have local travel only (and lower amounts of change when compared across clusters), but are in the middle when comparing clusters that have members with more-focused local travel. Cluster 1 members have a wider distribution of total number of trips than other clusters (specifically 4, 5, and 6). Cluster 2 comprises many people who have regional travel (although it also includes some national travelers). There are higher total distances traveled for cluster 2 members cumulatively across days and for each day when compared to other clusters; however, these averages are not the highest. Cluster 2 members, similar to cluster 1 members, have a higher distribution of total trips when compared to other clusters. Cluster 3 members travel primarily in a local setting. When compared to cluster 1, cluster 3 members have lower means for total distance traveled and number of trips, and average distance per trip. The changes from day to day for cluster 3 members are the lowest changes across all clusters for all 3 days (with the exception of day 2 to 3 for cluster 5). Cluster 4 members have travel that spans across a majority of the State of California. As expected, the members of this cluster have higher changes in total distances and average distances per trip. Additionally, cluster 4 members have higher changes in total number of trips than other clusters. Cluster 5 members also have primarily local travel. One aspect of cluster 5 that differentiates it from other clusters is the fact that there are a high number of members who have one or more days without travel. Because of this, there are high changes in total number of trips and distances from day to day. Cluster 6 members have high changes in travel attributes from day to day. The members of this cluster have the largest distribution of distances. When examining the geographic distribution of activity locations, it becomes apparent that many of the members of this cluster have long-distance trips (nationally or internationally). Because of these long-distance trips, these individuals have large changes in their travel from day to day. Some notable sociodemographic indicators that are significant reveal that those who are employed full time are more likely to belong to cluster 2 or 4. Women are more likely to belong to cluster 5 and cluster 2. Retired persons are more likely to belong to cluster 4. Individuals who do not have a fixed location for work are more likely to belong to cluster 6. Changes were also examined geographically using a standard distance metric. The analysis of the change in the geographic density of activity locations solidifies conclusions made through the visual inspection of maps created of activity location, by cluster. Cluster 6 has a much wider distribution of standard distance, or the distribution of activity locations, while clusters 1, 3, and 5 have the lowest. Cluster 2 shows larger standard distance than clusters 1, 3, and 5, but less than cluster 4, highlighting the difference between regional and statewide travel.

The use of standard distances reiterates the variation in changes that exist across days for individuals in a sample. By using standard distance as a selection criterion, a smaller subset of respondents was selected to analyze activity behavior and change in frequencies of activity types. Results of a second cluster analysis show that there are again distinct types of variation in day-to-day behavior, this time highlighting changes in activity. For instance, sociodemographics used as covariates show that age is correlated with variability type. This result might be expected due to the likelihood of older individuals being retired and having a more flexible schedule. Drastic changes in day-to-day behavior (such as cluster 5 in both models) illustrate the range of behavioral variation that exists. It has been shown through the use of covariates that those who are employed or are students are less likely to have the type of variation that results from many days with no travel.

2.7 Conclusions

The preceding analysis illustrates the necessity of examining both interperson and intraperson variation in daily behavior. The development of clusters based on change from one day to another in the survey permits development of different variability types. These variability types distinguish between local, regional, national, and international travel when using attributes of trips. Although the distance of trips and activity locations was revealed as a large contributor to the development of different variation types, there are also three clusters in the first cluster model that highlight differences in behavior within a local travel context. These three clusters highlight the differences between those who have days with no travel, those who have little change, and those who have larger change in day-to-day behavior. In addition, the variability types are correlated with spatial attributes of travel behavior, as shown using standard distance measures. A comparison of clusters in a geographic context furthers understanding of the latent variables that manifest in observed behavior. In addition, GPS data allow for a deeper investigation of behavior types by using place information to provide activity type and time use context.

While this chapter's research contributes to the body of knowledge regarding interperson and intraperson variation, there are a few caveats and areas for further improvement and future research. First, the designation of activity type was necessarily course. The nature of many of the GPS points collected during the survey limited the ability of any geocoding process to correctly attribute an activity location to a respondent's destination. Although it was attempted in this research to use a probabilistic mechanism to select the most likely destination and attach detailed place information to GPS coordinates, this attempt was not used in any analysis. After a manual inspection of the results of this probabilistic assignment process, it was concluded that the amount of error that would be entered into any statistical analysis was greater than any conclusion that could possibly be reached from analysis. Further refinement of a probabilistic selection of activity type could include time-of-day and hazard modeling to select a likely destination based on the larger context of the activity. This was not attempted. Additionally, this research did not investigate time use associated with each destination. This is an area for continued research, as the duration of an activity type and total amount of time used from day to day for different activity types has a large influence on day-to-day variation. In addition to these areas of further research, lessons learned from behavioral analysis should be applied to forecasting and travel demand modeling, where practicable. Operationalizing the findings and applying them in current modeling efforts is necessary if modeling sophistication is to increase.

These findings can be used in two ways. First, results can be used to inform surveys and exploratory analyses to develop theoretical models of intraperson variation. Although the existence of latent factors that are responsible for differences among individuals in intraperson variation are acknowledged and modeled, the "why" cannot be known until the question is asked and research is undertaken. Designing and implementing surveys to understand the psychological, physical, and social needs related to variation will better accommodate the reality of behavior in models. Second, the results of this and other similar studies can be used to further refine the random error term in current models. Models can be specified to better include heterogeneity across people and within an individual. Third, this chapter's finding and the findings of others add to the surmounting evidence that a one-day travel diary is not sufficient to accurately model some aspects of behavior. Adding support and evidence to this claim aids in bringing the need of data spanning a longer time frame to the forefront of discussions in large-scale data collection projects. The appropriate length of surveys and data collection efforts will likely always be a topic of debate. With limits on resources, there are tradeoffs between level of detail collected during the survey and length of time or duration of data collection. The research presented here illustrates the need to capture variation across days, as individuals do not vary in travel behavior from day to day in a homogenous manner. These variations were observable across a three-day period; however, it is likely that a 5- or 7-day survey would reveal even more similarities and differences among respondents. One aspect of this research that could be further explored is the impact of work versus nonwork days and week versus weekend travel behavior. A full-week survey would have allowed this comparison more completely.

As mentioned at the beginning of this chapter, this work was conducted through the Transportation Secure Data Center (TSDC), hosted by the National Renewable Energy Laboratory (NREL). Several household travel surveys are hosted on this server, including the California Household Travel Survey used in this analysis. The secure portal environment is well equipped with software necessary for analysis. Additionally, the environment is intuitively organized, and tutorials and examples are included that have been created using datasets hosted on the data center. The members of the NREL data center support team are quick to respond to any queries and research needs, and are knowledgeable and efficient. Several aspects of this chapter's study required additional resources outside of the TSDC environment (such as the use Google Places API). This was accomplished with additional approval and required complete anonymization of the data. For Google Places, coordinate pairs of a location were exported without any additional information in order to obtain place information. Place names and types were obtained outside of the TSDC and then joined to the database within the secure environment for further analysis. Although this is more involved than working solely on one's own personal computer, the process caused minimal interruption in this research's workflow.

2.8 Acknowledgements

The author would like to thank Michelle Lee and Mark Bradley for feedback and support in this research project, the Federal Highway Administration and Elaine Murakami for financial support and project guidance, and the NREL, TSDC (Evan Burton and Jeff Gonder in particular), for technical support and responsiveness in system administration for this research.

2.9 References

1. Auld, J., Mohammadian, A. and Doherty, S. (2008). Analysis of Activity Conflict Resolution Strategies. In *Transportation Research Record: Journal of the Transportation Research Board*, 2054, 10-19.
2. Auld J., Williams C., Mohammadian, A, and Nelson, P. (2009). An Automated GPS-based Prompted Recall Survey With Learning Algorithms. In *Transportation Letters*, 1, 1, 59-79.
3. Axhausen KW, Zimmermann A, Schönfelder S, Rindsfüser G & Haupt T (2000) Observing the rhythms of daily life: A six-week travel diary. *Transportation* 29, 2, 95–124.
4. Buliung, R. N., Roorda, M. J., & Remmel, T. K. (2008). Exploring spatial variety in patterns of activity-travel behaviour: initial results from the Toronto Travel-Activity Panel Survey (TTAPS). *Transportation*, *35*(6), 697-722.
5. Chen, C., Garling, T., and Kitamura, R. (2004). Activity rescheduling: reasoned or habitual? In *Transportation Research Part F*, 7, 351-371.
6. Clark, A. and Doherty, S. (2008). Examining the Nature and Extent of the Activity-Travel Preplanning Decision Process. In *Transportation Research Record: Journal of the Transportation Research Board*, 2054, 83-92.
7. Elango, V., Guensler, R. and Ogle, J. (2007). Day-to-day Variability in the Commute Atlanta Study. Presented at the Annual Meeting of the Transportation Research Board, January 2007, Washington D.C.
8. 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, 31-48.
9. Hanson S., and P. Hanson (1981). The Travel-Activity Patterns of Urban Residents: Dimensions and Relationships to Sociodemographic Characteristics. *Economic Geography*, 57,332-47

10. Hanson, S & J.O. Huff (1982). Assessing day-to-day variability in complex travel patterns. *Transportation Research Record* 891: 18-24.
11. Hanson, S. & J.O. Huff (1986). Classification issues in the analysis of complex travel behavior. *Transportation* 13, 271-293
12. Hanson, S and J.O. Huff (1988). Systematic Variability in repetitious travel. *Transportation* 15, 111-135.
13. Hatcher S.G, and Mahmassani H., (1992). Daily Variability of Route and Trip Scheduling Decisions for the Evening Commute. Transportation Research Record, 1357: p. 72-81.
14. Hirsh, M., Prashker, J. and M. Ben-Akiva (1986). Dynamic model of weekly activity pattern. *Transportation Science* 20, 24-36
15. Huff, J.O. & S. Hanson (1986). Repetition and variability in urban travel. *Geographical Analysis* 18, 97-114.
16. Kitamura, R. (1987). An analysis of weekly activity patterns and travel expenditure, in Golledge, R. and H. Timmermans (Eds.) Behavioral modeling approaches in geography and planning. Croom Helm, London, 399-423.
17. Li H, G.R., and Ogle J, (2005). Analysis of Morning Commute Route Choice Patterns using Global Positioning System-Based Vehicle Activity Data. Transportation Research Record, 1926: p. 162-170.
18. Ma, J. and K. G. Goulias (1997) A dynamic analysis of person and household activity and travel patterns using data from the first two waves in the Puget Sound Transportation Panel. *Transportation* 24, 309-331.
19. Mohammadian, K. and Doherty, S. (2005). Mixed Logit Model of Activity-Scheduling Time Horizon Incorporating Spatial-Temporal Flexibility Variables *Transportation Research Record: Journal of the Transportation Research Board*, No. 1926, pp. 33–40.
20. NuStats (2013) 2010-2012 California Household Travel Survey Final Report: *Version 1.0. June 14. Submitted to the California Department of Transportation. Austin, TX.*
21. Pas, E. (1988) Weekly travel-activity behavior. *Transportation* 15, 89-109.
22. Pendyala, R.M. and E.I. Pas (2000) Multiday and Multiperiod Data for Travel Demand Modeling. Invited Resource Paper in Transport Surveys: Raising the Standard, Proceedings of an International Conference on Transport Survey Quality and Innovation. Transportation Research Board E-Circular Number E-C008, Transportation Research Board, National Research Council, pp. II-B/1 II-B/22

23. Puget Sound Regional Council. Puget Sound Transportation Panel 1989-2002. Summary Document http://www.psrc.org/assets/1484/PSTP_summary.pdf Accessed September, 2014.

24. Schlich, R., Axhausen, K.: Habitual travel behaviour: evidence from a six-week travel diary. Transportation 30, 13–36 (2003)

25. Shapcott, M. (1978). Comparison of the use of time in Reading, England with time use in other countries. *Transactions of the Martin Centre for Architectural and Urban Studies* 3, 231-257

26. Stopher P.R. and Zhang Y. (2010). Is Travel Behaviour Repetitive from Day to Day? Proceedings of the 33rd Australasian Transport Research Forum ATRF 2010, Canberra, Australia, 1st October 2010

27. Susilo, Y.O. and RK. Kitamura (2005). An analysis of the day-to-day variability in the individual's action space: An exploration of the six-week Mobidrive travel diary data, *Transportation Research Record*, 1902, pp. 124 – 133

28. Vermunt, J.K. and Magidson, J. (2005). Technical guide for latent gold 4.0: basic and advanced. Statistical Innovations Inc., Belmont, MA.

29. Viti, F., C.M.J. Tampère, F. Walle & M. Castaigne (2010), Analyzing weekly activity-travel behavior from behavioral survey and traffic data, Proceedings of 12th World Conference on Transport Research, 11-15 July, 2010, Lisbon.

Chapter 3.0 Capturing Personal Modality Styles Using Multiday GPS Data—Findings from the San Francisco Bay Area

By Yanzhi “Ann” Xu, PhD & Randall Guensler, PhD (Trans/AQ, Inc.)

3.1 Executive Summary

This report presents findings from an analysis of individuals’ modality styles based on the three-day wearable GPS sample from the San Francisco Bay Area collected as part of the California Household Travel Survey (CHTS). This represents one of the first studies of personal modality styles in the United States that employs multiday GPS data. Modality styles are categorized as unimodal auto, unimodal green (i.e., walk, bike, and transit), multimodal auto, and multimodal green. Analyses focus on the relationship between modality styles and demographic, socioeconomic, and transit accessibility characteristics. The findings are consistent with existing knowledge about personal modality styles based on studies from Europe and the United States. Household vehicle ownership, home location access to transit, and individuals’ possession of a driver’s license have strong and statistically significantly correlation with modality style. Individuals with vehicle availability and a driver’s license are more likely to be unimodal or multimodal auto. Individuals whose home locations are transit accessible are more likely to utilize alternative modes. Younger adults tend to be more multimodal, with the exception that people age 65 and older are more likely to exclusively rely on alternative modes. Household presence of children is positively associated with individuals’ likelihood to rely on the auto mode, especially if an individual is employed. Household income, number of workers, individual education level, employment status, and gender are also correlated (to varying degrees) with modality styles.

Multiday GPS data can enrich the understanding of modality styles, but potential data quality issues remain a challenge in the analytical process. This chapter has outlined procedures that could prove useful in similar studies using multiday GPS data from other regions. Recommendations for further study are provided and are based upon current and previous experience analyzing multiday GPS data for modality styles.

3.2 Introduction

In urban areas of the United States, excepting New York City, most trips are made as a driver or passenger in a private vehicle. Only a small percentage of trips are made using another travel mode, such as walking, biking, or using transit, often referred to as alternative modes. It is important to note that even this relatively small percentage results from analysis of data across households, rather than within households (i.e., 5% of trips in a region may be made by transit, but the majority of these total trips may be made by regular transit users).

The goal of this project is to provide evidence of the value of GPS datasets housed at the Transportation Secure Data Center (TSDC) for use in multiday travel behavior analysis. In this paper, the focus is on examining variability in mode choice in multiday activity data. One of the limitations of most household travel surveys is that they are usually limited to only one day; as a result, meaning that modes used only occasionally or with high variability are often missed. This multiday assessment includes day-to-day variability in mode choice by household members as a function of demographic characteristics and trip purpose and—to the extent that relevant supplemental data could be obtained—as a function of transit access and land use.

After basic data exploration, the research team focused on the GPS subsample sponsored by the Metropolitan Transportation Commission (MTC) of the San Francisco Bay Area. The Caltrans multiday data for the Bay Area provided the best most comprehensive subset of data with continuous multiday, multimodal activity for use in the analyses. The entire Caltrans GPS subsample includes 3,871 households that completed all survey components. Of the 3,871 households, 3,429 households (9,141 respondents) are in the MTC subsample. Therefore, the Bay Area data constitute nearly 90% of all multiday data in the database. Given the level of effort that would be required to process the remaining 10% of the data for other regions for underlying demographic characteristics, roadway network, and transit networks, only the Bay Area data are employed in these analyses. Three days of electronic monitoring data were collected from members (age 16-75) from each Bay Area household; however, only one day of traditional travel diary data were recovered for each household. Detailed trip elements associated with mode choice and trip purpose are only available for one-third of the data.

Household recruitment strata were designed to collect data from representative households across California. The primary stratification variable was geographic, with 30 strata representing counties and urban areas across the state. Random stratified sampling within cross-classified demographic groups does not appear to have been undertaken during household recruitment (e.g., classifications by household size, crossed with number of workers, and then vehicle ownership, and income group, and transit access, etc.). Recruitment checks were in place to assess whether the final recruited households were representative within each individual variable alone (e.g., by household size alone, by income alone, etc.). In the California study, weighting was used to adjust outcomes for nonresponse bias. In the analyses presented in this chapter, household weights are not utilized, as the aim of this research is not to extrapolate the results to generate regional estimates; rather, the goal is to identify significant associations between modality styles and household characteristics.

3.3 Data Analysis

QA/QC of Data for Analysis

For the purposes of the GPS-based multimodal analyses, the starting point for data screening is the full set of Bay Area GPS travel data. Each GPS file in the Caltrans study was processed using the NuStatsTrip Identification and Analysis System (TIAS) software to identify potential trip ends. According to the documentation (NuStats, 2013a), the criterion for identifying potential trips was set to 120 seconds of dwell time (at a single location). "GPS trip data were then visually reviewed by analysts to screen out traffic delays and other falsely identified stops with dwell times of 120 seconds or more, as well as to add stops that had dwell times of less than 120 seconds but had clear "stop" characteristics (NuStats, 2013a)." The documentation indicates that chained trips that were not identified by the GPS dwell time criterion, but were reported in the diaries, were broken into trip legs (e.g., pick-up and drop-off trips, at school, ATM stops, etc.). It is unclear if other trips were manually identified by the data processing team. Furthermore, travel diary data are only available for one of the 3 days for each household that participated in GPS data collection. For the other 2 days in each household sample, it is unclear whether analysts used the patterns noted in the diary day to similarly break chained GPS trips into trip legs, assuming that the legs would also not be caught by the GPS dwell time criterion. Assessing the quality of the GPS data stream is necessary given the procedures employed, and decisions were made as to which data should be screened based upon the QA/QC analyses. Data screening removed some households, travel days, household travel days, individuals, individual travel days, and trips. However, given the lack of documentation as to how the second-by-second data were processed, much of the remaining assigned mode data were assumed to be accurate.

Analysis of Travel Modes—One-Day Diary Analysis

The one-day travel survey reported 42,460 total trips for the San Francisco Bay Area households. Data were coded into 29 modes, of which 31,034 trips (73.1%) were conducted either by an auto driver or an auto passenger (modes 5 and 6), while 7,393 (17.4%) of all trips were walk trips (mode 1) and 1,107 (2.6%) were bike trips (mode 2). The remaining trips were a mix of local transit, regional transit, commuter rail, vehicle for hire, and other alternative modes (6.9%).

Table 3-1 shows the household distribution and person distribution by number of modes used, as reported in the one-day travel dairy. At the household level, 3,260 of the 3,429 households (95.1%) reported travel on their assigned travel diary day. Among the 3,260 households who reported travel, 932 (28.6%) used only one mode, 1,084 (33.2%) used two modes, and 1244 (38.1%) used more than two modes (including walk travel). At the person level, 7,945 out of the 9,141 survey respondents (86.9%) reported travel. Among the 7,945 persons who reported travel, 5,145 (64.8%) used only one mode, 1,984 (24.6%) used two modes, and 816 (10.3%) used more than two modes (including walk travel).

Table 3-1: Number of Persons and Households, by Number of Modes Used

Number of Modes Used	Number of Households	Number of Persons
1	932	5,145
2	1,084	1,984
3	700	625
4	355	150
5	145	35
6	34	5
7	8	1
8	2	0
Total Reported Travel	**3,260**	**7,945**
Total in Sample	**3,429**	**9,141**

The large difference in the distribution indicates that household-level data are likely confounded by the presence of multiple individuals that use only one mode (i.e., adults and children).

The 29 travel modes—including walking and biking, personal automobile, heavy-rail, and a variety of other transit options—associated with the full Caltrans travel diary dataset can be found in the California Department of Transportation 2010-2012 Travel Survey Final Report Appendix (NuStats 2013b). Some of these modes are applicable to distinct population subsets and are not likely to provide a representative sample of trip chaining in the Bay Area alone. For example, the disability community participates in travel diary data collection, and many participants use the wheelchair mode; however, these households are not surveyed in sufficient numbers to represent the activities of this group. Some of the other transportation modes are only available in Los Angeles. Finally, the coding of trip modes, based on observed travel characteristics (such as speed or distance), needed to be assessed. Before conducting detailed analysis of the multiday trip data, a variety of data screening methods were applied to identify potential issues in the dataset and to eliminate households and trips that might be considered nonrepresentative from further analyses.

Data Screening

Data screening was conducted in two stages. The first stage focused on QA/QC. Multiple quality check criteria related to trip distance, duration, and speed were applied to retain only reasonable GPS trip records. The following criteria were used in the QA/QC stage:

- Walk trips shorter than 150 feet (45.7 m) were removed. This criterion follows suggestions by Clifton and Muhs (2012). 788 observations (0.7%) were removed in this step.
- Walk trips less than 5 minutes were removed. This criterion is based on the specifications listed on page 318 of the Appendix of the CHTS documentation (NuStats, LLC, 2013b). Appendix A discusses walk trip issues in much more detail. 10,988 observations were removed in this step (10.2%).
- Walk trips with speeds exceeding 8 mph were removed. A typical jogger jogs at 6 mph and a world-class runner can run approximately 13 mph (Grava, 2002). Only five observations were removed in this step (0.0%).
- Loop trips with different origin and destination types were removed. 58 observations were removed (0.1%).
- For nonwalking trips, the distance should be greater than 300 feet and duration should be at least 1 minute. Nonwalk trips that not meeting these criteria were removed. 1,172 observations were removed (1.1%).
- Nonplane trips with speed greater than 70 mph were removed (i.e., average speed for entire trip). 220 observations (0.2%) were removed.
- Plane trips with distance less than 100 miles were removed—five trips (0.0%) were removed in this step.

The second stage of screening pertained to the scope of analysis. Long-distance travel, persons with disabilities, and light travelers were excluded from further analysis.

- Based upon previous experience in Atlanta's Commute Atlanta study, the authors wanted to remove all households that employ a vehicle for commercial use from the database (Xu, 2010) as discussed in Appendix B. However, the Caltrans household travel diary survey did not ask such a question during recruitment.
- Long-distance travel—households that recorded trips more than 100 miles one-way were excluded from analysis (See Appendix C). This step removed 162 households (5.0% of all households).
- Persons with disabilities—households with individuals who recorded trips with modes 3 (wheelchair, 30 trips) and 21 (paratransit, 9 trips) were excluded from analysis. This step removed 10 households (0.3% of all households).
- Light travelers—only individuals who made at least three trips during the 3-day period were included in the modality analysis. 253 individuals were removed in the step (2.8% of all individuals).

After data screening, 87,854 trips from 6,094 persons of 3,128 households remained in the Caltrans analytical dataset for the San Francisco Bay Area. However, nine persons did not have matching personal information records in the database and two persons were younger than 16. These 11 participants were also excluded from data analysis, leaving 6,083 persons. With respect to the age of each participant, 23 people reported "don't know" (age=998) and 172 people declined to answer but are marked as "between 16 and 75." Unfortunately, these 195 persons could not be recoded as adults for the analysis because they may have contained individuals who are younger than 18 (and child travelers could not be used in the analysis). In addition, 264 people reported ages of 16 and 17, and were also excluded from the analysis. With respect to analysis of adult travel, the final working dataset contained 5,624 adult participants from 3,040 households making 81,839 trips.

Development of Mode Typologies

The original Caltrans Household Travel Survey data employed 29 modes, which are listed in Appendix D.

Each of the Caltrans Modes is examined in Appendix D, with details reasons provided as to why the authors concluded that certain modes should be excluded, or combined for the analyses presented in this chapter. The final aggregated modes employed in the analyses are defined as follows:

- w = Walk trips (Caltrans Mode 1).
- b = Bike trips (Caltrans Modes 2 and 4).
- a = Auto trips (Caltrans Modes 5-8, and Mode 10).
- t = Transit trips (Caltrans Mode 9 and Modes 11 to 29).

Caltrans Modes 3 and 21 (Wheelchair/Mobility Scooter and Dial-A-Ride/Paratransit) were not included as they are considered special modes to accommodate persons with disabilities. Persons with disabilities have different travel needs and travel patterns and their activity should be analyzed separately. Therefore, the researchers had excluded all persons with trips marked as modes 3 or 21 from the analyses.

For the multimodal analyses in this report, every trip mode was recoded into one of the aforementioned modes (i.e., walk, bike, automobile, or transit). Some trips were undertaken through more than one mode. The distribution of the trips across the four main modes and mode combinations is summarized in Table 3-2.

Table 3-2: Number of Trips, by Mode

Mode(s) Used	Number of Trips
Auto Trips (a)	43,454
Auto and Walk Trips (aw)	25,325
Walk Trips (w)	7,768
Transit Trips (t)	3,276

Mode(s) Used	Number of Trips
Bicycle Trips (b)	1,619
Bicycle and Walk Trips (bw)	334
Transit Trips and Walk Trips (tw)	54
Auto and Bicycle Trips (ab)	6
Auto and Bicycle and Walk Trips (abw)	3

Based on Table 3-2, a main mode is further assigned to each trip. Intuitively, the main mode of a trip can be assigned according to the mode that covered the longest distance of the trip. However, the distance of each trip segment, by mode, is not available in the dataset. Therefore, the main mode is assigned based on the mode with the highest average speed, as outlined below:

- Auto is assigned as the main mode for auto & walk (aw), auto & bike (ab), and auto, bike, & walk trips (abw).
- Bike is assigned as the main mode for bike & walk (bw) trips.
- Transit is assigned as the main mode for transit & walk (tw) trips.

Table 3-3 summarizes the final distribution of trips, by main mode.

Table 3-3: Distribution of Trips, by Main Mode

Main Mode	Number of Trips	Percentage
Auto (a)	68,788	84.1%
Bicycle (b)	1,953	2.4%
Transit (t)	3,330	4.1%
Walk (w)	7,768	9.5%
Total	**81,839**	**100.0%**

In reviewing the Caltrans survey documentation (NuStats, 2013a and 2013b), the research team identified a potential problem with mode coding that warrants further investigation. Appendix E shows that the 22 modes listed in the main Caltrans report (NuStats, 2013a) do not match the 29 modes reported in the report Appendix (NuStats, 2013b). Presumably, the data collected for the Bay Area were coded using a single standard, where mode definition remained consistent within the samples. Modes 1–9, which include walk, bike, and all personal vehicles appear to be unaffected by any potential coding difference. However, local bus, school bus, connecting public bus, express bus, light rail, heavy-rail, and ferries might be affected. An initial check of trip duration and trip distance for the modes did not identify any immediate coding issues (i.e., searching for significantly different average speed slope patterns within one mode dataset). However, it would be difficult to differentiate between some of the potential mode mismatches without conducting a detailed analysis of specific transit routes and GIS spatial analysis, which is beyond the available *resource budget of this project. Therefore, the research team operated under the assumption that the modes in the database were properly coded.*

The Multimodal Activity vis-à-vis Modality Styles

The research team sought to assess potential relationships between demographic and socioeconomic characteristics and traveler multimodality. For the analyses in this paper, four modality styles were defined, following the definitions previously employed in Vij, et al. (2011):

- Unimodal auto (uni_a)—a person who is predominantly an automobile user, with an auto mode share of 90% or above.
- Unimodal green (uni_g)—a person who travels predominantly by alternative modes, with a walk mode share of 80% or above, or a bike mode share of 80% or above, or a transit mode share of 80% or above.
- Multimodal auto (mul_a)—a person who is not unimodal, with an auto mode share of 10% or above (but less than 90%).
- Multimodal green (mul_g)—a person who is not unimodal, with an auto mode share of less than 10%.

The subsequent analyses assess traveler membership across these modality styles.

Cautionary Notes Based Upon Data Coding Review

In reviewing the data descriptions, the following additional notes apply to the analyses:

- Travel days include weekends.
- The Household vehicles variables (HHVEH, household vehicles, and VEHOP, operable household vehicles) include all vehicles owned, leased, and available, including all motorcycles and scooters.
- The coding of origin-destination (O-D) types in the GPS dataset may not be reliable enough for analysis across actual trip purpose. Place types are limited to six options (i.e., home, primary job, school, second job, transit stop, and other place [please specify]). Activity codes for trips at a transit stop place include changing modes (APURP=21), and dropping off and picking up transit passengers (APURP=22), but also includes drive-through meals (APURP=23) and other drive-through activities such as an ATM stop (APURP=24). The team assessed mode/place/activity combinations, but some uncertainty was anticipated due to potential subjective coding of GPS trips in the absence of travel diary data for two-thirds of the days.
- Households were allowed to participate in the standard travel diary data collection even if they refused to participate in GPS elements of the study. An analysis is probably warranted across the three response codes for the Carbon Vehicle (CVHGPS) variable and to compare baseline one-day travel diary data for GPS and non-GPS participants across the same demographic strata:

- As presented in Table 3-4, only 5.8% (2459) of households do not have a household vehicle. Of these households, 10.6% (310 households) indicate that they do not possess a vehicle because they do not need a car and 13.1% (384 households) can make all their trips by alternative mode (NuStats, 2014b, Table F.1.21). The remaining households that do not own vehicles appear to be captive to alternative modes.

Table 3-4: Reasons Given for Not Owning Vehicles in California Household Study

Reason for Zero Vehicle Ownership	Frequency	Percent
Do not need a car	310	10.6%
Too expensive to buy	575	19.6%
Too expensive to maintain	385	13.1%
Health/age-related reasons	240	8.2%
Cannot get insurance	22	0.8%
Concerned about environment	105	3.6%
Get rides from other people	106	3.6%
No place to park	68	2.3%
Public transit/car share/bike/walk	384	13.1%
No driver's license	237	8.1%
Cannot drive	291	9.9%
Other	186	6.3%
DK	16	0.5%
RF	6	0.2%

Source: NuStats, 2014b, Table F.1.21

Modality Style Analysis—Multiday GPS Sample

Each subsection of this section presents the analytical results for modality style across a variety of independent socioeconomic variables taken one at a time, and some in combination for certain variables of interest. Analyses pertain only to the 5,631 adults (age>=18) in the final analytical dataset. The authors report analyses of multimodality across household characteristics, including number of workers, income, vehicle ownership, presence of children, household structure, and home location access to transit. Analyses also examine individual characteristics, including gender, education, possession of a driver's license, employment status, and age group. Table 3-5 summarizes the overall shares of modality styles in the MTC 3-day wearable GPS sample.

Table 3-5: Overall Sample Distribution of Modality Styles

Modality Style	# Persons	% of Total	# Trips in 3 Days	# Auto Trips	# Bike Trips	# Transit Trips	# Walk Trips	# Daily Trips/Person	Average Daily Distance/Person (miles)	Average Daily Duration/Person (hrs)
Unimodal Auto	3,536	63%	51,947	51,125	81	45	696	4.90	29.52	1.12
Multimodal Auto	1,806	32%	26,661	17,570	1,238	2,270	5,583	4.92	25.76	1.33
Multimodal Green	166	3%	2,202	52	355	858	937	4.42	14.32	1.13
Unimodal Green	116	2%	1,029	41	279	157	552	2.96	4.90	0.75
Total	**5,624**	**100%**	**81,839**	**68,788**	**1,953**	**3,330**	**7,768**	**4.85**	**27.36**	**1.18**

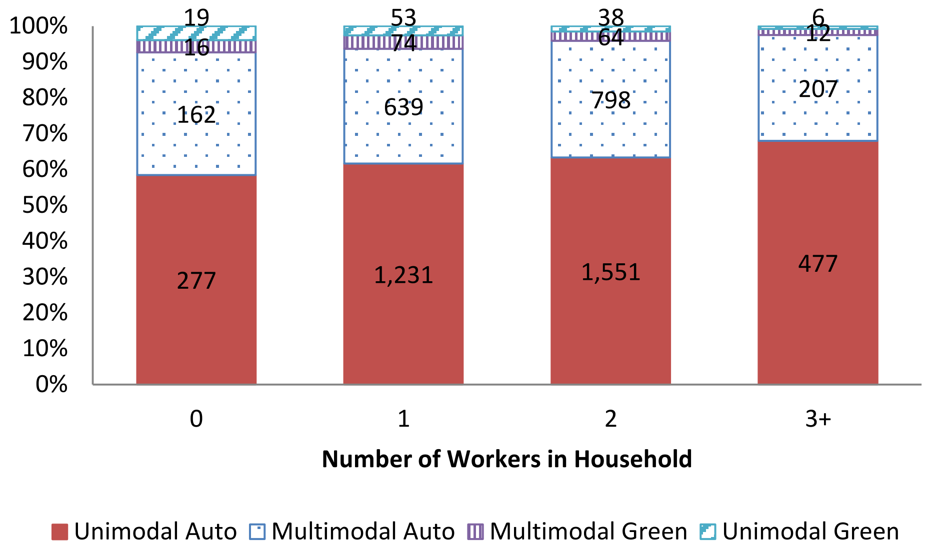
Modality Styles, by Household Number of Workers

Table 3-6 and Figure 3-1 present modality styles, by number of household workers (0, 1, 2, and 3+). As number of workers increases, the unimodal auto share increases.

Table 3-6: Modality Styles, by Number of Household Workers (0, 1, 2, 3+)

Number of Workers in the Household	Code	Unimodal Auto	Multimodal Auto	Multimodal Green	Unimodal Green	Total
0	0	277	162	16	19	474
1	1	1,231	639	74	53	1,997
2	2	1,551	798	64	38	2,451
3+	3	477	207	12	6	702
Total		**1,806**	**3,536**	**166**	**116**	**5,624**

Figure 3-1: Number of Persons in Each Modality Style, by Number of Household Workers



Modality Styles, by Household Income

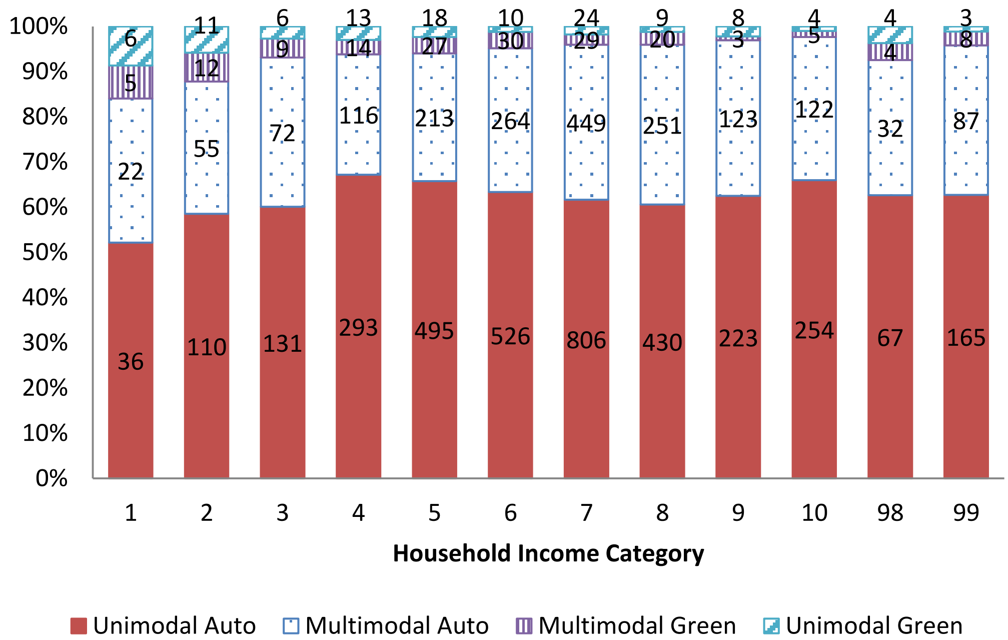
Table 3-7 provides the breakdown of persons in each household income group.

Table 3-7: Number of People in Each Household Income Category

Income Group	Income Range	Number of People	Percent of Sample
1	$0 to $9,999	69	1%
2	$10,000 to $24,999	188	3%
3	$25,000 to $34,999	218	4%
4	$35,000 to $49,999	436	8%
5	$50,000 to $74,999	753	13%
6	$75,000 to $99,999	830	15%
7	$100,000 to $149,999	1308	23%
8	$150,000 to $199,999	710	13%
9	$200,000 to $249,999	357	6%
10	$250,000 or more	385	7%
98	Don't Know	107	2%
99	Refused	263	5%
Total		**5624**	**100%**

Figure 3-2 shows an interesting correlation between household income and multimodality. Within the low to medium household income groups (up to $50,000), the share of unimodal auto individuals increases as household income increases. Between $50,000 and $200,000, the share of unimodal auto individuals decreases as household income increases. In the highest household income groups (above $200,000), the share of unimodal auto individuals increases and the share of unimodal green and multimodal green individuals decreases as income increases.

Figure 3-2: Personal Modality Styles, by Household Income

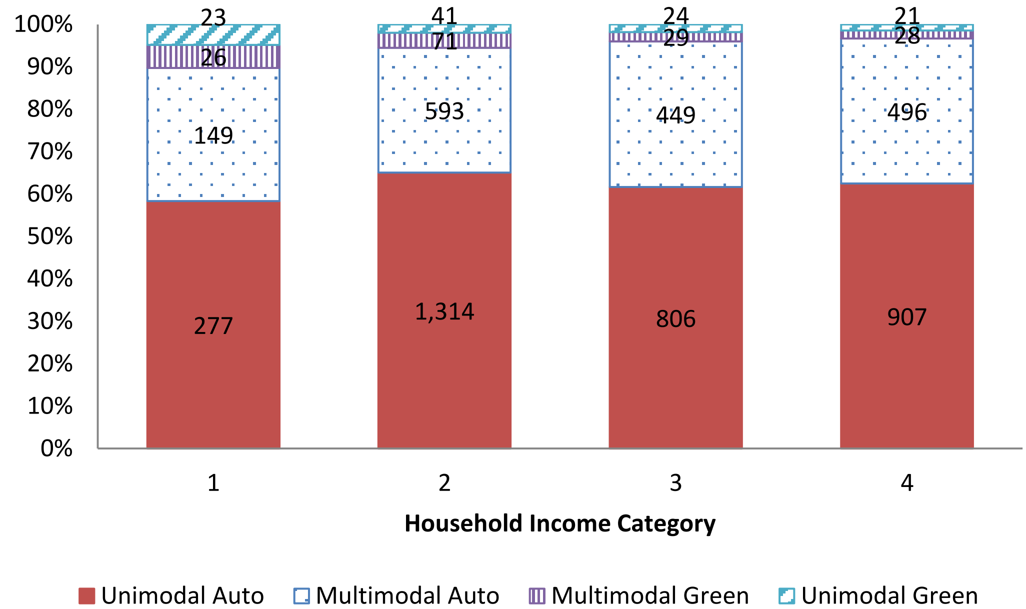


Modality Styles, by Collapsed Household Income Category

The research team clustered income groups together to produce Figure 3-3. As this figure shows, there are no striking modality share differences other than perhaps a slight uptick in unimodal auto selection for middle-income participants. For this analysis, income categories were defined as follows:

- 1 = Low Income—$0 to $35k (475 adults).
- 2 = Medium Income—$35k to $100k (2,019 adults).
- 3 = High Income—$100k to $150k (1308 adults).
- 4 = Higher Income—$150+ (1,452 adults).

Figure 3-3: Personal Modality Styles, by Household Income Category



Modality Styles, by Household Operating Vehicles

Figure 3-4 presents the distribution of individuals across modality styles by the number of operating vehicles in the household (including automobiles, trucks, motorcycles, and motor scooters). Issues with vehicle ownership in the database are discussed in more detail in Appendix F. Operating vehicle categories are established as 0, 1, 2, 3, and 4 or more vehicles per household (see bullet summary below):

- 0 vehicles—190 adults.
- 1 vehicle—1,232 adults.
- 2 vehicles—2,636 adults.
- 3 vehicles—1,145 adults.
- 4+ vehicles—421 adults.

As expected, ownership of zero vehicles is correlated with alternative mode use. The share of unimodal auto style increases with vehicle ownership, with a decreasing rate as ownership increases. Approximately 5% of individuals from households with zero operating vehicles still fall into the unimodal auto category and 40% of individuals from households with zero operating vehicles fall into the multimodal auto category. Only 55% of individuals from households that do not have operating vehicles fall into the unimodal green or multimodal green categories. This figure does not control for household size, income, or other variables at this point in the analyses. Figure 3-5 breaks down the relationship by income category. Figure 3-6 breaks down the relationship by number of workers in household. Across income categories and the number of workers in each household, the trend holds true that the share of unimodal auto style increases with vehicle ownership, but the effect sizes differ with each one-vehicle increase in the household.

Figure 3-4: Personal Modality Styles, by Household Vehicle Ownership

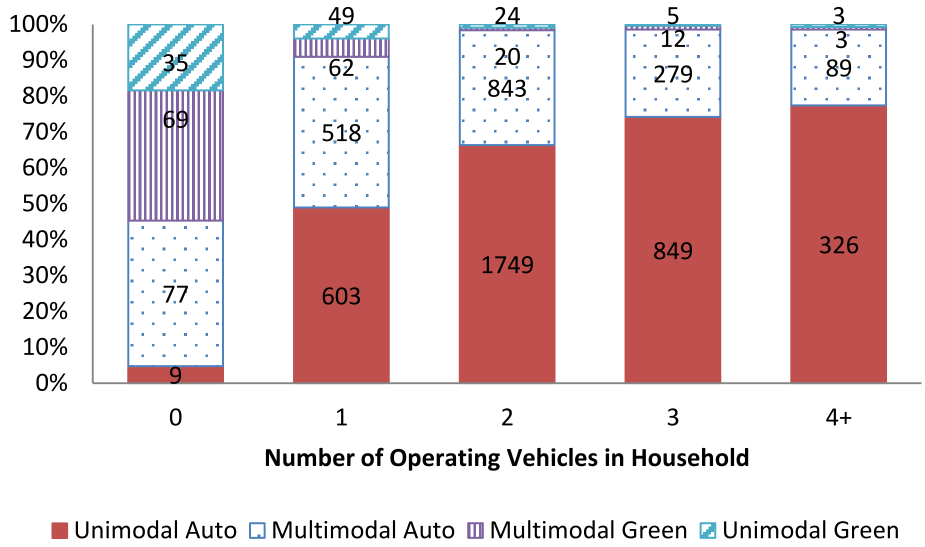


Figure 3-5: Personal Modality Styles, by Vehicle Ownership (inner x-axis) by Income Category (outer x-axis)

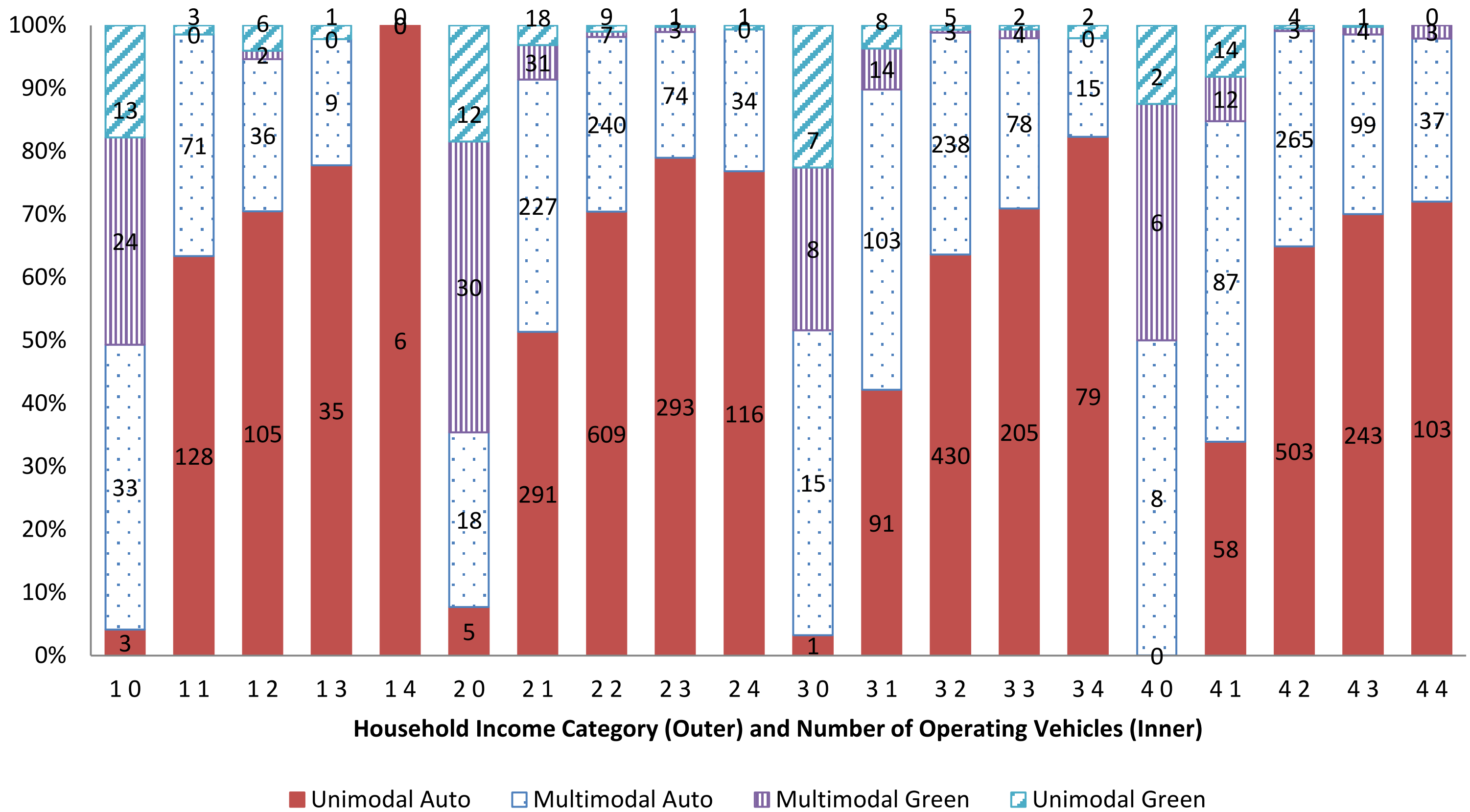
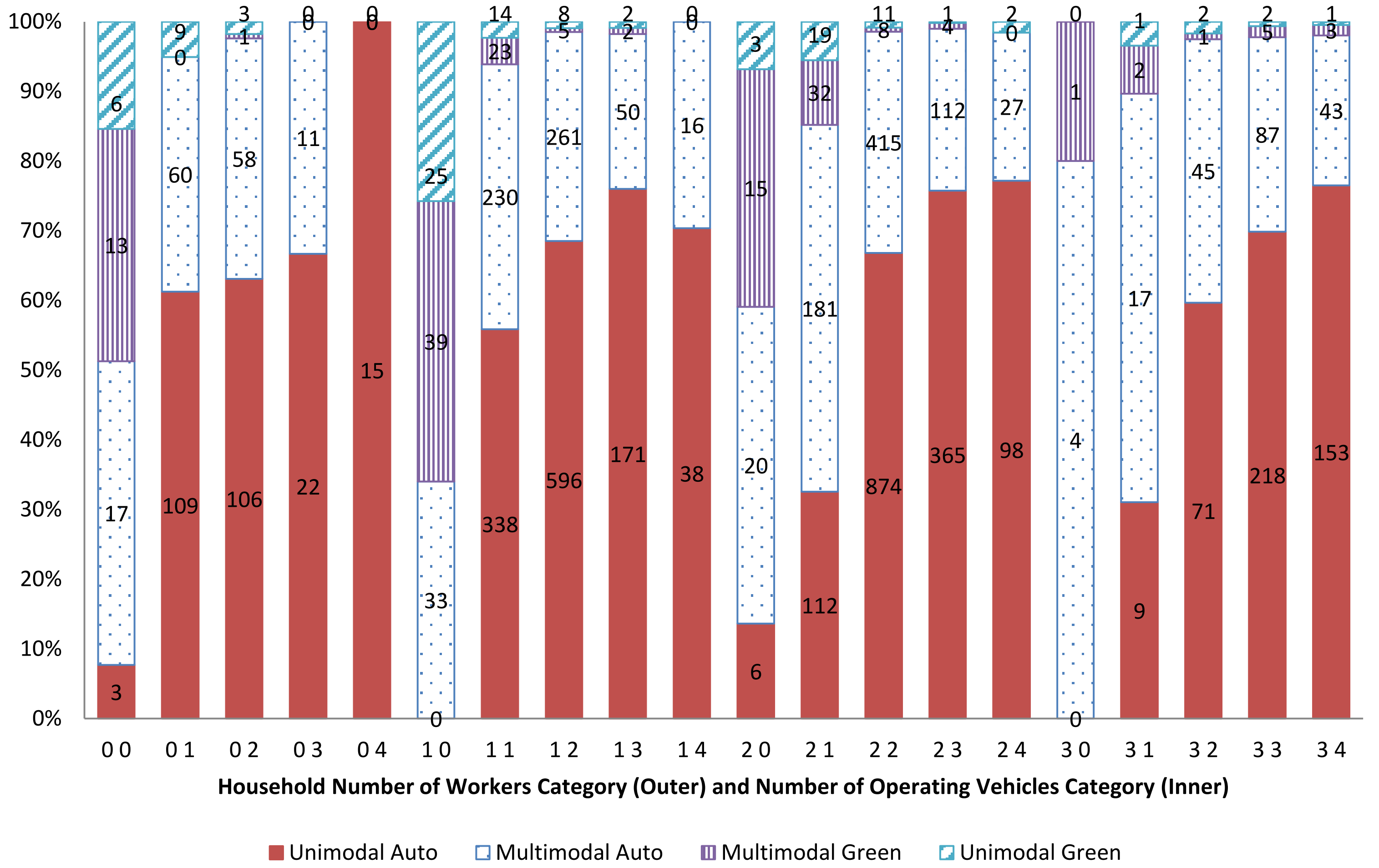


Figure 3-6: Personal Modality Styles, by Vehicle Ownership (inner x-axis) by Household Workers Category (outer x-axis)



Income vs. Number of Workers

One of the potential confounding variables in the previous analyses is the number of workers in the household (HHEMP_CAT) versus the number of persons in the household. That is, high household incomes may not be as significantly related to alternative mode activity as the average income per worker in the household (only information about income at a household level—not at the person level—is available). The number of workers per household (see coding below) is plotted against household income category in Figure 3-7. As suspected, higher-income households tend to have significantly more workers per household, with most of the increase appearing in the shift from one-worker households to two-worker households as household income category increases. Modality style is then plotted for number of household workers and household income category in Figure 3-8. Within the high-income group especially, unimodal auto tends to increase with number of workers:

- 0—0 Workers.
- 1—1 Worker.
- 2—2 Workers.
- 3—3+ Workers.

Figure 3-7: Income vs. Number of Household Workers

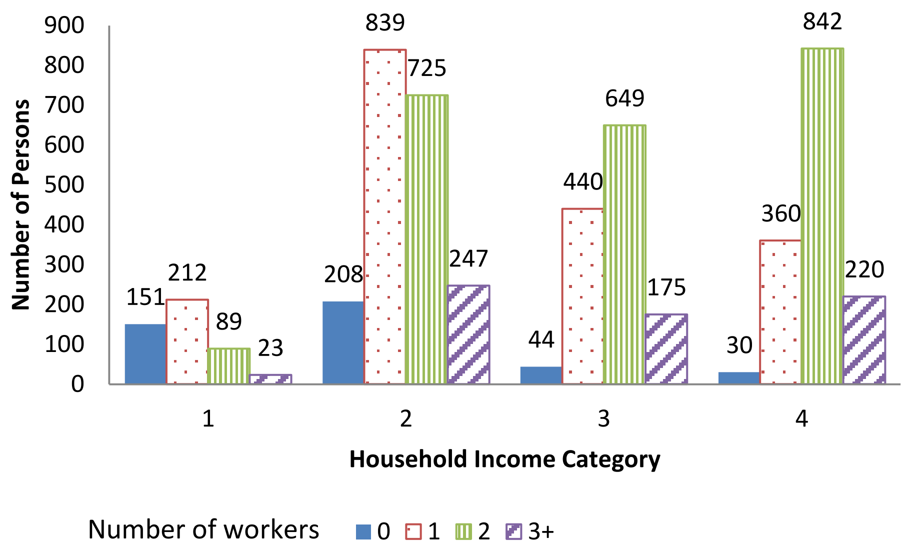
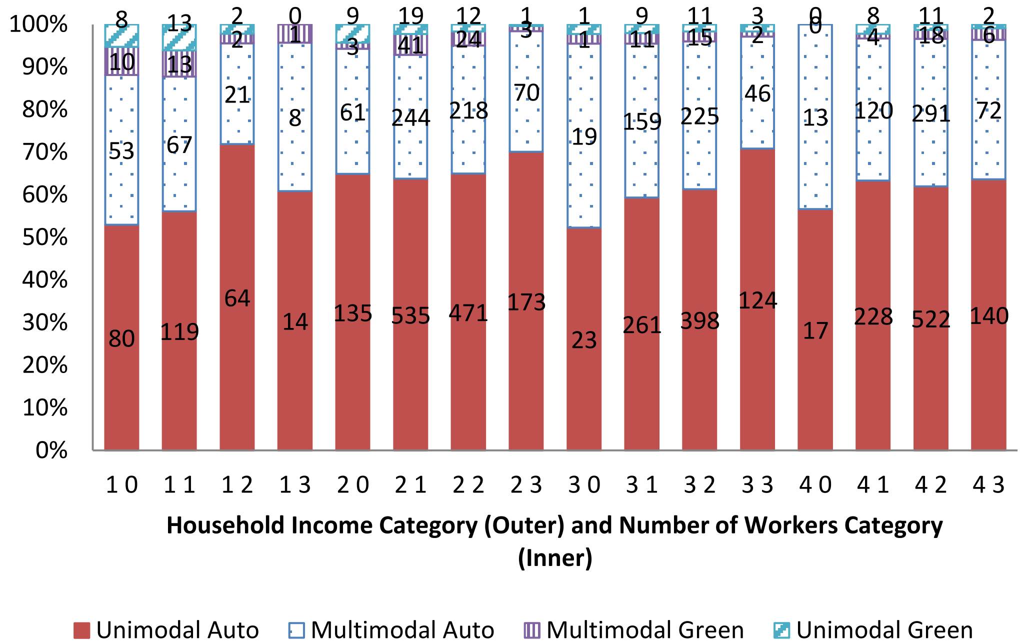


Figure 3-8: Income vs. Number of Household Workers



Modality Styles, by Household Presence of Children

Presence of children in the household is hypothesized to impose a constraint on multimodality in that it is typically easier to travel with children by automobile than by alternative modes. On the other hand, recreational travel by walking and biking may also increase. The sample contained 3,530 adults without children (coded as 0) and 2,094 adults with children (coded as 1). Figure 3-9 illustrates the four modality types for adults without children and with children. As expected, the presence of children is correlated with higher unimodal auto group membership and decreased multimodal auto and multimodal green group membership. Figure 3-10 further breaks down this relationship by subcategorizing by household income. The same general relationship is true across the income groups, with perhaps a more pronounced shift noted in the lower- and middle-income groups. However, it is important to note that correlations between income, household structure, and residential location may all be interacting with respect to this observed change.

Figure 3-9: Personal Modality Styles, by Presence of Children

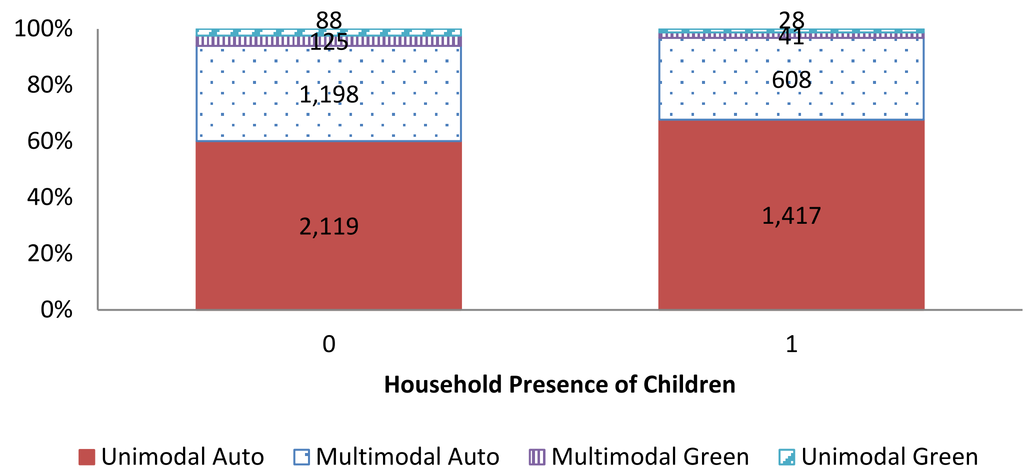
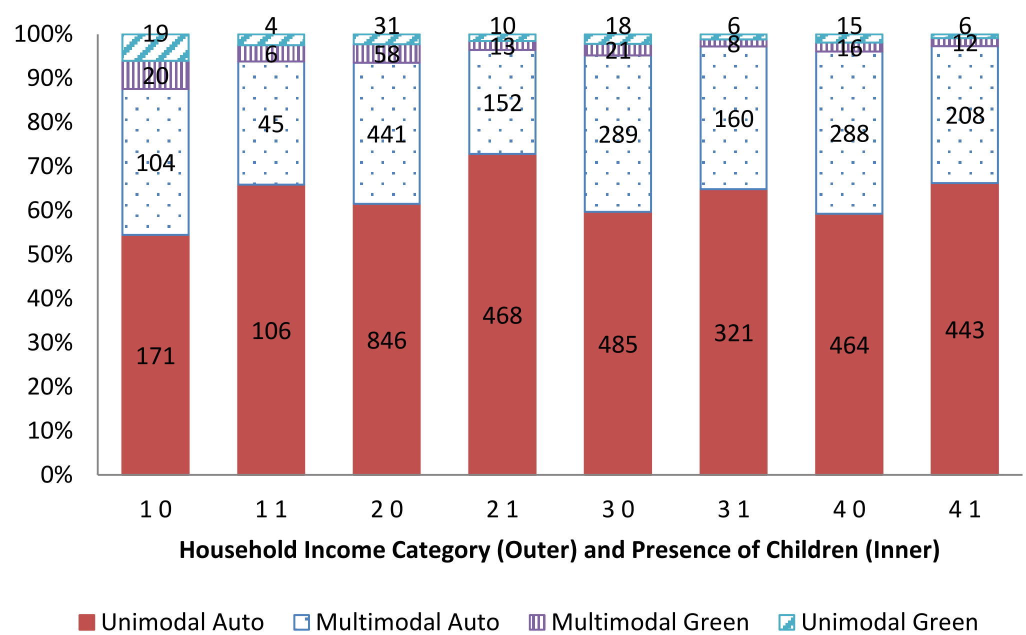


Figure 3-10: Personal Modality Styles, by Presence of Children (inner x-axis) and Household Income Category (outer x-axis)



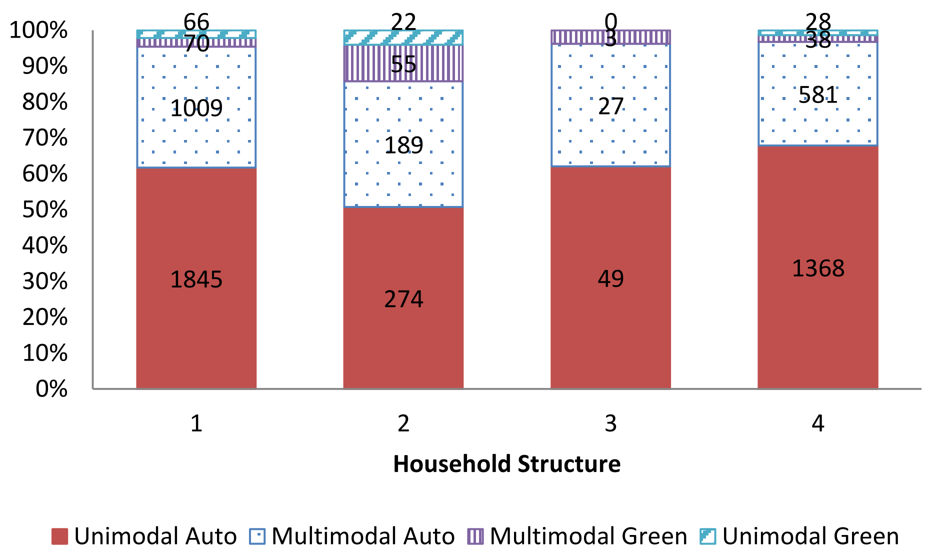
Modality Styles, by Household Structure

Table 3-8 and Figure 3-11 illustrate personal modality styles, by household structure. The four categories were selected to assess the potential interaction of both presence of children and presence of children in single-parent households on modality. Single individuals are hypothesized to have significant flexibility in travel with a propensity toward multimodality, while single adults with children are hypothesized to have less flexibility and may derive higher utility from unimodal auto activity. Single adults with no children exhibited significantly higher multimodality and unimodal green membership. Single adults with children did not differ greatly from other modality categories; however, the sample was small.

Table 3-8: Modality Styles, by Household Structure

Life cycle description	Code	Unimodal Auto	Multimodal Auto	Multimodal Green	Unimodal Green	Total	Percent
2+ adults, zero children	1	1,845	1,009	70	66	2,990	53%
1 adult, zero children	2	274	189	55	22	540	10%
1 adult, with children	3	49	27	3		79	1%
2+ adults, with children	4	1,368	581	38	28	2,015	36%
Total		**3,536**	**1,806**	**166**	**116**	**5,624**	**100%**

Figure 3-11: Personal Modality Styles, by Household Structure



Modality Styles, by Home Transit Access

For each household, the geocoded locations of home and work were identified in the database. The General Transit Feed Specification (GTFS) data for all of the major transit systems were then accessed and imported into the analytical database. Appendix G provides Microsoft MapPoint maps of the transit stop locations for the transit systems in the San Francisco Bay Area. Transit stop locations (latitude, longitude) for all major transit systems were available. Data for two smaller systems, Napa Vine and Solano County (SolTrans), were not available. However, given the overall sample size, the authors do not believe that the transit accessibility provided by these two smaller systems would affect the outcomes if they were included.

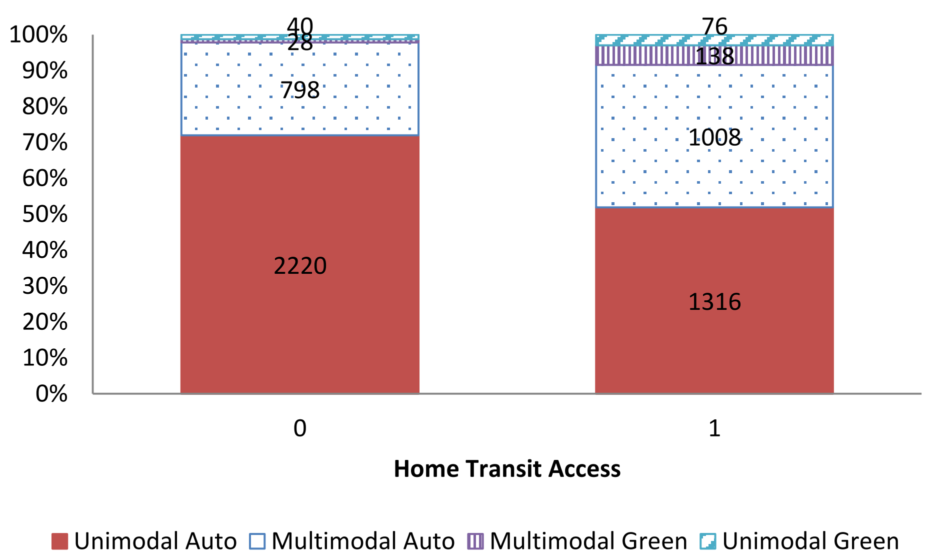
When the home location was located within one-quarter mile of a transit stop, a new variable, "Transit Access" for commutes, was defined as "Yes" and coded as "1."

For households where the home location was not located within one-quarter mile of a transit stop, the new variable, "Transit Access" for commutes, was defined as "No" and coded as "0." Home transit access for trips was significantly associated with unimodal green, multimodal green, and multimodal auto group membership (see Table 3-9). In essence, the unimodal auto share was significantly lower for households where home locations were "transit accessible." It is important to note that transit access is fully correlated with the decision by transit agencies to provide service to the home locations (and work locations, too), where such decisions are also presumably based upon transit service feasibility. Service feasibility is related to transit passenger demand, which is also a function of land-use density, intensity, mix of uses, etc. The presence of a transit stop at a location does not cause travel to be undertaken. Nevertheless, given the existing transit systems, where existing transit service and selection of stop locations are to some extent market-driven to enhance farebox recovery, the presence of transit stops and home locations appears to be an important factor in multimodality in the Bay Area. Future work will further incorporate work location transit access into the analysis.

Table 3-9: Sample Distribution of Modality Styles, by Home Transit Access

Transit Access	Unimodal Auto	Multimodal Auto	Multimodal Green	Unimodal Green	Total
no (0)	2,220	798	28	40	3,086
	72%	**26%**	**1%**	**1%**	**100%**
yes (1)	1,316	1,008	138	76	2,538
	52%	**40%**	**5%**	**3%**	**100%**

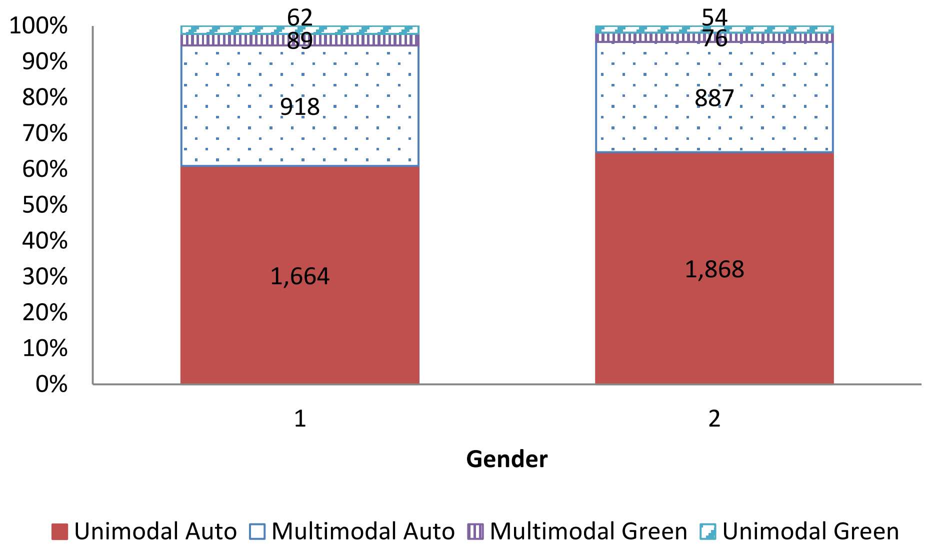
Figure 3-12: Personal Modality Styles, by One-Quarter Mile Transit Stop Access



Modality Styles, by Gender

The sample of adults contains 2,733 males and 2,885 females (one person declined to reveal her or his gender). Females have a slightly greater propensity to be unimodal automobile and slightly less multimodal green (Figure 3-13).

Figure 3-13: Personal Modality Styles, by Gender (male = 1, female = 2)



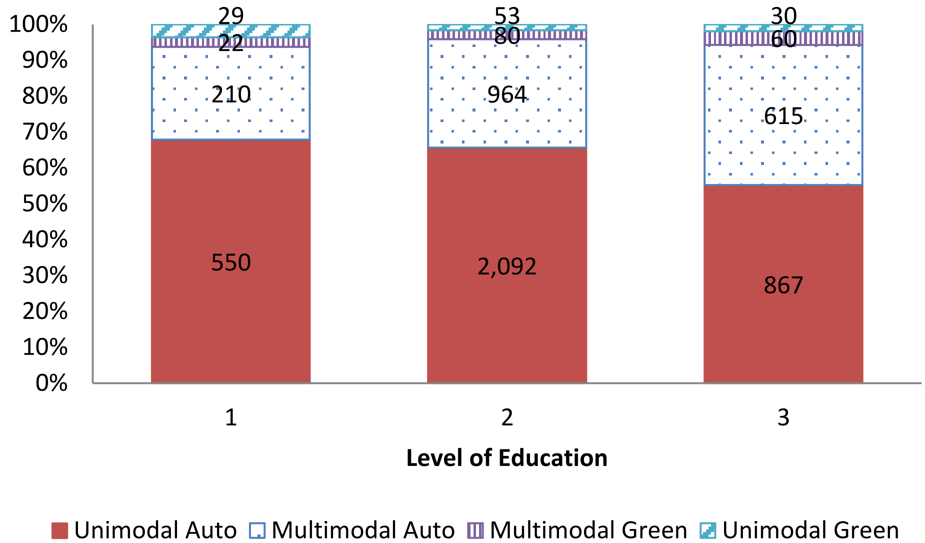
Modality Styles, by Education Level

Figure 3-14 presents the modality styles across driver education levels. The level of education completed by the driver was coded as follows in the Caltrans dataset: 1 = Not a high school graduate, grade 12 or less (which also includes young children); 2 = High school graduate (high school diploma or GED); 3 = Some college credit but no degree; 4 = Associate or technical school degree; 5 = Bachelor's or other undergraduate degree; 6 = Graduate degree (includes professional degree like MD, DDS, JD, etc.); 7=Other (specify); 8=DK; and 9=RF. The education variable was recoded as follows:

- 1 = High school degree or less (811 adults).
- 2 = College degree or some college beyond high school degree (3,189 adults).
- 3 = Advanced degree or education beyond college (1,572 adults).
- 0 = Other, did not know or refused to answer (52 adults).

Higher education levels are clearly correlated with multimodal activity, both in multimodal that is predominantly by automobile and by alternative modes. The increase in multimodality comes from a significant decrease in unimodal automobile activity. However, education levels are correlated with many other sociodemographic variables, such as employment status, income, and home location choice. One cannot conclude from Figure 3-14 that education level infers a causal relationship with respect to multimodality. However, the correlation is significant.

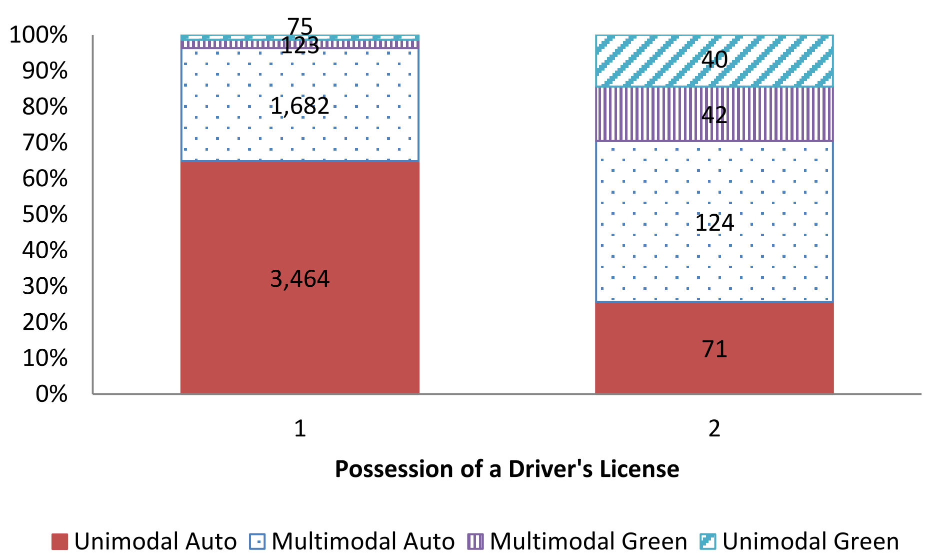
Figure 3-14: Personal Modality Styles, by Individual Education Level



Modality Styles, by Possession of a Driver's License

Figure 3-15 presents modality styles across the individual's possession of a driver's license. The coding for a valid driver's license was set as 1=yes and 2=no. Among the adults, 5,344 people reported having a valid license, whereas 277 reported not having a license. Two people reported "don't know" and one person refused to answer (neither group is included in Figure 3-15). Individuals with a driver's license are much less likely to utilize alternative modes.

Figure 3-15: Personal Modality Styles, by Possession of a Driver's License (1 = yes, 2 = no)



The Caltrans household survey did not ask follow-up questions designed to identify the reason why the respondent does not possess a driver's license. This is in contrast to questions asked of travelers who do not own a household vehicle (HHNOV), where the survey specifically requests additional information about why the household does not own a vehicle. HHNOV allows respondents to select reasons why they do not own an automobile from all that apply (see Table 3-10).

Table 3-10: Reasons that can be Selected by Participants for not Owning an Automobile

Code	Reasons for Not Owning an Automobile
01	Do not need a car—I can do what I need without a motor vehicle
02	Too expensive to buy
03	Too expensive to maintain (gas/insurance/repairs)
04	Health-/age-related reasons
05	Cannot get insurance
06	Concerned about impact on environment
07	Get rides from other people
08	No place to park
09	Use public transit/car share/bike/walk
10	No driver's license
11	Cannot drive
12	Other
98	I do not know
99	I prefer not to answer

For the 278 individuals without a license, additional data are necessary to assess why they do not have driver's license. As it stands, it is not possible to assess whether there is a causal link between a lack of vehicle ownership, lack of a drivers' license, and the need for either a vehicle or license. In other words, it is not possible to ascertain whether the traveler does not have a license and own a vehicle by choice (e.g., they live and work downtown and choose not to have a license and not to own a vehicle) or because they have somehow been forced into this situation. Future surveys should include a question designed to ascertain why an individual does not have a driver's license given the correlation with alternative mode use.

Modality Styles, by Employment Status

One's employment status could potentially affect his or her modality style. Among the 5,631 adults in the dataset, 4,093 reported that they were employed and 1,523 were not employed. Five people reported "don't know" and three people refused to answer. Figure 3-16 shows the modality style shares among the adults who reported employment status. Not much practical difference is noted in automobile categories. A shift is noted between multimodal green and unimodal green, with an increased unimodal share for nonworkers. Figure 3-17 then breaks down the relationship by income category. The results are mixed. Within the low-income group, the fact that an individual is employed is positively associated with his or her likelihood to be unimodal auto. In the medium-income group, the correlation is reversed. In the high- and higher-income groups, employment status does not seem to have an impact on one's modality style.

Figure 3-16: Personal Modality Styles, by Individual Employment Status (1 = yes, 2 = no)

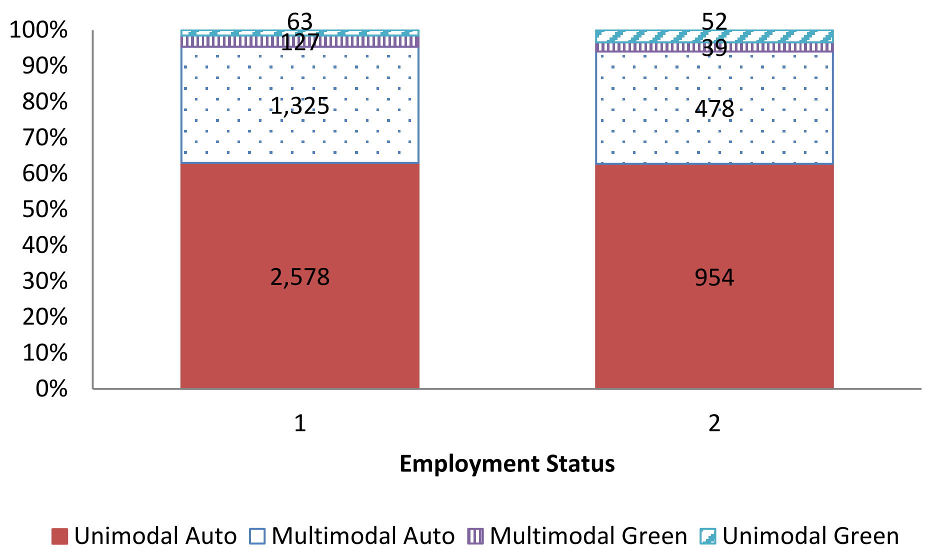


Figure 3-17: Personal Modality Styles, by Individual Employment Status (inner x-axis) and Household Income Category (outer x-axis)

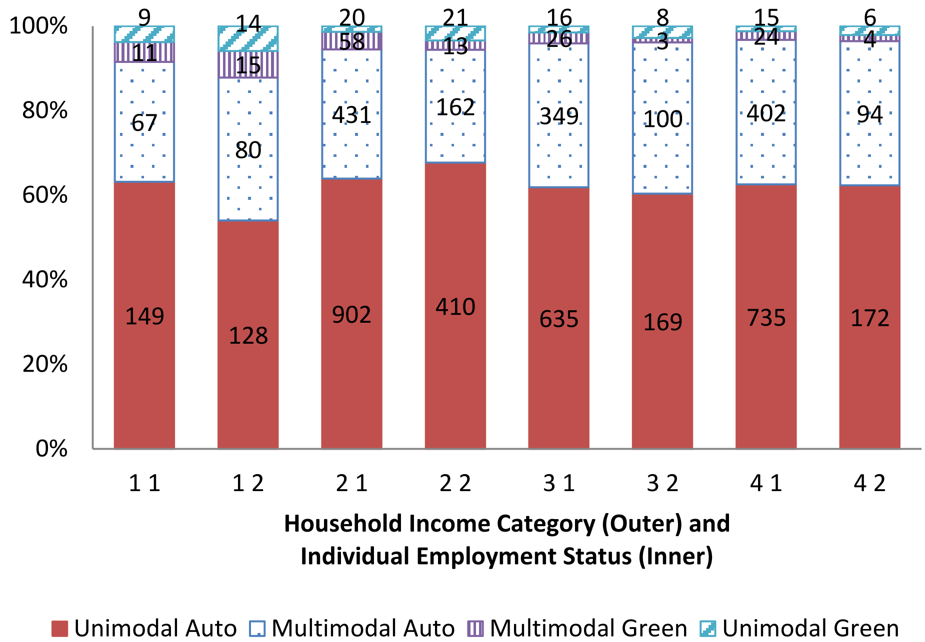
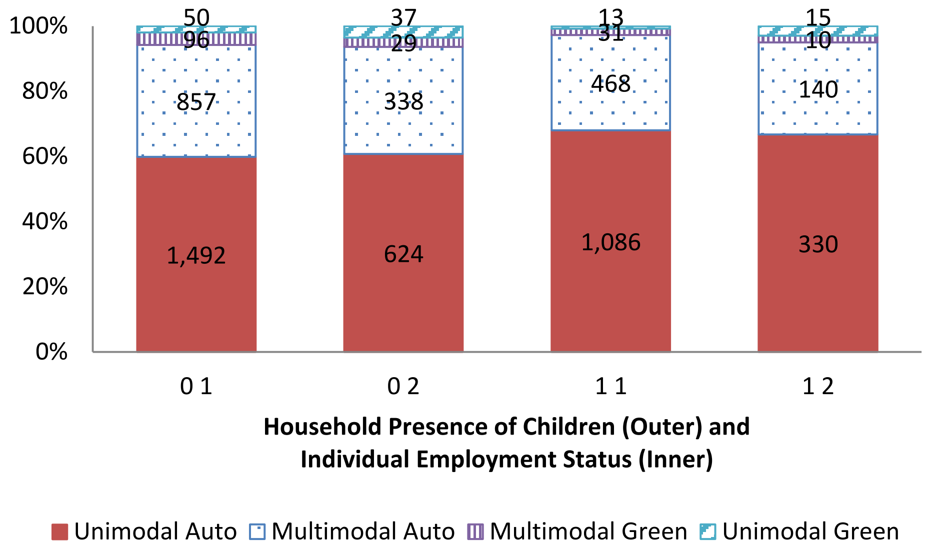


Figure 3-18 shows the interaction of effects between the presence of children in a household and an individual's employment status. When an individual comes from a household without children, there is no pronounced difference in modality style shares, whether or not this person is employed. On the contrary, when there is at least one child in the household, an individual is more likely to be auto-oriented (i.e., unimodal and multimodal auto) if he or she is employed than his or her unemployed counterparts. This is an intuitive observation, considering the scheduling constraints of a working parent.

Figure 3-18: Personal Modality Styles, by Employment Status (inner x-axis) and Presence of Children (outer x-axis)



Modality Styles, by Age Group

This section explores shares of modality styles by age group. Table 3-11 lists the definitions of the age groups and the number of individuals in each age group. Figure 3-19 provides a graphical representation of the modality style shares across age groups. The results indicate that younger adults have smaller combined shares of unimodal and multimodal auto; hence, younger people may be less auto-oriented and more multimodal.

Table 3-11: Definitions and Sample Distribution of Age Groups

Age Group	Code	Unimodal Auto	Multimodal Auto	Multimodal Green	Unimodal Green	Total
18–24	1	248	128	16	11	403
25–34	2	319	167	25	19	530
35–49	3	1,018	505	52	25	1,600
50–64	4	1,560	799	63	45	2,467
65+	5	391	207	10	16	624
Total		**3,536**	**1,806**	**166**	**116**	**5,624**

Figure 3-19: Personal Modality Styles, by Age Group

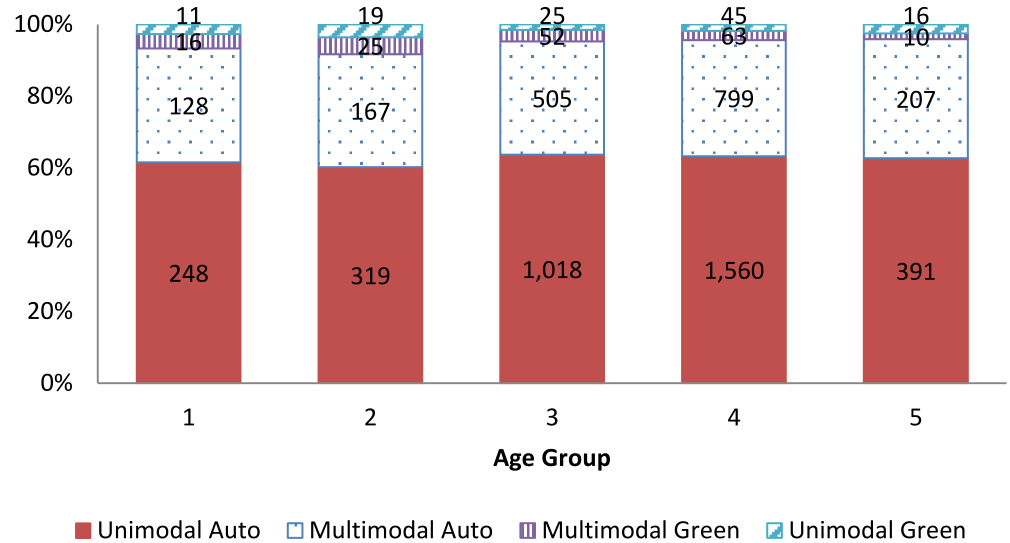
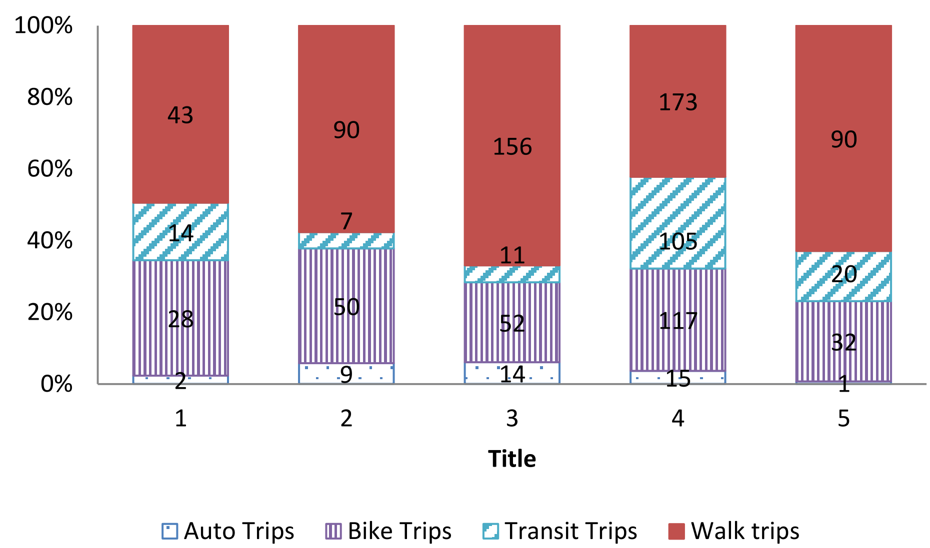


Figure 3-20 shows the number of trips, by mode, across age groups within the unimodal green group. No clear trends readily emerge, most likely due to the small sample size of the unimodal green group and the small sample size of the senior population group (age 65 and above).

Figure 3-20: Trip Distribution by Mode, by Age Group within Unimodal Green Style



Modeling of Group Membership

In this section, the research team modeled modality styles as group membership, where participation in any of the four groups group has some derived utility for the participant. A multinomial logit choice model is applied to assess the probability of group membership for each individual as a function of the demographic, sociodemographic, and transit access variables described in earlier analyses. The multinomial logit model employed four modality styles:

- 1—Unimodal auto (reference case).
- 2—Unimodal green.
- 3—Multimodal auto.
- 4—Multimodal green.

The predictor variables related to demographic characteristics include:

- Gender (1 – male, 2 – female);
- Age group (Young – 18 to 34, Middle Young – 35 to 49, Middle – 50 to 64, and Senior – 65 and older);
- Presence of children (0 – no, 1 – yes);
- Education (1 - high school degree or less, 2 - undergraduate, 3- graduate);
- Possession of a driver's license (1 – yes, 2 – no);
- Socioeconomic status measured through employment status (1 – yes, 2 – no);
- Household income (1 – Low, 0-35k; 2 – Medium, 35-100k; 3 – High, 100-150k; 4 – Very High, 150k+);
- Vehicle ownership (0, 1, 2, 3, 4+); and
- Land use captured by home transit access within one-quarter mile (0 – no, 1 – yes).

The presence of children was chosen as the predictor variable instead of the lifecycle category defined earlier, considering the small sample size of the single parents.

Home transit access is a significant variable and is positively associated with an individual's likelihood to be multimodal. If one's home location is within ¼ miles of a transit stop, this person is much more likely to be multimodal, compared to individuals whose home locations are not near a transit stop.

Table 3-12 presents the resulting coefficients for the multinomial logit model. All signs of the coefficients make intuitive sense.

Gender turns out to be a significant variable. Females are less likely to be multimodal than are males, which is in line with observations in Buehler & Hamre (2014). Age group is generally not significant, but the trends signs of the coefficients that older individuals are less likely to be multimodal, as expected from current consensus in existing literature (Kuhnimhof et al. 2006; Nobis, 2007; Kuhnimhof et al. 2012; Chlond, 2012; Buehler & Hamre, 2014). Individuals from households with children are less likely to be multimodal, and the effects are significant at 0.05 level when comparing unimodal green and multimodal auto to unimodal auto. In terms of education, the higher the education level, the more likely a person is to be multimodal auto or multimodal green, compared to unimodal auto, and the effects are significant. The possession of a valid license is significant in all comparisons.

Socioeconomic characteristics also play an important role in modality styles. Employment status is a significant variable when comparing unimodal green to unimodal auto. An individual who is not employed is about twice as likely to be unimodal green than unimodal auto. Higher income is associated with higher likelihood of being multimodal, but the effect size is small. Vehicle ownership is negatively associated with multimodality and the effects are significant across all comparison pairs. Individuals from households with no operating vehicles exhibit different modality styles than those from households with at least one operating vehicle.

Home transit access is a significant variable and is positively associated with an individual's likelihood to be multimodal. If one's home location is within ¼ miles of a transit stop, this person is much more likely to be multimodal, compared to individuals whose home locations are not near a transit stop.

Table 3-12: Multinomial Logit Regression Model Coefficients

		Multimodal Auto		Multimodal Green		Unimodal Green	
	(Intercept)	1.213	***	-1.275	.	0.314	
Gender	Female			-			
	Male	0.178	*	0.711	**	0.594	*
License	No			-			
	Yes	-1.324	***	-2.410	***	-2.536	***
Age Group	Senior			-			
	Middle	0.041		0.477		0.043	
	Middle Young	0.105		0.752		-0.261	
	Young	0.204		1.484	**	0.500	
Education	Graduate			-			
	Undergraduate	-0.341	***	-0.695	*	-0.326	
	High School	-0.594	***	-1.067	*	-0.096	
Employed	No			-			
	Yes	-0.025		0.020		-0.647	*
Household with Children	No			-			
	Yes	-0.289	**	-0.163		-0.361	
Household Income	Very High			-			
	High	-0.029		-0.154		-0.134	

		Multimodal Auto		Multimodal Green		Unimodal Green	
	Medium	-0.371	***	-0.182		-0.700	.
	Low	-0.578	***	-1.038	*	-1.458	**
Household Operating Vehicles	1			-			
	2	-0.532	***	-2.262	***	-1.739	***
	3	-0.879	***	-1.911	***	-2.607	***
	4+	-0.938	***	-1.956	*	-2.059	*
	zero	2.124	***	4.290	***	3.880	***
Home Transit Access	No			-			
	Yes	0.606	***	1.227	***	0.471	.

*Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Modeling with One-Day Sample

The same modeling exercise was performed using only the first day of the GPS data, as if the households had only been equipped for one day. The one-day sample contains 30,331 trips from 5,242 individuals, indicating that some individuals did not make a trip during the first day. Table 3-13 summarizes the distributions of trips by modes across the four modality styles from the one-day sample. Compared to results shown for the 3-day sample, the one-day sample shows a higher percentage of unimodal individuals. This difference is likely a reflection of under-representation of mode variability in a one-day sample.

Table 3-13: Summary Statistics, by Modality Style with One-Day Sample

Modality Style	# Persons	% of Total	# Trips in 3 Days	# Auto Trips	# Bike Trips	# Transit Trips	# Walk Trips	# Daily Trips/Person	Average Daily Distance/Person (miles)	Average Daily Duration/Person (hrs)
Unimodal Auto	3,456	66%	19,548	19,451	4	2	91	5.66	34.37	1.29
Multimodal Auto	1,296	25%	8,688	5,431	356	802	2,099	6.70	33.98	1.82
Multimodal Green	217	4%	1,217	3	188	477	549	5.61	17.35	1.42
Unimodal Green	273	5%	878	12	245	132	489	3.22	5.51	0.78
Total	**5,242**	**100%**	**30,331**	**24,897**	**793**	**1,413**	**3,228**	**5.79**	**32.07**	**1.40**

Table 3-14 summarizes the multinomial logit regression coefficients using the one-day sample. There are no stark differences in the sign and magnitude of the coefficients between the three-day and the one-day samples, but the p-values for the coefficients from the 1-day sample are generally larger than those from the 3-day sample. The lack of significant modeling improvements is likely due to the fact that a sampling period of 3 days is still quite short in the context of within-person variability. When the sampling period increases to 20 weekdays, there is considerable improvement in modeling efficiency (for a more detailed discussion, see Xu [2010]).

Table 3-14: Multinomial Logit Regression Model Results from One-Day Sample

		Multimodal Auto		Multimodal Green		Unimodal Green	
	(Intercept)	0.726	*	-0.451		1.053	*
Gender	Female			-			
	Male	0.149	.	0.566	**	0.425	*
License	No			-			
	Yes	-1.180	***	-2.288	***	-2.333	***
Age Group	Senior			-			
	Middle	0.144		0.153		0.124	
	Middle Young	0.206		0.337		-0.012	
	Young	0.365	.	0.865	.	0.115	
Education	Graduate			-			
	Undergraduate	-0.330	***	-0.647	**	-0.246	
	High School	-0.632	***	-1.003	*	-0.194	
Employed	No			-			
	Yes	-0.058		0.148		-0.609	**
Household with Children	No			-			
	Yes	-0.157		-0.282		-0.467	*
Household Income	Very High			-			
	High	0.011		-0.305		-0.349	
	Medium	-0.251	*	-0.493	.	-0.629	*
	Low	-0.541	**	-1.201	**	-0.870	*
Household Operating Vehicles	1			-			
	2	-0.547	***	-1.761	***	-1.219	***
	3	-0.880	***	-1.772	***	-1.843	***

		Multimodal Auto		Multimodal Green		Unimodal Green	
	4+	-0.948	***	-2.657	**	-1.679	***
	zero	1.662	***	3.563	***	2.908	***
Home Transit Access	No			-			
	Yes	0.382	***	1.242	***	0.378	.

*Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

3.4 Conclusions and Recommendations

This study is the first to analyze modality styles in the United States using multiday GPS data. Previous studies focusing on modality styles in the United States have employed data collected from phone interviews (Buehler and Hamre, 2014) and postal surveys (Diana and Mokhtarian, 2009). The use of multiday GPS data in modality studies has the potential to enhance such analyses, presumably because GPS recorded data are not as susceptible to respondent underreporting (Ogle et al. 2005). In the meantime, multiday GPS data also pose a great challenge with respect to data screening and analysis. This report has outlined procedures that could prove useful in similar studies using multiday GPS data from other regions. The results from the data analysis align well with findings from previous studies on multimodality in Europe and the United States, adding to the literature the understanding of the correlation between modality style and traveler characteristics. This section first summarizes key findings, and then discusses the areas of analytical uncertainty that could limit the accuracy of the results. Finally, recommendations for further study are provided, based upon current and previous experience analyzing multiday GPS data for modality styles.

Key Findings

Many of the findings based on the 3-day wearable GPS sample collected in the San Francisco Bay Area are consistent with findings from previous studies. This study has shown the applicability of multiday GPS data in the analysis of modality styles, even if the sampling period is relatively short. Key findings on the relationship between modality styles and travel demographic and socioeconomic characteristics are summarized below.

- **Household vehicle ownership**. The MTC wearable GPS sample confirms that household vehicle ownership has a strong correlation with an individual's modality style. Individuals from households with zero vehicles show significantly different shares of modality styles than their counterparts from households with vehicles. The size of the effect tends to decrease as the number of vehicles increases and as household income increases. Further work should examine the reason for zero car ownership to distinguish between "choice" versus "captive" zero vehicle households.

- **Possession of a driver's license**. Similar to vehicle ownership, the possession of a driver's license also has a strong association with one's modality style. Adults without a driver's license are significantly more likely to use alternative modes such as walk, bike, and transit. Due to data constraints, however, one cannot differentiate between individuals who are not able to obtain a driver's license versus those who choose not to obtain a driver's license.
- **Transit access**. The MTC dataset adds evidence to the literature that an individual from a household located near a transit stop is more likely to be multimodal and utilize alternative modes. This effect is the most pronounced when comparing the multimodal green style to the unimodal auto style, indicating that individuals living close to a transit stop may have significantly higher use of alternative modes, but the likelihood that such individuals will completely rely on alternative modes is also low.
- **Household presence of children**. Individuals from households with children are more likely to be unimodal auto. The effect is more readily seen in the low- and medium-income groups than in the high- and higher-income groups, presumably because households with higher incomes have more flexibility in terms of work schedule and the ability to secure childcare services. The presence of children is also an important factor when examining modality styles by individual employment status. Employed individuals from a household with at least one child are more likely to be unimodal auto.
- **Age**. The results show a general trend that unimodal auto travel increases with age group, consistent with motorization as age and lifestyle change over time (e.g., Kuhnimhof, et al., 2006; Buehler and Hamre, 2014). However, this trend is not statistically significant for the most part. A larger sample of senior adults is needed for a meaningful examination of the effects of aging on modality styles.
- **Other factors and potential interactions**. Household income, number of workers, individual education level, employment status, and gender are also correlated with one's modality styles to varying degrees. Findings related to these variables are consistent with previous studies. However, caution should be exercised when interpreting results due to the potential collinearity and interactions among predictor variables.

Areas of Analytical Uncertainty

The trip data employed in this analysis are derived from the GPS data files for the Bay Area portion of the Caltrans statewide household travel diary study. According to the survey documentation, each trip was processed using the Trip Identification and Analysis System (TIAS) software to identify potential trip ends (NuStats, 2013a), where the criteria for identifying potential trips was set to 120 seconds of dwell time (at a single location). The documentation also indicates that "GPS trip data were then visually reviewed by analysts to screen out traffic delays and other falsely identified stops with dwell times of 120 seconds or more, as well as to add stops that had dwell times of less than 120 seconds but had clear 'stop' characteristics (NuStats, 2013a)." Hence, the documentation indicates that manual manipulation of data was also employed to break trips into trip legs (e.g., pick-up and drop-off trips, at school, ATM stops, etc.). However, no clear criteria are provided, or the criteria are considered proprietary and not reported. The proliferation of short walk trips in the dataset indicated to the research team that many of these trips were likely facilitative (i.e., moving from an origin to the transportation mode) and did not have independent utility. The authors performed data screening to eliminate these facilitative walk trips from the multimodality analysis. However, a more detailed assessment of trip coding in the Caltrans database appears warranted based upon this preliminary review. It is especially important to assess whether there are any significant differences between the days for which travel diary data are available and the 2 days for which diary data were not available, and analysts had to apply professional judgment to the trip identification process.

As shown in the data-screening section, there are many uncertainties around data quality, and the analysts had to implement screening criteria to delete certain trips. Ideally, these quality issues would be better addressed at the data collection stage, rather than at the analysis stage. The introduction of data-screening criteria potentially reduces the replicability of the analyses, since different analysts may employ different screening criteria.

Travel diary data are only available for one of the 3 days that each household had participated in GPS data collection. For those trips collected on a travel diary day, mode choice was provided by the participant. For the two other days, the data analyst coding each trip had to select the mode from one of 29 available modes. The Caltrans documentation contains no information as to how modes were assigned by analysts. Inferences can be drawn as to travel mode based upon trip start and end points (when trips occur along coded transit routes, especially when travel was made by the same self-identified mode on the travel diary day) and second-by-second travel speed and acceleration. Given the number of trips identified by the authors for which derived average trip speed did not appear reasonable (based upon trip time and distance), a more detailed assessment of travel mode assignment is probably warranted for these data.

Transit stops were integrated from the GTFS datasets available from the public transit agency websites. Only two of the smaller systems did not have available data. Although the likelihood that the lack of these data affect the transit-oriented analyses, there are a significant number of users that appear to live in the SolTrans (Solano Transit) service area based upon the zip code analysis. Adding the transit stop data for the missing systems could improve the analysis. In addition, the presence of a nearby transit stop does not infer the level of transit service available to a participant. It is encouraging to have identified a strong correlation between transit use and multimodality, but this is not too surprising. Supplemental analysis of transit service frequency and integration of a transit service frequency or level-of-service variable for home, work, school, daycare, and other primary travel locations could enhance the analysis.

A number of additional coding issues are discussed within the paper and report appendices. For example, analytical findings may also be affected by the presence of households in the dataset that use their personal vehicles for commercial purposes, coding of vehicle ownership and operable vehicles, aggregation of motorcycle and scooter modes, and coding of place and activity types (which can affect identification of individual trips, tours, and travel modes). Ensuring that the households in the Bay Area dataset are representative was beyond the scope of this analysis. However, based upon the low percentage of single-parent households in the dataset, a second look at this issue may be warranted.

The analyses presented in this chapter identify what appear to be significant and strong relationships between multimodality and sociodemographic and transit variables. Although there are some data and analytical uncertainties associated with the aforementioned issues, it seems unlikely that additional QA/QC of the Caltrans dataset will negate these basic findings.

Recommendations for Further Study

Multiday GPS data have proven to be useful in the analysis of personal modality styles. However, successful application of GPS data to modality studies relies heavily on careful planning in the survey design and meticulous screening of data in quality assurance.

To facilitate the understanding of modality styles, a few questions need to be added to the existing survey design. First, the survey should explicitly ask for the reason why a respondent does not possess a driver's license. Second, it is important to code home, work, school and any other locations that a respondent frequents. The analysis presented in this report was not able to differentiate modality styles by trip purpose due to concerns related to the coding of O-D types. Additional research into the relationship between modality style and trip purpose/activity elements may be warranted. That is, the predominant end-use activity or activities undertaken by the traveler may be confounded with some of the other variables employed in these analyses. Third, the existing survey does not have any direct information on whether an individual is married or a parent, yet prior research (Vij et al., 2011) has pointed out that these variables are significantly correlated with an individual's modality style.

The data QA/QC process plays a critical role in the accurate representation of the shares of modality styles. The quality of data related to GPS slow modes, such as walk trips, should be scrutinized. In general, the data coding issues identified in this chapter and its appendices warrant further investigation. The researchers recommend that three independent teams be contracted to reprocess the GPS data stream and recode trip purpose and travel mode data for comparative purposes.

Besides survey design and data quality, directions regarding further analyses include the following:

- A supplemental analysis should include land-use data, incorporating such parameters as land-use density, land-use intensity, land-use mix, number of employment opportunities within walking distance of transit, and a variety of other factors.
- A more detailed assessment of household and traveler vehicle ownership and use should be conducted to assess whether use patterns within households are significantly correlated with individual multimodality.
- Given the findings with respect to presence of children and correlation with multimodality, an enhanced analysis that looks at the ages of these children and school status would likely enhance the results.
- Family lifecycle stage (e.g., singles, young couples, young couples with young children, middle-aged couples with children, empty-nesters, retirees, etc.) would be an interesting addition to the analyses presented to date (but would require significant effort, given the need to process data with scripts on the remote server).

Finally, professional analysts generally desire to perform work using their own analytical systems. Researchers that have been in the business of analyzing second-by-second data have developed considerable expertise, software programming skills, and scripts that are of value to private industry. As such, performing analyses on remote servers is less than ideal from both an efficiency and intellectual property standpoint. Enhancing remote server data connections and analytical capabilities is paramount if the current setup is to serve as a model for data analysis. However, some guarantees will also likely need to be established to protect the intellectual property rights of the analysts who have no choice but to place their scripts onto the remote server to have them run.

3.5 References

Chlond, B. (2012). Making People Independent from the Car-multimodality as a Strategic Concept to Reduce CO2 Emissions. In: Zachariadis, T. (ed.) Cars and Carbon: Automobiles and European Climate Policy in a Global Context, pp. 269–293. Springer, New York (2012).

Chum, G.L., and M.W. Burris (2008). Potential Mode Shift form Transit to Single-Occupancy Vehicles on a High Occupancy Toll Lane. In Transportation Research Record: Journal of the Transportation Research Board, No. 2072, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 10–19.

Nobis, C.: Multimodality: Facets and Causes of Sustainable Mobility Behavior. Transp. Res. Rec. 2010, 35–44 (2007)

Clifton, K. and C. Muhs (2012). "Capturing and Representing Multimodal Trips in Travel Surveys." Transportation Research Record: Journal of the Transportation Research Board 2285(-1): 74-83.

Cosma, S. (2013). Classification and Regression Trees, Carnegie Mellon University, Lecture 22 - http://www.stat.cmu.edu/~cshalizi/350/lectures/22/lecture-22.pdf, Last Accessed 04/17/2013.

Elango, V., and R. Guensler (2010). "An Automated Activity Identification Method for Passively Collected GPS Data." 3rd TRB Conference on Innovations in Travel Modeling. Tempe, Arizona. May 10-12, 2010.

Elango, V., R. Guensler, and J. Ogle (2007). “Day-To-Day Travel Variability in the Commute Atlanta Study.” Transportation Research Record. Number 2014. pp. 39-49. National Academy of Sciences. Washington, DC. 2007.

Elango, V., R. Guensler, and J. Ogle (2007). “Day-To-Day Travel Variability in the Commute Atlanta Study.” Transportation Research Record. Number 2014. pp. 39-49. National Academy of Sciences. Washington, DC. 2007.

Elango, V., S. Khoeini, Y. Xu, and R. Guensler (2013). “Longitudinal GPS Travel Data and Breach of Privacy via Enhanced Spatial and Demographic Analysis.” Transportation Research Record. Number 2354. pp. 86-98. National Academy of Sciences. Washington, DC. 2013.

Grava, S. (2002). Urban Transportation Systems. McGraw Hill.

Guensler, R. and J. Wolf (1999). “Development of a Handheld Electronic Travel Diary for Monitoring Individual Tripmaking Behavior.” Published in the CD-ROM Proceedings of the 77th Annual Meeting of the Transportation Research Board. Washington, DC. January 1999.

Khoeini, S, Y. Xu, R. Guensler (2014 Manuscript). Development of the I-85 HOV-to-HOT Post-Conversion Carpooling Survey and Sampling Plan. Georgia Institute of Technology, School of Civil and Environmental Engineering.

Kuhnimhof, T., B. Chlond, and S. von der Ruhren (2006). "Users of Transport Modes and Multimodal Travel Behavior Steps toward Understanding Travelers' Options and Choices." Transportation Research Record: Journal of the Transportation Research Board 1985(-1): 40-48.

Kuhnimhof, T., R. Buehler, M. Wirtz, and D. Kalinowska (2012). Travel Trends among Young Adults in Germany: Increasing Multimodality and Declining Car Use for Men. J. Transp. Geogr. 24, 443–450 2012.

Li H., R. Guensler and J. Ogle (2006). “Impact of Objective Route Attributes on the Choice of Primary Morning Commute Route.” Published in the CD-ROM Proceedings of the 85th Annual Meeting of the Transportation Research Board. Washington, DC. January, 2006.

Li, H., R. Guensler, and J. Ogle (2005). "An Analysis of Morning Commute Route Choice Patterns Using GPS Based Vehicle Activity Data." Transportation Research Record. 1926. pp. 162-170. National Academy of Sciences. Washington, DC. 2005.

Li, H., R. Guensler, J. Ogle, and J. Wang (2004). "Using Global Positioning System Data to Understand Day-to-Day Dynamics of Morning Commute Behavior." Transportation Research Record. Number 1895. pp. 78-84. National Academy of Sciences. Washington, DC. 2004.

Li, H., R. Guensler, J. Ogle, J. Jun, and V. Elango (2007). "Descriptive Analysis of Long-Distance Travel by Personal Vehicles Using 2004 Commute Atlanta Data." 86th Annual Meeting of the Transportation Research Board, Washington, DC. January 2007.

Neter, J., W Wasserman, and M. Kutner (1990). Applied Linear Statistical Models, Regression, Analysis of Variance, and Experimental Designs, Third Edition, Irwin Inc., ISBN 0-256-08338-X, 1990.

NuStats, LLC (2013a). California Department of Transportation 2010-2012 California Household Travel Survey Final Report. Version 1.0. Arlington, TX. June 14, 2013.

NuStats, LLC (2013b). California Department of Transportation 2010-2012 California Household Travel Survey Final Report Appendix. Version 1.0. Arlington, TX. June 14, 2013.

Ogle, J., Guensler, R., and V. Elango (2005). "Commute Atlanta Value Pricing Program: Recruitment Methods and Travel Diary Response Rates." Transportation Research Record. Number 1931. pp. 28-37. National Academy of Sciences. Washington, DC. 2005.

Pearre, N., W. Kempton, R. Guensler, and V. Elango (2011). "Electric Vehicles: How Much Range is Required for a Day's Driving?" Transportation Research C. Volume 19, Issue 6, pp. 1171-1184. December 2011.

Pearre, N., W. Kempton, R. Guensler, and V. Elango (submitted). "Electric Vehicle Battery Size, Recharge Rate, and Charging Locations: Jointly Meeting Travel Requirements." Transportation Research C.

Santini, D., Y. Zhou, V. Elango, Y. Xu, and R. Guensler (2014). Daytime Charging: What is the Hierarchy of Opportunities and Customer Needs? Case Study based upon Atlanta Commute Data (14-5337). 93rd Annual Meeting of the Transportation Research Board. Washington, DC. January 2014.

Schönfelder, S., H. Li, R. Guensler, J. Ogle, K.W. Axhausen (2006). "Analysis of Commute Atlanta Instrumented Vehicle GPS Data: Destination Choice Behavior and Activity Spaces." Published in the CD-ROM Proceedings of the 83rd Annual Meeting of the Transportation Research Board. Washington, DC. January, 2006.

Vij, A., A. Carrel and J. L. Walker (2011). "Capturing Modality Styles Using Behavioral Mixture Models and Longitudinal Data." University of California at Berkley.

Wolf, J., R. Guensler, and W. Bachman (2001). "Elimination of the Travel Diary: Experiment to Derive Trip Purpose from GPS Travel Data." Transportation Research Record. Number 1768. pp. 125-134. National Academy of Sciences. Washington, DC. 2001.

Xu, Y. (2010). "Effective GPS-Based Panel Survey Sample Size for Urban Travel Behavior Studies." Doctoral Dissertation, Georgia Institute of Technology, May 2010.Xu, Y., and R. Guensler (2011). "Effective GPS-Based Panel Survey Sample Size Analysis for Before-and -after Studies using Generalized Estimating Equation Methods." 90th Annual Meeting of the Transportation Research Board, Washington, DC. January 2011.

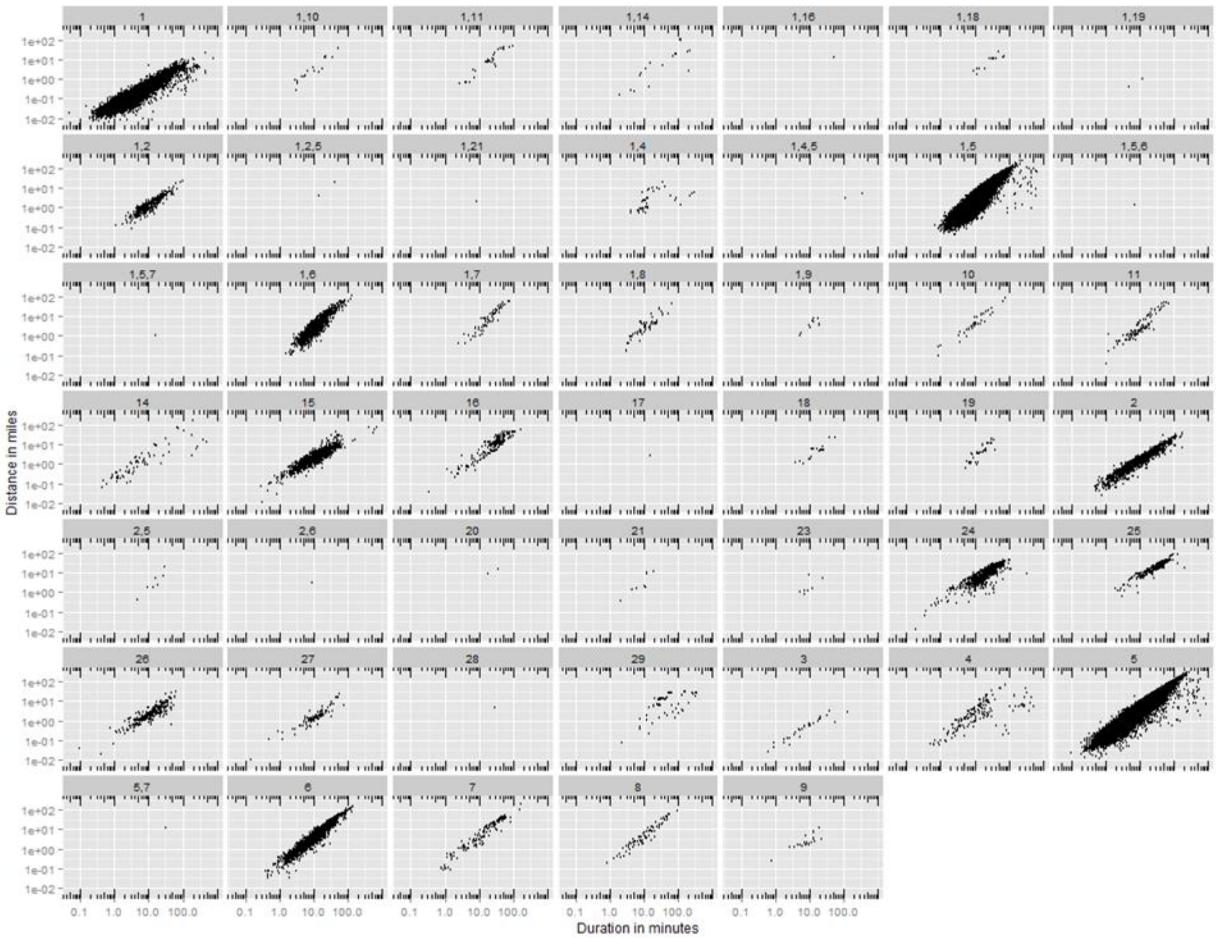
Appendix A. Assessment of Walk Trips

According to the Caltrans report (NuStats, 2013a) Section 5.5:

> "This data deliverable includes all 'GPS/DIARY complete' GPS households. There are a total of 3,491 GPS vehicles in the 1,866 complete vehicle and vehicle OBD GPS households and 8,202 GPS persons in the 3,871 wearable GPS households. The 3,491 GPS vehicles captured 12,380 GPS trips on the assigned travel days, compared to 11,609 reported trips for these same vehicles. The 8,202 GPS persons captured 45,986 GPS trips on their assigned travel day compared to 39,995 reported trips for these same participants. So, across all GPS samples, a total of 58,366 GPS trips were collected compared to 51,604 reported trips for the same vehicles or persons."

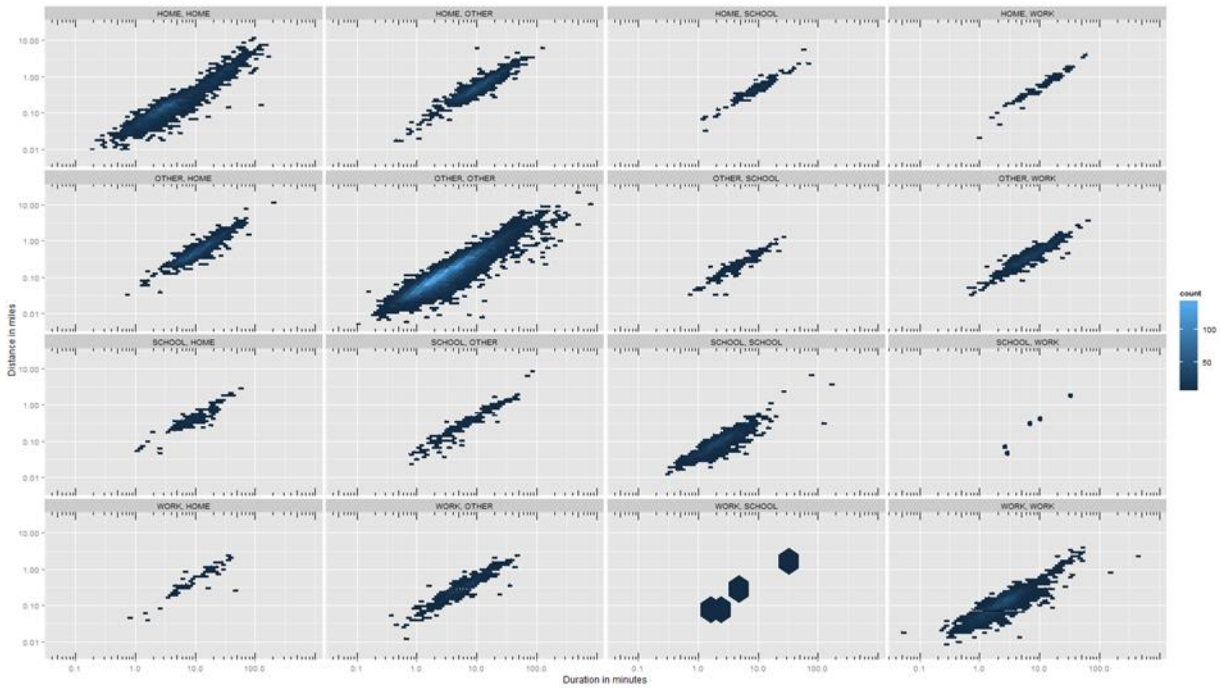
Given the issues related to GPS data collection at low speeds, walk trips tend to be problematic. At speeds less than 3 mph, latitude and longitude and GPS speed accuracy varies significantly (known as GPS wander). The team investigated walk trips in detail to make sure that false walk trips are removed. Figure 3-21 shows duration-distance scatterplots by unique modes of each trip. A visual assessment indicates that there are many walk trips (Mode 1) with distances traveled less than 0.1 miles (approximately 500 ft.). There are also walk trips with unreasonably high speeds (about 40 mph), and very long walk trips (close to 70 miles or more than 10 hours).

Figure 3-21: Duration-Distance Scatter Plot Paneled by Unique Trip Mode



Walk trips are further assessed by origin-destination (O-D) types, as shown in Figure 3-22. A large number (107,000) home-home, work-work, and school-school trips are included in the data.

Figure 3-22: Density Plot of Walk Trips Paneled, by Origin and Destination Types



The Caltrans report (NuStats, 2014a) contains a Reporting Exceptions subsection (Section 5.5.1), in which the authors indicate that additional typical unreported trip types include work-related trips from the office (discussed earlier), loop trips (i.e., those that start and end at the same location), and on-site travel (i.e. trips that are conducted on the premises of one property, like a hospital or apartment complex). In Section 5.1.1 of the Caltrans report, the authors indicates that GPS trips were flagged as loop trips whenever a GPS trip was detected in which the origin and destination were the same location. However, according to the coding rules, this should only occur when the Place Type is "Other."

Caltrans trip data are coded first by place type, and then trip purpose for each place type:

Trip Data

- Place name
- Address, including zip code
- Place arrival time
- Other individuals traveling with you

Place Type (PTYPE, O_PTYPE)

- 01 Home
- 02 Primary job
- 03 School
- 04 Second job

- 05 Transit stop
- 07 Other place, please specify

Activity Data

- What was your main activity there?
- Once you arrived, what activities did you/he/she do here? (up to three codes can be applied)

Activity Types (APURP) for the Home Place

- 01 Personal activities, such as sleeping, personal care, leisure, chores
- 02 Preparing meals/eating
- 03 Hosting visitors/entertaining guests
- 04 Exercise (with or without equipment)/playing sports
- 05 Study / schoolwork
- 06 Work for pay at home (telecommuting equipment)
- 07 Using computer/telephone/cell or smart phone or other communications device for personal activities
- 08 All other activities at my home

Activity Types (APURP) for the Work and Volunteer Place

- 09 Work/job duties
- 10 Training
- 11 Meals at work
- 12 Work-sponsored social activities, such as holiday or birthday celebrations, etc.
- 13 Non-work-related activities, such as social clubs, etc.
- 14 Exercise/sports
- 15 Volunteer work/activities
- 16 All other work-related activities at my work

Activity Types (APURP) for the School Place

- 17 In school/classroom/laboratory
- 18 Meals at school/college
- 19 After school or non-class-related sports/physical activity
- 20 All other after school or non-class related activities, such as the library, band rehearsal, clubs, etc.

Activity Types (APURP) for the Transit (Quick Stop) Place

- 21 Change type of transportation/transfer (walk to bus, walk to/from parked car)
- 22 Pick-up/drop-off passenger(s)
- 23 Drive-through meals (snacks, coffee, etc.)
- 24 Drive-through other (ATM, bank, etc.)

Activity Types (APURP) for the Other Place

- 25 Work-related (meeting, sales call, delivery)
- 26 Service private vehicle (gas, oil, lube, repairs)
- 27 Routine shopping (groceries, clothing, convenience store, household maintenance)
- 28 Shopping for major purchases or specialty items (appliance, electronics, etc.)
- 29 Household errands (bank, dry cleaning, etc.)
- 30 Personal business (visit government office, attorney, accountant)
- 31 Eat meal at restaurant/diner
- 32 Health care (doctor, dentist, eye care, veterinarian, etc.)
- 33 Civic/religious activities
- 34 Outdoor exercise (playing sports/jogging, bicycling, walking, walking the dog, etc.)
- 35 Indoor exercise (gym, yoga, etc.)
- 36 Entertainment (movies, watch sports, etc.)
- 37 Social/visit friends/relatives
- 38 Other, please specify
- 39 Loop Trip (for interviewer only - not listed on diary)

The Caltrans documentation (NuStats, 2014a) indicates that a significant cause of mismatch between travel-diary-reported trips and GPS-reported-trips includes "loop trips (i.e., those that start and end at the same location) and on-site travel (i.e. trips that are conducted on the premises of one property, like a hospital or apartment complex)." However, the activity type known as "Loop Trip" is only be provided for "Other" places; i.e., loop type purposes are not provided for round trips to home, work, and school places. There is no trip purpose coding option to indicate "on-site travel" for any place type. For home-home trips, trip purposes (APURP) may be identified as: 01 Personal activities, such as sleeping, personal care, leisure, chores; 04 Exercise (with or without equipment)/playing sports; or 08 All other activities at my home. However, there is no text field for the respondent to elaborate on the other activities code (APURP=08). For Work and Volunteer Places, such loop travel from work to work might relate to any of the available purpose codes: 09 Work/job duties; 10 Training; 11 Meals at work; 12 Work-sponsored social activities, such as holiday or birthday celebrations, etc.; 13 Non-work-related activities, such as social clubs, etc.; 14 Exercise/sports; 15 Volunteer work/activities; 16 All other work-related activities at my work. As noted for home trips, there is no text field for the respondent to elaborate on the other activities code (APURP=16). For School Places, such loop travel from work to work might relate to any of the available purpose codes: 17 In school/classroom/laboratory; 18 Meals at school/college; 19 After school or non-class-related sports/physical activity; and 20 All other after school or non-class related activities, such as the library, band rehearsal, clubs, etc.

Similarly, there is no text field for the respondent to elaborate on the other activities code (APURP=20).

> "According to the rules of this study, loop trips should have been reported whenever their purpose (e.g., exercise or walk the dog) was not tied to the purpose of the previous trip. This means that a Loop Trip made from home is a valid trip whereas a loop walk trip in a park preceded by a drive to the park for exercise purposes should not have been reported. A total of 2,637 loop trips were identified, 1,969 of which were reported by participants. Furthermore, 3,797 other non-transportation or on-site trips were found that were not required to be reported. (NuStats, 2014a, Section 5.1.1)

The above statement is true in that walking the dog from home should be reported as a trip according to the rules of the study. For the Home Place, a walk the dog trip could be identified as one of three options: 01 Personal activities, such as sleeping, personal care, leisure, chores; 04 Exercise (with or without equipment)/playing sports; or 08 All other activities at my home. Given that only a small fraction of home, work, and school trips (fewer than 2,000 out of more than 107,000 trips) are flagged as loop trips for trip purpose, it does not appear that "GPS trips were flagged as loop trips whenever a GPS trip was detected in which the origin and destination were the same location." That is, the trip purpose codes do not appear to have been changed. This is as it should be given that a Loop Trip purpose is not available home, work and school trips, and QA/QC analysis should be able catch the mismatch.

The Caltrans report and Appendix by NuStats do not provide enough details to assess how the trip purpose coding was handled for home-home, work-work, and school-school trip in the GPS analysis, given the accuracy and sensitivity of the three different GPS units deployed and the variety of purposes that are available for these round trips. The team is still assessing this issue.

The results of the day-to-day variability in mode choice are a direct function of how the analytical team defines a true vs. false trip, especially for walk trips. The team used the following criteria to remove walk trips (mode ID = 1) from analysis:

1. Walk trips with travel distance less than 150 feet (one-half block) are defined as On-site Trips and removed from the analysis (per Clifton and Muhs, 2012).
2. Walk trips with duration less than 5 minutes are defined as non-trips in accordance with survey coding rules (NuStats, 2014b; Page 318) and removed from the analysis.
3. Walk trips with speeds exceeding 8 mph (a typical jogger can run approximately 6 mph) are removed from the analysis as a trip mode coding error.

4. If the origin and destination of a coded Loop Trip for Other Trip Purpose are different, a coding error is assumed and the trip are flagged for removal.

Appendix B. Commercial Vehicle Households

The Caltrans report (NuStats, 2013a) indicates that in some household travel surveys, by design and travel diary instruction, work-related trips (i.e., commercial use of personal auto) may not reported in the travel diary and not collected during the retrieval call. Unfortunately, the GPS data do contain work-related trips and will affect household- and vehicle-level statistics. This is one source of mismatch between diary trips and GPS trips reported (NuStats, 2013a). Previous research in the Commute Atlanta study indicated that households and persons that reported using their vehicles on a regular basis for commercial purposes had such a significant impact on daily household and vehicle travel activity, that a ‘commercial vehicle use’ variable should be employed in household sample stratification (Elango, et al., 2007). Such a variable is not present in the Caltrans Household Travel Survey for the households or individual vehicles (this variable would appear in Recruitment Script Section 3.0, Vehicle Roster) and therefore cannot be used to flag households for removal. A large difference between GPS reported data and diary reported data for a household may be a reason enough to flag the household for removal from day-to-day mode choice variability analysis. Further work will have to address whether it will be possible to identify households that use their vehicles extensively for work-related trips through a combination of diary and GPS analysis based upon household travel diary data. The Caltrans documentation indicates that of the 3,055 work-related GPS trips (presumably based upon departure from the work location and return to the work location), 2,066 (68%) had been reported in travel diaries and 989 had not (NuStats, 2014a). Hence, it may be possible to flag high work-use households using the travel diaries. Given that the use of the vehicle for work purposes significantly affects mode choice, not just for the one vehicle, but potentially for the entire household due to joint decision-making processes, the goal is to remove these household days from the analytical data set.

Appendix C. Long-Distance Travel

The Caltrans Report Section 5.5.1 indicates that in some household travel surveys, external to external trips (i.e., those that have origins and destinations outside of the planning regions) are not reported in the travel diary and not collected during the retrieval call, and that this was by design, per the travel diary instructions (NuStats, 2013a). However, these data should be present in the GPS data set. This makes the analysis problematic because days that include long-distance and extra-regional trips are not likely to include representative intra-regional variability in mode choice (i.e. the mode selection is constrained by the long-distance travel).

The team is excluding travel conducted outside the region by eliminating all households in which long-distance trips occurred. A long-distance trip is defined as a trip longer than 100 miles one-way, in line with Diana and Mokhtarian (2009).

Appendix D. Treatment of Travel Modes in the Analyses

The travel mode definitions for the Caltrans travel diary data set can be found in the California Department of Transportation 2010-2012 Travel Survey Final Report Appendix (NuStats 2013b). The full list of the 29 modes is listed below.

- Walk (Mode 01)
- Bike (Mode 02)
- Wheelchair/Mobility Scooter (Mode 03)
- Other Non-Motorized, including skateboards, etc. (Mode 04)
- Auto/Van/Truck Driver (Mode 05), Auto/Van/Truck Passenger (Mode 06)
- Motorcycle, Scooter, Moped (Mode 08)
- Taxi, Hired Car, Limousine (Mode 09)
- Rental Car/Vehicle (Mode 10)
- Private Shuttle, e.g. Super Shuttle, Employer shuttles, hotel shuttles, etc. (Mode 11)
- Greyhound Bus (Mode 12)
- Airplane (Mode 13)
- Other Private Transit (Mode 14)
- Local Bus/Rapid Bus (Mode 15)
- Express Bus/Commuter Bus - AC Transit, Golden Gate Transit, etc. (Mode 16)
- Premium Bus, e.g. Metro Orange and Silver Lines in Los Angeles (Mode 17)
- School Bus (Mode 18)
- Public Transit Shuttle, e.g. Dash, Emery Go-Round, etc. (Mode 19)
- AirBART/LAX Flyaway (Mode 20)
- Dial-A-Ride/Paratransit, access services, etc. (Mode 21)
- Amtrak Bus (Mode 22)
- Other Bus Rail/Subway (Mode 23)
- BART, Metro Red, Purple Line (Mode 24)
- Amtrak, Caltrain, Coaster, Metrolink (Mode 25)
- Light Rail - Metro Blue/Green/Gold Line, Muni Metro, Sacramento Light Rail, San Diego Sprinter/Trolley/Orange/Blue/Green, VTA Light Rail (Mode 26)
- Street Car or Cable Car (Mode 27)
- Other Rail (Mode 28)
- Ferry/Boat (Mode 29)

Each of the modes is examined in the bullets below and decisions are made as to whether to retain, exclude, or combined certain modes for the analyses presented in this report.

- Walk (Mode 01) - For all persons that are fully mobile, every trip begins and ends with a walking end, whether it is from home to driveway to gain access to a car, or from home directly to work on foot. Walk trips are tracked as a mode in all analyses. However, accounting for walk trips raises some significant potential issues:

- Walk trips are important elements in trip chaining, whether the walk legs of the trip are independent of other modes or are simply providing access to another mode. By definition, any trip that is independent of another mode from one location to another, where an activity is undertaken at that destination, can be defined as a walking trip. For example, walk from home to shopping and walk from shopping to home include two walk trips (or one shopping walk tour). However, some walk trips can be defined as incidental. For example a walk trip to transit can be defined as separate legs of a transit trips or as a combined mode (transit mode with walk access). For example, a three-component trip chain that includes only: 1) walk to transit station, 2) transit to transit station, and 3) walk to work, could be defined as including zero walk trips as none have independent utility. In this case, the two walking elements simply provided access and egress modes to a transit work trip and were not undertaken with an independent trip purpose. However, a four-component trip chain that includes 1) walk to transit station, 2) transit to transit station, and 3) walk to shopping, and 4) walk to work, definitely contains two independent walk trips (walking to the shopping location for shopping and walking to work from shopping). The question that arises is whether the first leg is also coded as a walk trip, or a transit trip with walk access mode.

- Bike (Mode 02) - Bicycle trips can extend the range of transit trips and are tracked as a mode in all analyses.
- Wheelchair/Mobility Scooter (Mode 03) - The number of disabled individuals in the Bay Area study are not likely to be representative of the range of multimodal activity in region. Separate analyses are needed for the disability community. As such, household using this mode are included in the assessment.
- Other Non-Motorized, including skateboards, etc. (Mode 04) - As with wheelchairs, the number of non-motorized users in the Bay Area study are not likely to be representative of the range of multimodal activity in region. This mode is grouped with Mode 02.
- Auto/Van/Truck Driver (Mode 05), Auto/Van/Truck Passenger (Mode 06), and Carpool/Vanpool (Mode 07) - These three modes are all undertaken in private vehicles. The difficulty in differentiating between the three modes is that there is no clear definition of carpool provided in the Caltrans report, or appendices (NuStats 2013a, 2013b). Interview scripts and written documentation do not clearly define how respondents were asked to define the term "carpool" when reporting trip modes. The very low frequency of carpooling (0.6% of trips) reported in Mode Choice Table 8.3.1 indicates that

respondents may not have interpreted the term consistently across the sample. The average travel distances are longer for the carpool/vanpool mode.

- Motorcycle, Scooter, Moped (Mode 08) - Although two-wheel motor vehicles can extend the range of transit trips and could be tracked as a separate mode in all analyses, the authors have elected to cluster this mode with automobiles as has been done in previous research.
- Taxi, Hired Car, Limousine (Mode 09) - Vehicles for hire constitute an active alternative transportation mode in the Bay Area and are included in the modal analyses.
- Rental Car/Vehicle (Mode 10) - This mode is grouped as auto travel.
- Private Shuttle, including Super Shuttle, Employer shuttles, hotel shuttles, etc. (Mode 11) - This mode is grouped as transit.
- Greyhound Bus (Mode 12) - This mode is grouped as transit.
- Airplane (Mode 13) - Because long-distance travel is not being included in the assessment, households using this mode will not be included in the analysis.
- Other Private Transit (Mode 14) - This mode is grouped as transit Local Bus/Rapid Bus (Mode 15) - Local buses will be tracked as an alternative mode in the analyses prepared for this report.
- Express Bus/Commuter Bus, e.g. AC Transit, Transbay, Golden Gate Transit, etc. (Mode 16) - Commuter buses will be tracked an alternative mode in the analyses prepared for this report.
- Premium Bus, e.g. Metro Orange and Silver Lines in Los Angeles (Mode 17) - These modes are not used in the Bay Area and will be excluded from the analyses.
- School Bus (Mode 18) - Given the large number of school bus trips in the data set, this mode will be tracked ad analyzed in this report.
- Public Transit Shuttle, e.g. Dash, Emery Go-Round, etc. (Mode 19) - These are considered alternative modes.
- AirBART/LAX Flyaway (Mode 20) - This mode is grouped as transit, but because long-distance travel is not being included in the assessment, this mode is not included in the analysis.
- Dial-A-Ride/Paratransit, access services, etc. (Mode 21) - As with wheelchairs, the number of samples from the disability community in the Bay Area study are not likely to be representative of the range of regional multimodal activity. Separate representative sampling in the disability is required. As such, households using this mode are not included in the assessment.
- Amtrak Bus (Mode 22) - The Amtrak bus serves as a connection to Amtrak and can be an alternative mode itself. This mode will be omitted from the analyses in this report.

- Other Bus Rail/Subway (Mode 23) - The mode is not sufficiently defined and may be confounded with bus, light rail, and heavy-rail. This mode is grouped as transit.
- BART, Metro Red, Purple Line (Mode 24) - BART will be tracked as a heavy-rail mode in this report. Although Muni provides more frequent and diverse rail service, some users with BART passes may use BART for short distance travel in San Francisco, but this should not significantly impact the analyses.
- Amtrak, Caltrain, Coaster, Metrolink (Mode 25) - Caltrain serves as a heavy-rail commuter line and this mode is aggregated with BART under heavy-rail service.
- Light Rail, including Metro Blue/Green/Gold Line, Muni Metro, Sacramento Light Rail, San Diego Sprinter/Trolley/Orange/Blue/Green, VTA Light Rail (Mode 26) - The light rail modes tend to provide more.
- Street Car or Cable Car (Mode 27) - Street cars, and to some extent cable cars, provide similar service to Muni and is aggregated with other light rail modes.
- Other Rail (Mode 28) - Other rail is not sufficiently defined to ensure that the mode selection is not confounded with other heavy-rail modes. This mode is grouped as transit.
- Ferry/Boat (Mode 29) - Water transport (ferries) is an active alternative transportation mode in the Bay Area and is included in the modal analyses.

Appendix E. Discussion of Mode IDs

The 22 modes listed in the main NuStats report (NuStats, 2013a) in Table 5.5.1.1 (Travel Modes Included in Matching Process) do not match the 29 modes in the main report Appendix (NuStats, 2013b). In the Appendix data dictionaries, Mode ID 10 is Rental Car; this mode does not appear in main report Table 5.1.1.1. The insertion of additional modes in the Appendix results in different modes being assigned to the majority of Mode IDs (10-29) as illustrated in Table 3-15.

Table 3-15: Differences Noted in Caltrans Reported Mode ID Assignments

Mode ID	22 Modes Employed in Main Report Table 5.5.1.1 Mode Descriptions	29-Modes Employed in Appendix Data Dictionaries
1	Walk	Walk
2	Bike	Bike
3	Wheelchair / Mobility Scooter	Wheelchair / Mobility Scooter
4	Other Non-Motorized	Other Non-Motorized (Skateboard, Etc.)
5	Auto / Van / Truck Driver	Auto / Van / Truck Driver
6	Auto / Van / Truck Passenger	Auto / Van / Truck Passenger
7	Carpool / Vanpool	Carpool / Vanpool
8	Motorcycle / Scooter / Moped	Motorcycle / Scooter / Moped
9	Taxi / Hired Car / Limo	Taxi / Hired Car / Limo
10	Private Shuttle (Super Shuttle, Employer, Hotel)	Rental Car / Vehicle
11	Greyhound Bus	Private Shuttle (SuperShuttle, Employer, Hotel, etc.)
12	[Not reported in Table 5.5.1.1]	Greyhound Bus
13	Other Private Transit	Airplane
14	Local Bus, Rapid Bus	Other Private Transit
15	Express Bus / Commuter Bus (Golden Gate, AC Trans)	Local Bus / Rapid Bus
16	[Not reported in Table 5.5.1.1]	Express Bus / Commuter Bus (Ac Transbay, Golden Gate Transit, Etc.)
17	School Bus	Premium Bus (Metro Orange / Silver Line)
18	Public Transit Shuttle	School Bus

Mode ID	22 Modes Employed in Main Report Table 5.5.1.1 Mode Descriptions	29-Modes Employed in Appendix Data Dictionaries
19	[Not reported in Table 5.5.1.1]	Public Transit Shuttle (Dash, Emery Go-Round, Etc.)
20	Dial-a-Ride / Paratransit (Access Services)	AirBART / LAX Flyaway
21	Amtrak Bus	Dial-A-Ride / Paratransit (Access Services, Etc.)
22	Other Bus	Amtrak Bus
23	[Not reported in Table 5.5.1.1]	Other Bus Rail/Subway:
24	[Not reported in Table 5.5.1.1]	BART, Metro Red / Purple Line
25	[Not reported in Table 5.5.1.1]	Ace, Amtrak, Caltrain, Coaster, Metrolink
26	[Not reported in Table 5.5.1.1]	Metro Blue / Green / Gold Line, Muni Metro, Sacramento Light Rail, San Diego Sprinter / Trolley / Orange/Blue/Green, VTA Light Rail
27	[Not reported in Table 5.5.1.1]	Street Car / Cable Car
28	[Not reported in Table 5.5.1.1]	Other Rail
29	[Not reported in Table 5.5.1.1]	Ferry / Boat

Appendix F. Vehicle Ownership Issues

Household vehicle ownership (HHVEH) is defined in the survey questions as:

- "How many motor vehicles are owned, leased, or available for regular use by the people who currently live in your household? Please be sure to include motorcycles, mopeds, and RVs."

The number of operational vehicles during the study period (VEHOP) is asked of households that report more than zero vehicles owned leased, or available:

- "[IF HHVEH>0] How many of these vehicles are operational and used regularly during the week?"

NOTE: No travel diary question is asked as to why such vehicles are not operational, meaning that the reason could be temporary (flat tire), semi-permanent (operable and stored in the backyard for future rehabilitation), or permanently disabled (scrap supply). Future surveys should clarify whether the vehicle disability is temporary or permanent.

The team noted that vehicle ownership corrections in the script are prompted when the number of vehicles reported as operational is greater than the number of vehicles owner, leased or available. These corrections were one-way, in that as long as the value for operational is less than total vehicles, total vehicle ownership is never questioned. As such, the team believes that the VEHOP variable is likely to be more accurate than the HHVEH variable. As such, the analyses use VEHOP rather than HHVEH in classification analyses.

Table 3-16 provides a summary of the operational vehicle variable (VEHOP), derived from Caltrans Appendix Table F.1.22, with the additional 2459 households that reported HHVEH=0 added to the first row. These 2459 households were presumably not asked the operational vehicle question because they had reported having zero vehicles owned, leased, or available for regular use. This resulting statewide table appears to contain extra households: 42786-42431= 355 households. Furthermore, the VEHOP table reports only the totals for coded numeric values, and does not report the number of households reporting 98/99. The team reviewed the VEHOP variable for the Bay Area sample but could not identify any reasons for discrepancies.

Table 3-16: California Sample, Number of Operating Vehicles (VEHOP) per Household

Number of Operating Vehicles	Frequency	Percent
n/a (HHVEH=0)	2,459	5.7%
0	354	0.8%
1	12,671	29.6%

Number of Operating Vehicles	Frequency	Percent
2	18,663	43.6%
3+	8,639	20.2%
Total	**42,786**	**100%**

Appendix G. Incorporation of Bay Area Transit Stop Data

The GTFS data for each of the transit providers in the San Francisco Bay Area were procured from all each agency's websites and coded by latitude and longitude for use in geospatial analysis. The authors coded more than 18,500 transit stops for the following systems: AC Transit, BART, Caltrain, Central Contra Costa, Fairfield-Suisun, SF Ferry Systems, Marin Transit, SamTrans, Santa Rosa City, SF Muni, Sonoma County Transit, and VTA. Figure 3-23 illustrates the stops for the entire Bay Area, and Figure 3-24 through Figure 3-29 provide some of the more detailed slides by area and service. Almost every location in the City of San Francisco is located within one-quarter mile of a transit stop.

Figure 3-23: Map of Bay Area Transit Stops

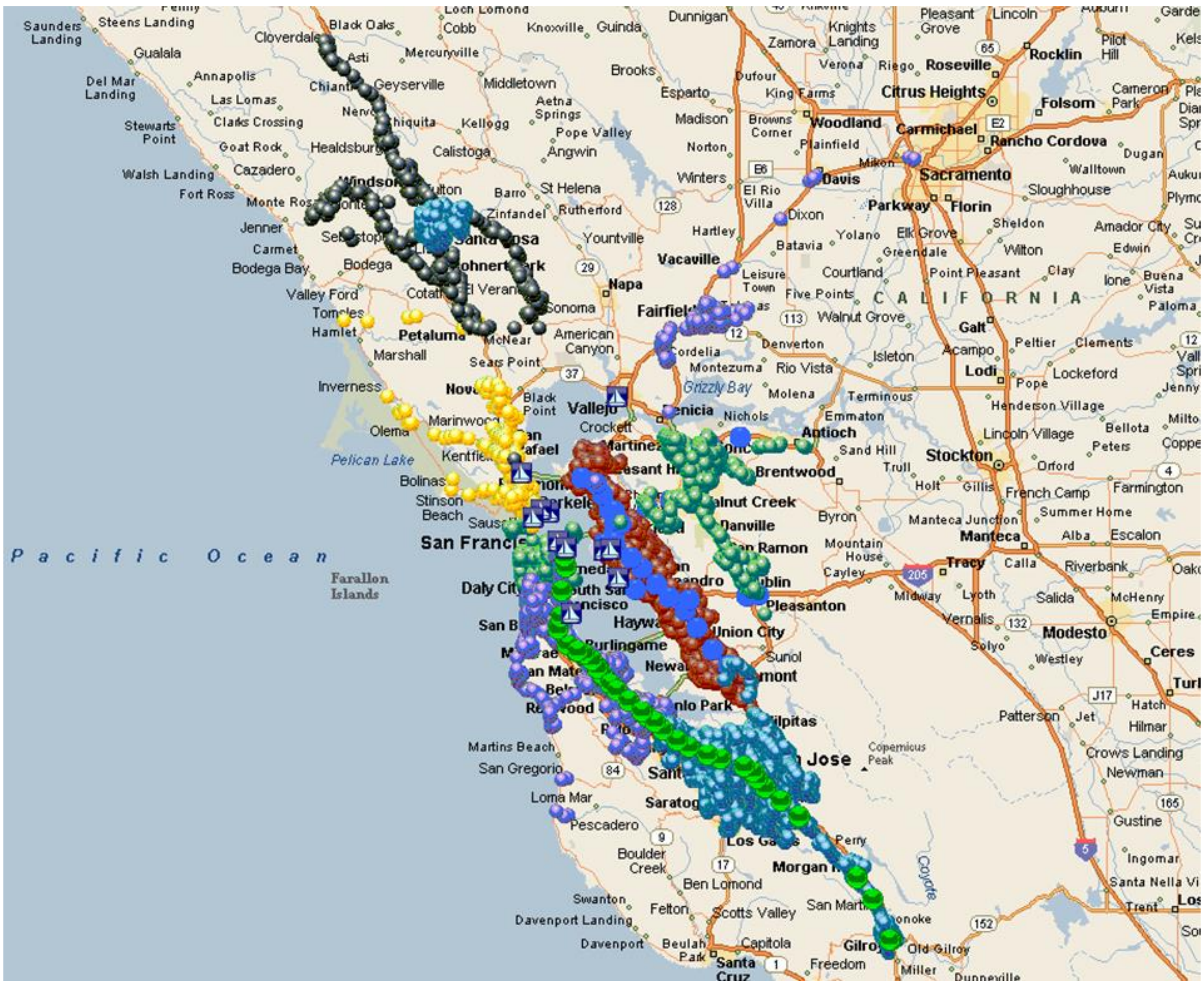


Figure 3-24: Map of Bay Area Transit Stops by Region—San Francisco MTA Stops

Figure 3-25: Map of Bay Area Transit Stops by Region—BART Station Locations

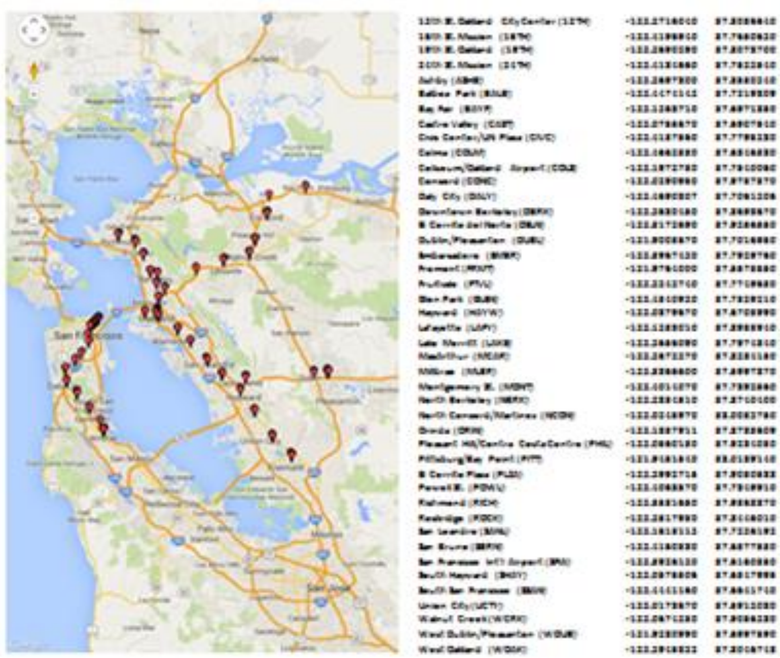


Figure 3-26: Map of Bay Area Transit Stops by Region—AC Transit and Central Contra Costa Transit Stops

Figure 3-27: Map of Bay Area Transit Stops by Region—South Bay Transit Stops

Figure 3-28: Map of Bay Area Transit Stops by Region—Caltrain Stations

Figure 3-29: Map of Bay Area Transit Stops by Region—North Bay Transit Stops

Chapter 4.0 An Empirical Study of the Deviation between Actual and Shortest-Travel-Time Paths

By Wenyun Tang, PhD (candidate)[5] & David M. Levinson, PhD[6]
(University of Minnesota)

4.1 Abstract

Few empirical studies of revealed route characteristics have been reported in the literature. This study challenges the widely applied shortest-path assumption by evaluating routes followed by residents of the Minneapolis–St. Paul metropolitan area, as measured by the GPS Component of the 2010 Twin Cities Travel Behavior Inventory conducted by the Metropolitan Council. It finds that most travelers used paths longer than the shortest path. This is in part a function of trip distance, trip circuity, number of turns, and age of the driver. The same traveler often used multiple routes between home and work on different days. Some reasons for these findings are conjectured.

4.2 Introduction

Few empirical studies of revealed route characteristics have been reported in the literature. Previous research by the authors [29] found fewer than 40% of commuters took the shortest paths, though 90% of subjects took routes that were within 5 minutes of the shortest paths. Other researchers have found similar results [1,15,18]. The reasons for this are several, but the simplest explanation is that people care about things in addition to and other than average travel time.

Previous research finds travelers care about monetary cost [3], avoiding stops [27], travel time reliability [2], and aesthetics [26]. Travelers might misperceive travel times [17]. They also might not want to engage in route search, and instead want to remain on habitual routes. Mismeasurement of the shortest path is also a possibility, though this has been disproved as the dominant reason.

This chapter discusses a study that tested the widely applied shortest-path assumption by evaluating routes followed by residents of the Minneapolis–St. Paul metropolitan area, as measured by the GPS Component of the 2010 Twin Cities Travel Behavior Inventory. Some of the deviation between the actual and shortest-travel-time paths is explained using a regression model, with network structure and sociodemographic factors used as independent variables.

[5] Wenyun Tang, PhD, candidate, Southeast University, 2# Sipailou, Nanjing, Jiangsu, China, Visiting Scholar: University of Minnesota, Department of Civil, Environmental, and Geo- Engineering.
[6] David Levinson. Professor and Braun/CTS Chair in Transportation, University of Minnesota. 500 Pillsbury Drive SE, Minneapolis, MN 55455. +01 (612) 625-6354. dlevinson@umn.edu.

This chapter addresses data, methodology, analysis of within travelers and between travelers differences, models, and results. This chapter then discusses possible causes for the observations that people do not choose the shortest path.

4.3 Data

Several sources of data were used for this study, including data, travel speeds, and base network data. All of these sources are discussed in greater detail in the following sections.

Travel Data

The first data source is from the Travel Behavior Inventory (TBI), conducted by the Metropolitan Council for the Twin Cities (Minneapolis–St. Paul region) in 2010 and 2011. A GPS component of the survey was used in this analysis; the GPS data were collected from a subset of individual subjects from 250 households. These households were issued GPS units to carry for a seven-day period. In addition, the same subjects also completed a travel survey for one weekday. This is detailed in the TBI report for Metropolitan Council[16], and the data are available at the Transportation Secure Data Center.

Valid GPS data were collected from 278 persons from the 250 households surveyed as part of the TBI. Trip exclusions are shown in Table 4-1. The small sample collected for this study avoided “false positives,” as the commute trips identification constraint condition was strict. However, this constraint may have excluded real work trips in which:

- There are errors in the longitude and latitude of some origins or destinations because of the accuracy of GPS devices (if one of the errors is greater than 500 meters, then it would not be considered a commute trip);
- The location of home and work differs significantly from the parking location, and a break was detected in the GPS data at the point of parking; and
- The GPS tracks started in the middle of a trip because users may have forgotten to turn the GPS on at the origin, or the GPS signal may not have been located until the trip was underway.

If these issues occurred near the origin or destination, then these trips were not identified as commute trips, which inflates the numbers of Home-to-Other (H2O), Other-to-Home (O2H), and Other-to-Other (O2O) trips.

Travel Speeds

The second data source is the TomTom road speed network data for 2010, which was acquired by the Metropolitan Council for the TBI [6]. To understand these data better, travel times for the first two data sources were compared. Since the two data sources are mapped to different networks, these sources had to be harmonized.

As shown in Figure 4-1, TomTom data are largely consistent with GPS data for the same links, though TomTom's times are a bit lower (speeds higher) on average. Potential causes for this discrepancy include differences in definition, sampling, and the treatment of traffic signals, and the possibility that some subjects made small stops that were not identified as distinct trips. A nontrivial number of TBI GPS links had travel times significantly higher than TomTom data. The TomTom methodology for averaging link travel times (and the number of observations used to construct those averages) is proprietary. As a result, the GPS data may have included short stops (engine running) or weather conditions that were not accounted for as part of this study; similar trips may have been filtered out by TomTom.

Base Network

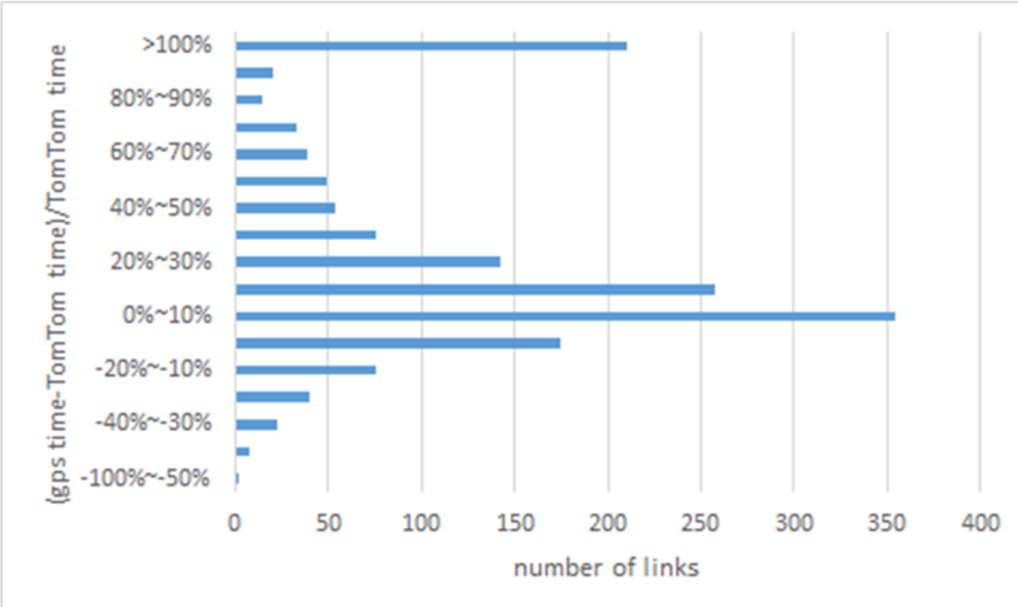
The third source of data was The Lawrence Group (TLG) base network, which includes 290,231 links. This data source has been described in previous studies and is maintained by the Metropolitan Council and TLG. It covers the seven-county metropolitan Minneapolis area and is considered by local planners to be the most accurate Geographic Information System (GIS) street map of the regional network to date. TomTom and TBI GPS data were mapped onto this network as part of this study.

Table 4-1: TBI GPS Trips Used/Unused for Analysis, by Reason of Exclusion

Steps	Number of Trips	Description
Origin Trips	16,902	• The identification was based on the time gap between two successive GPS points • If the dates of two GPS points were different and were not at midnight, then the latter point was considered to be the origin of the next trip • If the dates of two GPS points were the same, then the time was checked • If the time gap was greater than a predefined threshold (300 seconds), then these points were assigned as different trips
1	12,572	• Remove trips in which speed is always zero
2	8,461	• Remove trips where trip duration was less than 5 minutes
3	4,895	• Because in some trips the speed is “2” or “0” with no other numbers, remove the trips with average speed less than 2
H2W, Auto	142	• Use the method in the report to identify trips
H2W, Auto	124	• Destinations of two of the trips are not in the Twin Cities GIS network, so these were excluded • Some of the trips involved indirect travel from home-to-work; indirect trips were excluded from the H2W category, and were

Steps	Number of Trips	Description
		instead included in H2O

Figure 4-1: Travel Time Comparison on Links between TomTom GPS and TBI GPS Data (on average, TomTom travel times are lower than observed in the TBI)



4.4 Methodology

During the GPS data processing phase for the TBI, three steps were considered: 1) trips identification; 2) mode classification; and 3) trip purpose identification.

Trip Identification

Trips were tracked by GPS devices by finding origin and destination points for each trip. The identification was based on the time gap between two successive GPS points. If the dates of two GPS points were different and were not at midnight, then the latter point was considered as the origin of next trip. If the dates of two GPS points were the same, then the times of successive GPS points were compared. If the time gap (start of GPS data [possible trip] $n+1$ – end of GPS data [possible trip] n) was greater than a 300-second (5-minute) threshold, then they were also considered different trips; however, if the points were within 300 seconds, then they were considered part of the same trip. GPS points at the start or end of trips that showed no spatial movement, such points or trips, were removed.

Mode Classification

Mode classification is an important assessment in the use of GPS data. For this study, as set of mode identification rules were developed based on the literature [5, 10, 28] and expert assessment (these rules are shown in Table 4-2). Note that these rules are not complete in the sense that they do not guarantee that a segment will be classified into a mode. Instead, these rules aim to identify trips unambiguously by mode; ambiguous trips may have no identified mode.

Visual inspection of the individual trip records suggested that they plausibly reflect the actual modes taken; however, a fast bike and a slow car remain indistinguishable using this method. From the perspective of this study, focusing on automobile users, the most important task was to avoid "false positives" (nonauto trips showing up as auto), rather than worrying about "false negatives" (auto trips excluded from the sample set). Other studies may have different objectives with regard to modal classification, and typically identifying transit is more difficult.

Trip Purpose Identification

Trip purposes are identified based on the relative location of the GPS trip origin and destination (start and stop point) and the subject's known home and work location, as detailed in Table 4-3. In order to identify whether a trip is traveling from the origin to the destination directly without multiple purposes, the trip angles were calculated at 5 and 10 minutes after leaving and before arriving, respectively. A schematic diagram is given in Figure 4-2.

If the angles at 5 and 10 minutes are both greater than 90 degrees, a trip was considered to have other stops on the way to the destination (e.g., dropping children off at school on the way to work). Again, avoiding false positives (misidentifying nonwork trips as work trips) was more of a concern than the converse, since introducing nonwork trips into the sample were expected to create more bias than excluding random work trips.

After being divided into trips, modes, and trip purposes, auto commute trips were identified. As shown in Table 4-4, the GPS data contains 232 drive commute trips (124 H2W and 90 W2H) belonging to 58 travelers from 51 households. No W2W trips were identified, so those have been excluded. Several round trips from home without stops (H2H) trips were identified. Persons with no work address were identified as nonworkers. Trips to destinations other than the main work address were classified as nonwork (Other) trips, even if the function of the trip was for work, as that could not be determined from the GPS data.

Auto commute GPS data were then matched to TLG Twin Cities network. This method snapped all points to the nearest (by distance) link, ensuring:

- The link was a through link with no broken ends except for origin and destination links;
- The link was in the same travel direction as the GPS data; and
- There were no cycles (routes using the same link multiple times) in the network route.

The shortest-distance route was then developed on the TLG network, while the shortest-time route was computed using the TomTom network. TomTom speed data includes seven periods in a 24-hour day:

- Early Morning (AM1).
- Late Morning (AM2).
- Midday (MD).
- Early Afternoon (PM1).
- Late Afternoon (PM2).
- Evening (EV).
- Night (NT).

Link travel speed was selected based on the trip period's start time in GPS data. This was then compared to the total distance between the actual route and the shortest-distance route as well as the shortest-time route. The total travel time between the actual route and shortest-time route was also compared. The total overlap distance was calculated using the actual route, the TomTom shortest-travel-time-path route, and the shortest-distance route. An example of one subject's shortest-travel-time path (using TomTom data), shortest-distance path, and actual path is shown in Figure 4-3.

Besides finding the general difference between the actual path and the shortest-time route, the trip circuity and the number of turns were also compared. (The circuity is the ratio of network to Euclidean, or straight-line, distance.) Through this process, the relationship between time difference and circuity, and the difference between time and the number of turns on the actual route, is found.

Table 4-2: Trip Classification Rules

Walk
• Maximum speed of all points ≤ 20*km/h*
• Duration > 60*s*
• Percentile of speed of all points ≤ 10*km/h*
• Average speed of all points ≤ 6*km/h*
Rail
• Distance from first point of speed accelerates to 10*km/h* to the nearest rail station <150*m*
• Distance from last point that speed is greater than 10*km/h* to the nearest rail station <150*m*
• Average speed of all points >10*km/h*
Bus
• Distance from first point of speed accelerates to 10*km/h* to the nearest bus stop <50*m*
• Distance from last point that speed is greater than 10*km/h* to the nearest bus stop <50*m*
• Average speed of all points >10*km/h*
Bicycle
• 85th percentile of speed of all points ≥10*km/h* and <20*km/h*

- Max speed of all points ≤30*km/h*

Car
- The remaining trip segments with average speed of all points >10*km/h*

Table 4-3: Definitions of Trips Based on Relative Location of Trip Origin and Known Home and Work Locations

	Origin				
Destination	Worker			Nonworker	
worker	*H* ≤ 500*m*	*W* ≤ 500*m*	*H*+*W* > 500*m*	*H* ≤ 500*m*	*H* > 500*m*
H ≤ 500*m*	H2H	W2H	O2H	-	-
W ≤ 500*m*	H2W	W2W	O2W	-	-
H+*W* > 500*m*	H2O	W2O	O2O	-	-
nonworker					
H ≤ 500*m*	-	-	-	H2H	O2H
H > 500*m*	-	-	-	H2O	O2O

Note: Location tested for proximity to Home, Work and Other in sequence

Destination location identified after Origin

Where: H2W = Home-to-Work, H2O = Home-to-Other, and so on

Table 4-4: Number of Trips, by Travel Mode and Trip Purpose

	H2W	H2O	O2H	W2H	W2O	O2W	O2O	H2H	Total	Percentage
Walk	1	24	3	0	0	17	67	26	138	2.82
Train	0	0	0	0	0	0	1	0	1	0.02
Bus	8	26	104	14	12	14	110	0	288	5.88
Bike	0	13	8	2	0	4	36	0	63	1.29
Drive	124	969	982	90	68	85	1073	10	3419	69.85
Not identified	43	260	313	12	15	53	308	0	986	20.14
Total	**176**	**1292**	**1410**	**118**	**95**	**173**	**1595**	**36**	**4895**	**100.00**
Percentage	**3.60**	**26.39**	**28.80**	**2.41**	**1.94**	**3.53**	**32.58**	**0.74**	**100**	

Where: H2W = Home-to-Work, H2O = Home-to-Other, and so on

Figure 4-2 displays the calculation of trip angles at 5 and 10 minutes after leaving and before arriving. Trips where the direction of travel was in the opposite direction from the origin were assumed to have side activities.

Figure 4-2: Measuring Trip Angles

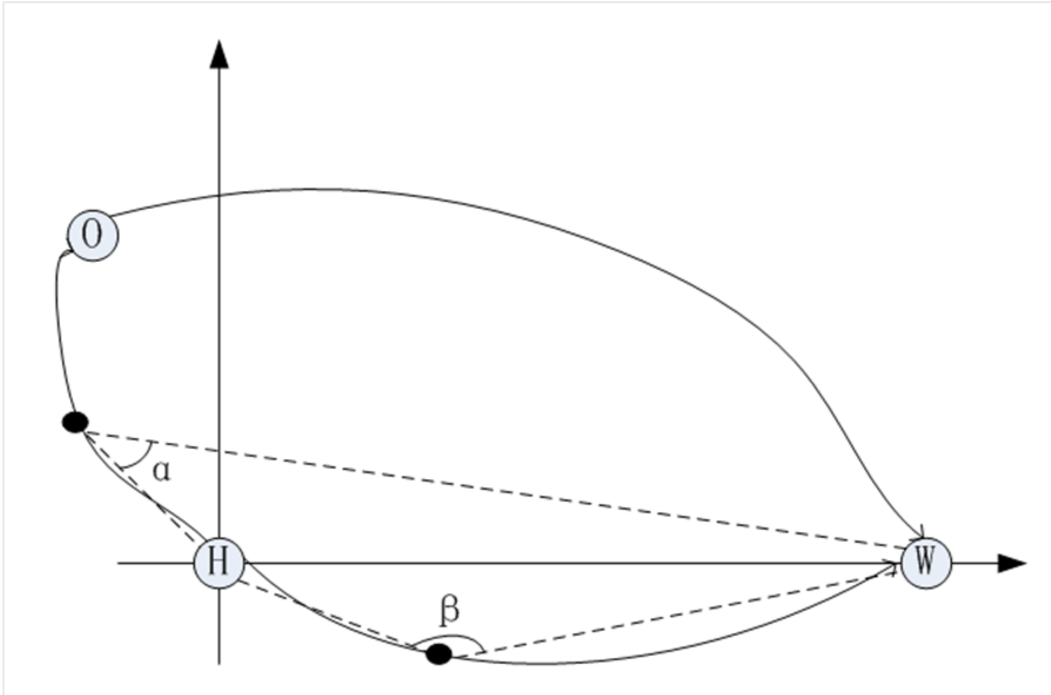
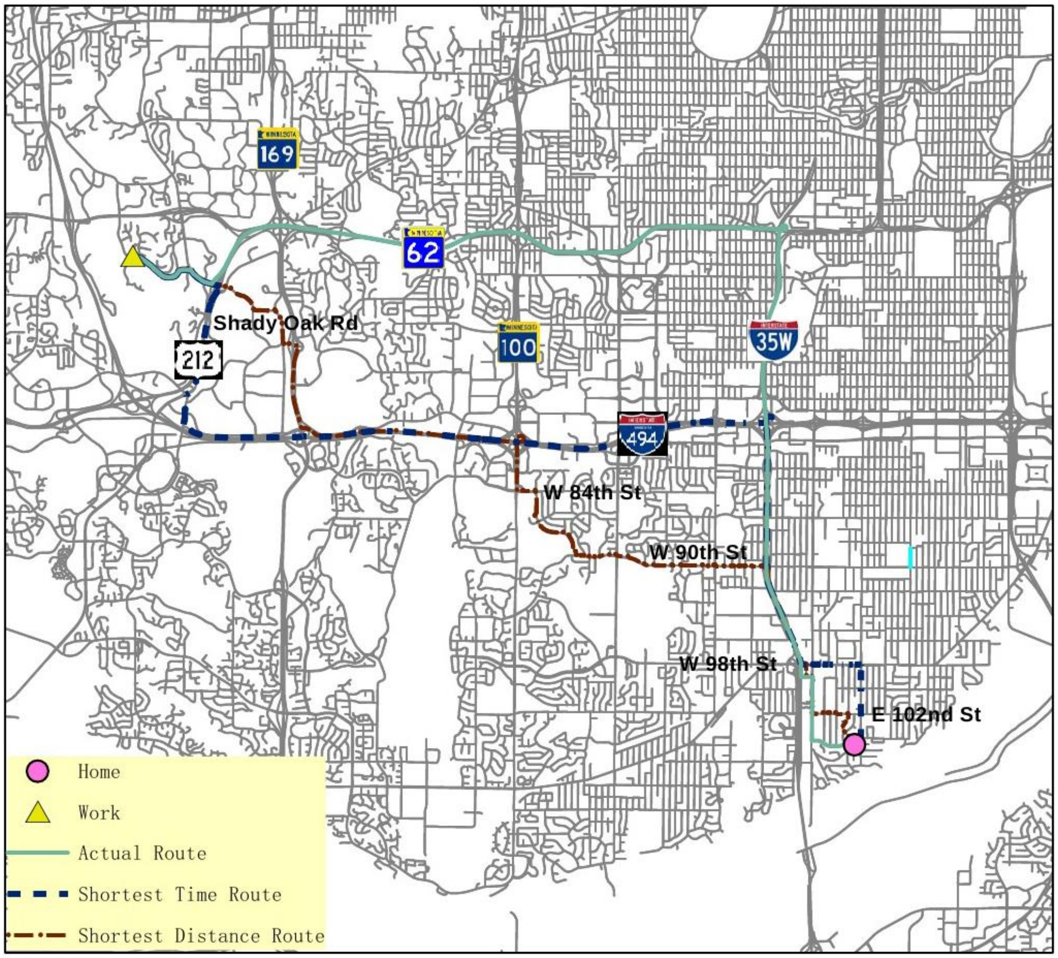


Figure 4-3: Example of Shortest Distance, Shortest Travel Time, and Actual Routes on the TLG GIS Network



4.5 Descriptive Statistics

Figure 4-4 shows that actual trip length and duration are correlated with difference between TomTom GPS-based shortest travel time (t_{sp}) and actual travel time from the TBI (t_{GPS}). As the difference increases, the length and duration both increase. Both the trend of average and standard deviation of length and duration are similar. Examined from another perspective, while the trip duration gets longer, the difference is greater (Figure 4-5). As the percentage difference between the two datasets increases, the length and duration both increase (Figure 4-4). As trip duration increases, the difference between the two datasets is also larger in percentage terms.

Figure 4-4: Summary Information for Each Difference Interval $(t_{GPS}-t_{sp})/t_{sp}$

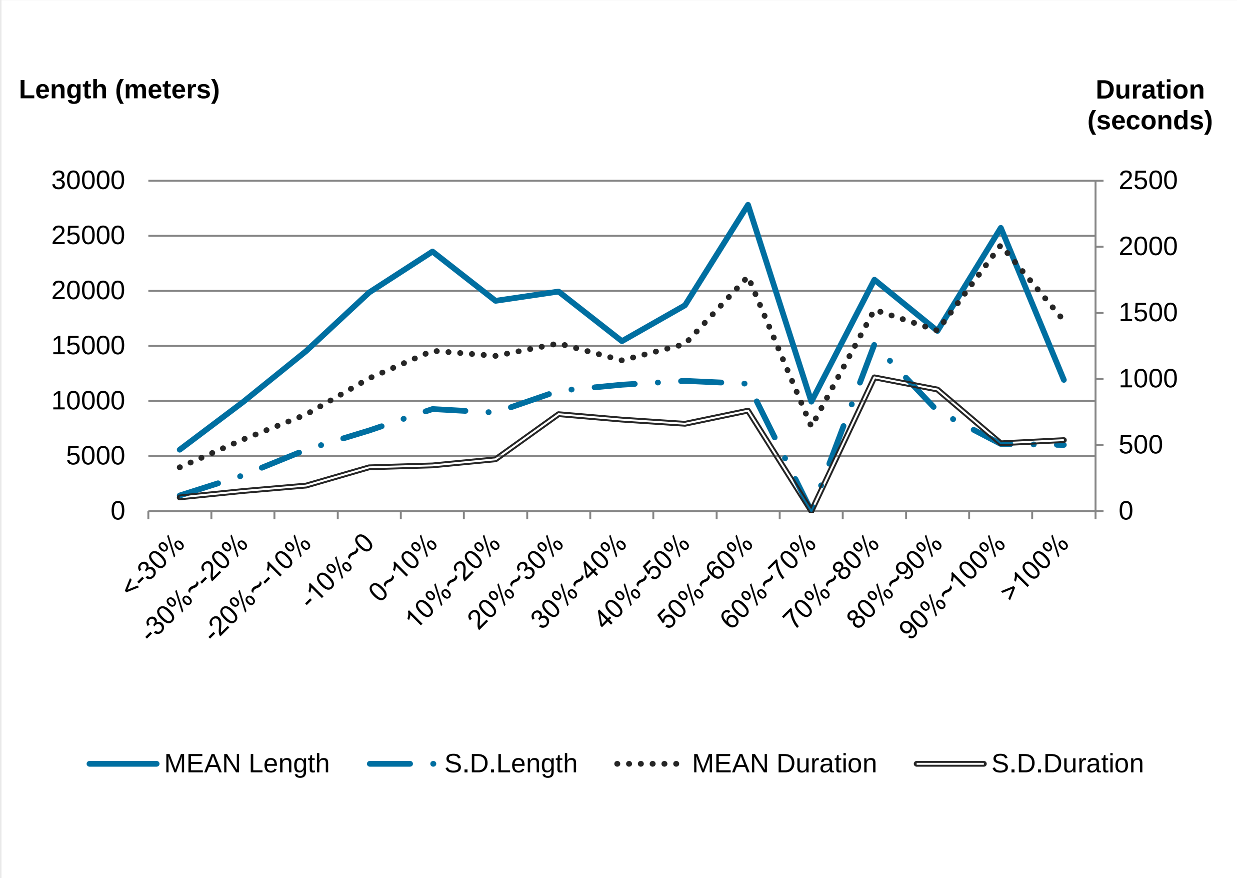
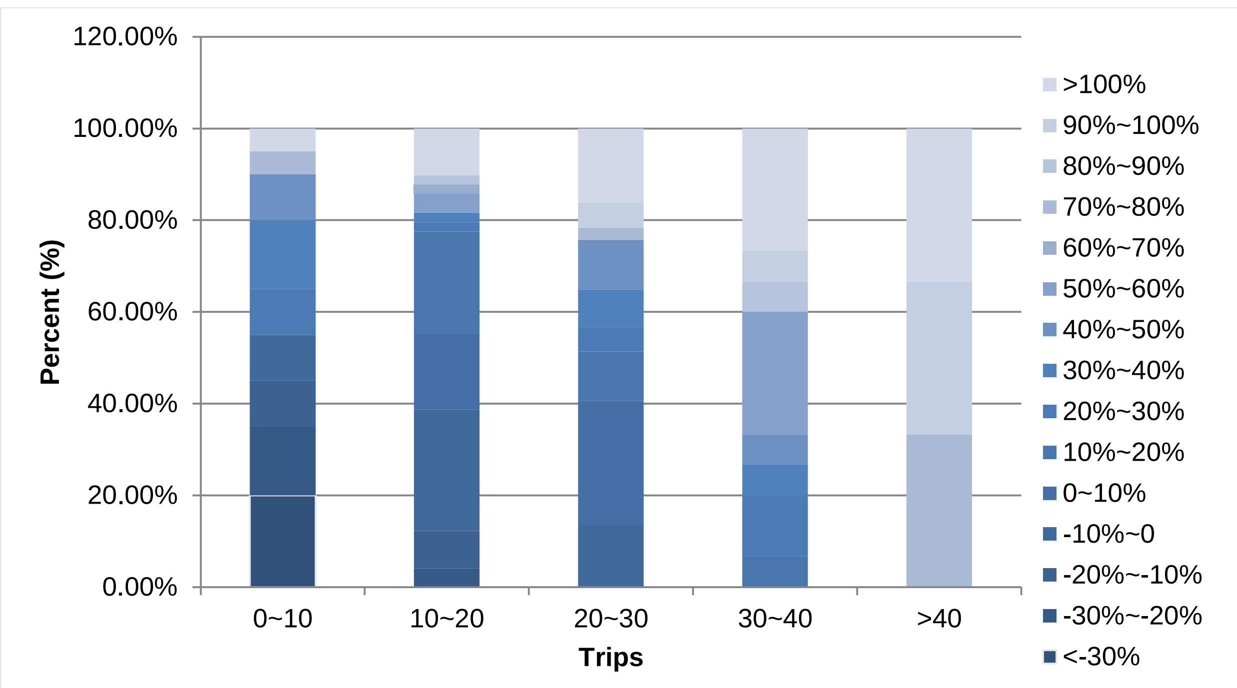


Figure 4-5: Percent of Trips in Each Difference Interval for Each Trip Duration Interval (min.)



4.6 Descriptive Results (Within Travelers)

One of the theoretical advantages of multiday GPS data for the same travelers is the ability to compare results for individual travelers between days. While the sample for this study is relatively small, the results for this analysis are presented as follows.

Thirty-five travelers in the sample had multiple home-to-work trips, averaging 3.11 trips for each traveler, and resulting in 109 trips. Among these multiple-trip travelers:

- Each day, 26 of 35 travelers (comprising 83 trips, or 74%) took the same route. These routes are shown in Figure 4-6. Of these:
  - The shortest-travel-time paths—called Same-Route-Shortest-Path (SRSP) travelers—were taken by 7 of the 26 travelers (comprising 25 trips, or 27%).
  - The shortest path was not used by 19 of the 26 travelers (or 73%). These are Same-Route NOT Shortest-Path (SRNSP) travelers.
- The same route was not taken each day by 9 of the 35 travelers (comprising 26 trips, or 24%). These are Not-Same-Route (NSR) travelers and they are shown in Figure 4-7.

Equation 3 describes the overall difference (Delta) between routes taken by one traveler day by day.

Equation 3: Overall Difference (Delta) between Routes Taken by One Traveler Day by Day

$$\Delta = \frac{1}{N}\sum_{N}\left(1 - \frac{D_{overlap}}{D_{shorter}}\right)$$

where:

D_overlap is distance overlap between two actual routes

D_shorter is a distance of the shorter route between the origin and destination

N is the number of routes for a particular traveler

A higher Delta means more overlap. A delta of zero means no overlap (the routes are almost completely different).

As seen in Figure 4-7, the difference between everyday-actual route and shortest-travel-time route fluctuates. Table 4-5 shows that the difference among actual routes fluctuates. There are many possible reasons for any given fluctuation, including a continuing search for the shortest path, unobserved (in the TomTom data) day-to-day congestion fluctuations, unobserved side trips, or a quest for variety.

Figure 4-6: Percentage of Overlap between the Actual Route and Shortest-Path, Same-Route (SR) Travelers

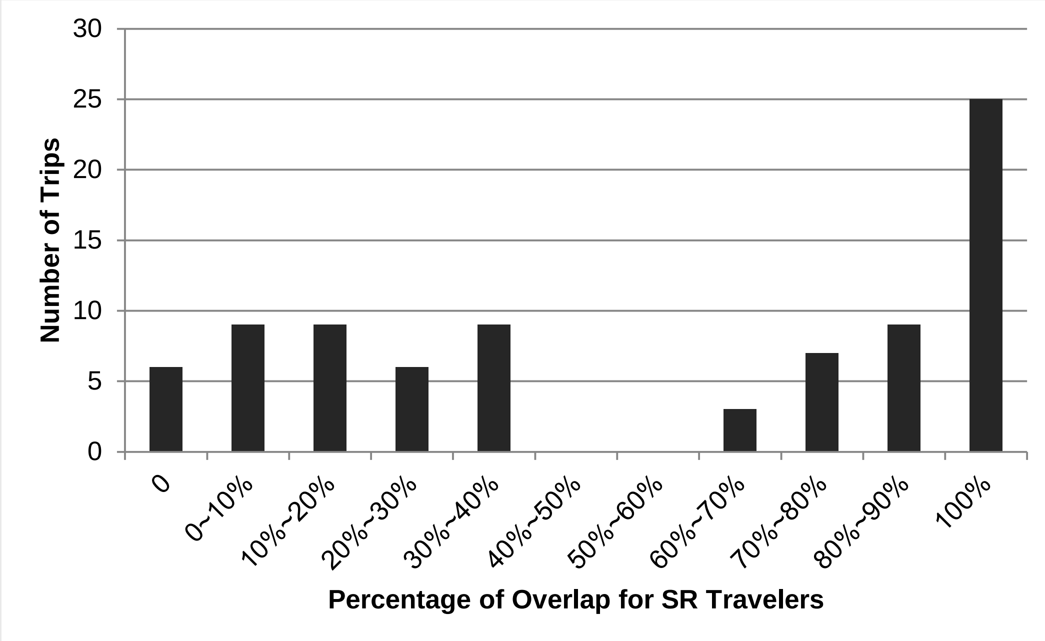


Figure 4-7: Percentage of Overlap between the Actual Route and Shortest-Path, NSR Travelers

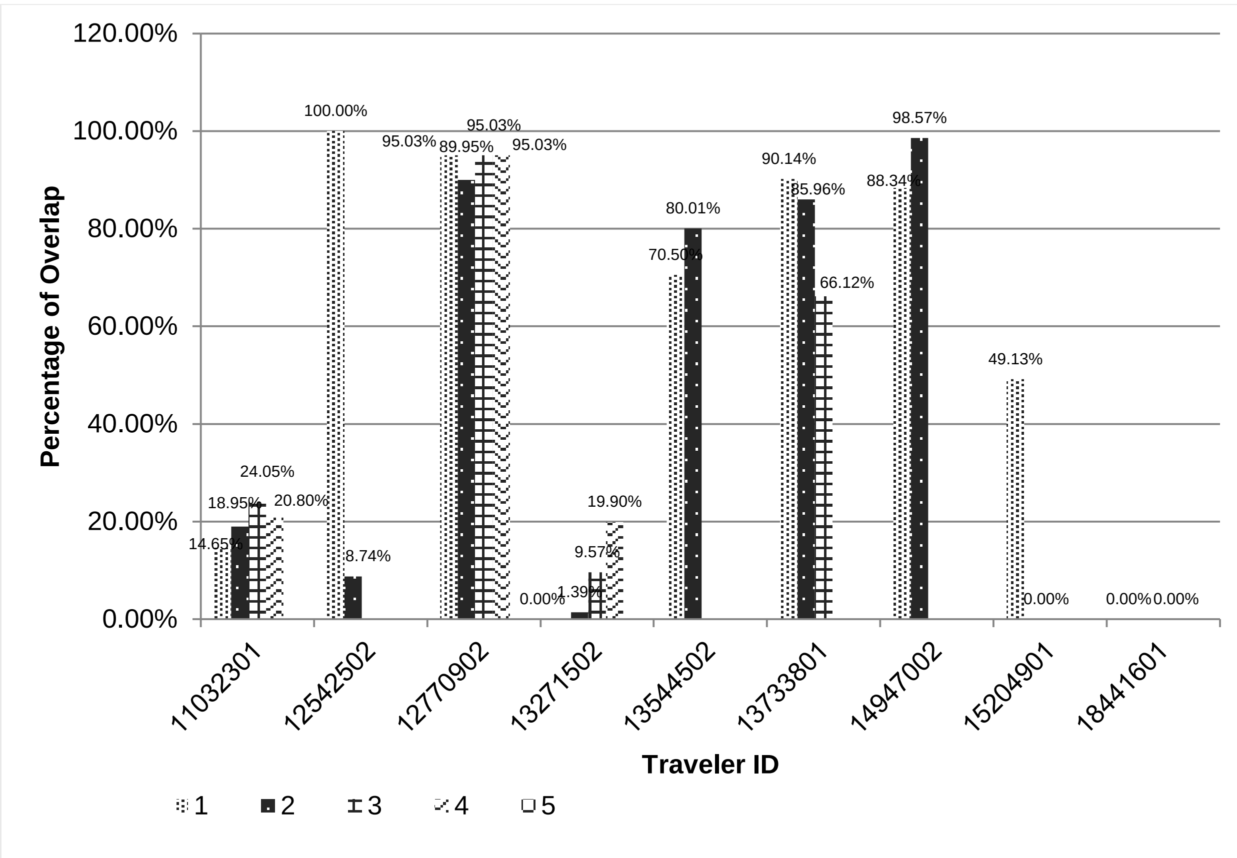


Table 4-5: Percentage of Overlap for Each Pair of Trips, Overall Difference (Delta) and Standard Deviation of NSR Travelers

Person ID	Trips ID	Trips ID					Delta	SD
11032301		3	8	18	23		0.12	0.339
	3	-	-	-	-			
	8	0.16	-	-	-			
	18	0.18	0.11	-	-			
	23	0.17	0.04	0.07	-			
12542502		14	27				0.91	0.645
	14	-	-					
	27	0.91	-					
12770902		21	26	47	66	85	0.38	0.419
	21	-	-	-	-	-		
	26	0.11	-	-	-	-		
	47	0	0.11	-	-	-		
	66	0	0.11	0	-	-		
	85	0.83	0.96	0.83	0.83	-		
13271502		2	11	34	48		0.76	0.091
	2	-	-	-	-			
	11	0.69	-	-	-			
	34	0.86	0.83	-	-			
	48	0.53	0.80	0.85	-			
13544502		17	25				0.09	0.067
	17	-	-					
	25	0.09	-					
13733801		1	9	86			0.23	0.128
	1	-	-	-				
	9	0.29	-	-				
	86	0.09	0.30	-				
14947002		2	6				0.11	0.072
	2	-	-					
	6	0.11	-					
15204901		28	107				1	0.347
	28	-	-					
	107	1	-					
18441601		13	42				0.25	0
	13	-	-					
	42	0.25	-					

Note: Standard deviation of percentage of overlap between actual route and shortest-travel-time

4.7 Descriptive Results (Between Travelers)

Figure 4-8 compares (in percentage terms) actual GPS travel times with estimated TomTom times on the shortest path. As can be seen, almost all trips had travel times longer than the TomTom shortest path (a few are shorter because the TomTom network does not have speeds on some local roads). This is in part because end-of-trip details (e.g., parking) are not a part of the TomTom network. More than half the trips were longer (30% or more) than the estimated shortest-travel-time path.

Figure 4-9 displays the absolute difference in minutes. More than half of all auto commute trips in the sample are more than 5 minutes longer than the shortest path. The highest travel time differences occur for trips with low overlap. For trips with a high overlap rate, the time differences are not as large, but are still far from zero. However, when compared to the shortest-distance route (Figure 4-10), the percentage of overlap between the actual route and the shortest-time route (Figure 4-11) is higher. However, only about one quarter (35 out of 124) chose a route that has a high degree of overlap (>90%) with the shortest-travel-time route.

Figure 4-12 and Figure 4-13 demonstrate that when the circuity of the actual route increases, the difference in times decreases; as the number of turns rises, the difference shrinks.

The average difference for males is nominally higher than females (male:273*s*; female:254*s*), and the standard deviation for males is also higher than for females (male:402*s*; female:382*s*). These differences do not appear to be significant.

Figure 4-8: Travel Time Comparison (percentages) between TBI GPS Time (t_{GPS}) on Actual Routes and TomTom GPS Time (t_{sp}) on Shortest-Time Route

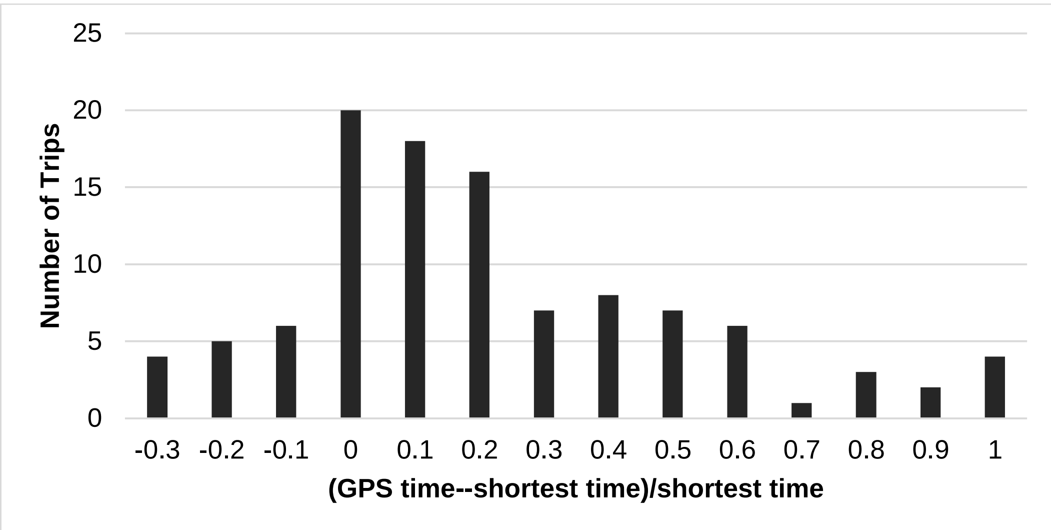


Figure 4-9: Travel Time Difference (minutes) between TBI GPS Time (t_{GPS}) on Actual Route and TomTom GPS Time (t_{sp}) on Shortest-Time Route

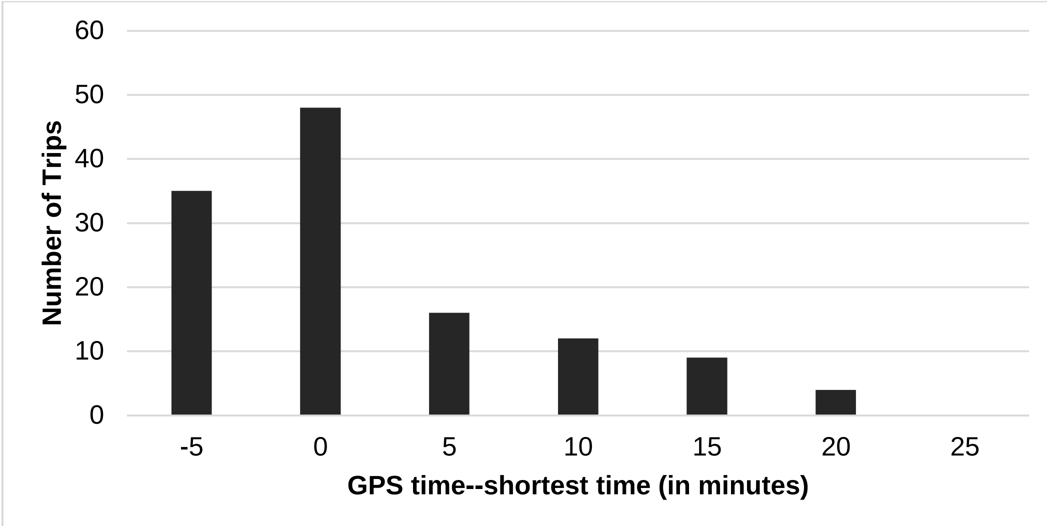


Figure 4-10: Percentage of Overlap—Difference between Actual Route and Shortest-Distance Route

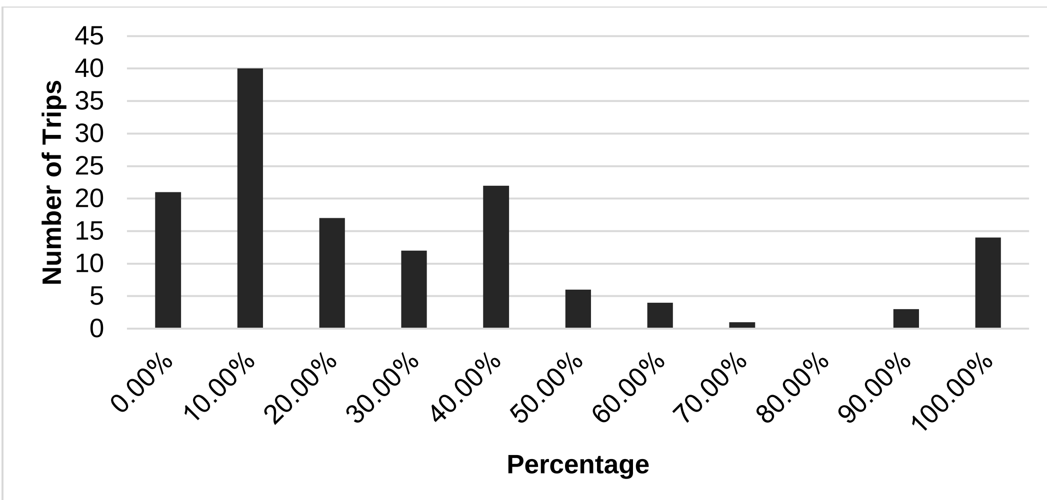


Figure 4-11: Percentage of Overlap—Difference between Actual Route and Shortest-Travel-Time Route

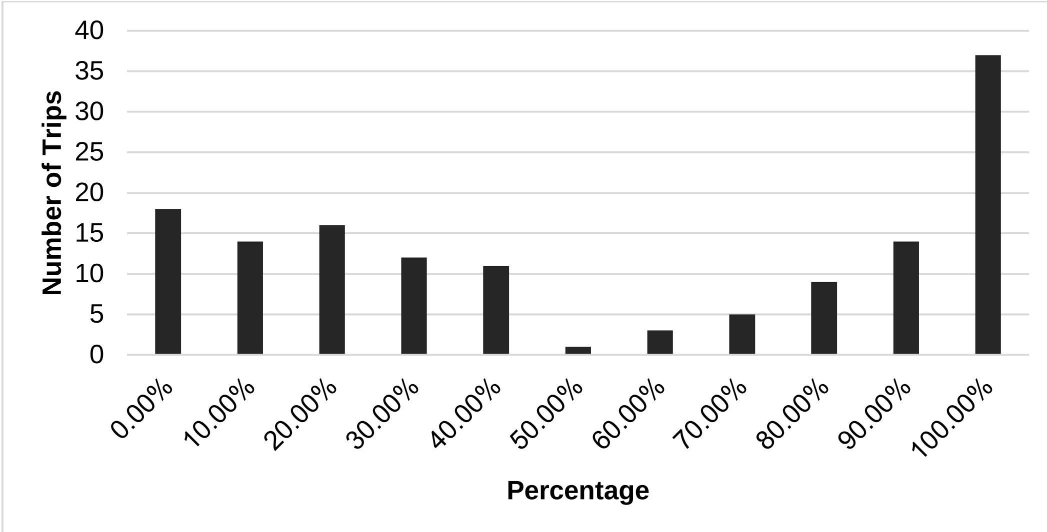


Figure 4-12: The Relationship between Time Difference and Circuity ($D_{Network}/D_{Euclidean}$) of Actual Route

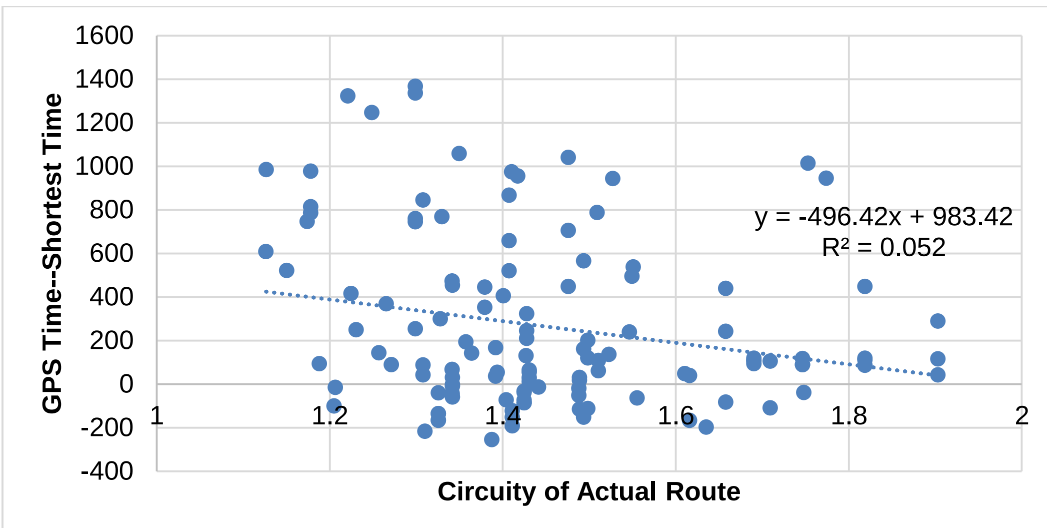
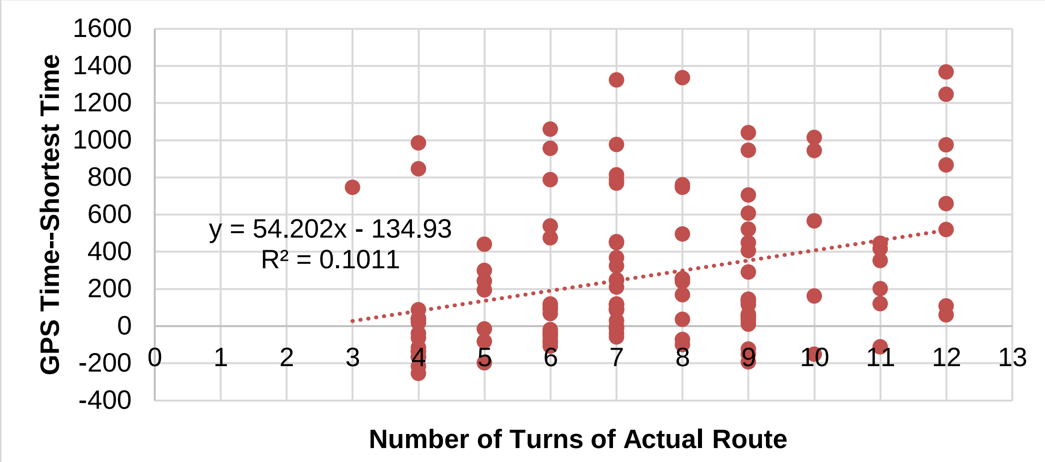


Figure 4-13: The Relationship between Time Difference and Number of Turns on the Actual Route



4.8 Models

As part of this study, models were built to reveal the relationship between the ratio of observed GPS time to the estimated shortest path time from TomTom data and circuity to the number of turns for actual route as well as sociodemographic characteristics. As a first step, a simple Ordinary Least Squares (OLS) linear regression was constructed for each attribute.

Equation 4: OLS Linear Regression

$$Y = aX + b$$

These models then revealed how attributes affect the time difference between GPS time (t_{GPS}) on the actual route and TomTom time on the shortest-time route (t_{sp}). Figure 4-8 through Figure 4-13 displays these results. A multivariate linear regression model was then built.

Generally, the ratio of observed to shortest-path travel time (τ) is expected to be a function of network characteristics (N) and individual characteristics (S).

Equation 5: Ratio of Observed to Shortest-Path Travel Time, Specified as General Function of Network Structure and Individual Characteristics

$$\tau = f(N, S)$$

The variables of interest can then be further specified:

Equation 6: Ratio of Observed to Shortest-Path Travel Time, Dependent on Network Distance, Circuity, Number of Turns, and the Age and Gender of the Traveler

$$\tau = f(D_{sp}, C_{GPS}, C_{sp}, T_{GPS}, A_i, G_i)$$

where:

$\tau = t_{GPS}/t_{sp}$

C = Circuity = $D_{Network}/D_{Euclidean}$

T = Number of turns

A_i = Age of the individual traveler

G_i = Gender of the individual traveler

Undoubtedly, other variables may play in this, including the location of traffic signals or crash data, but they were not available or tested in this analysis. These results are presented in the next section.

4.9 Regression Model Results (Between Travelers)

The final regression results are shown in Table 4-6. The dependent variable is the ratio of the time (τ) on the chosen path (t_{GPS}) from the TBI and the estimate of the travel time on the shortest path (t_{sp}) from TomTom data.

This ratio decreases with distance. Longer trips are more likely than short trips to follow the shortest path. Longer trips may also need to be more efficient because of their length (and associated time budgets).

All else being equal, it is to be expected that more circuity would add to total travel time on the shortest path. However, longer-distance trips are less circuitous, both because the network structure enables movement to routes higher in the hierarchy of roads, which are traversed for long distances, and because short trips often have to travel on curvilinear subdivision streets (especially in the postwar suburbs) [9]. In addition, people select for less circuitous routes when choosing where to live relative to their work location (or vice versa) [12]. In this study, more circuitous routes were found to be associated with lower travel time ratios. This is a complex result without an easy intuitive explanation.

In regard to ratios, the proportion by which the actual travel time exceeds the shortest-path time drops as the circuity increases. While this is not at all surprising for the denominator (shortest-path time), it is more surprising for the numerator (actual [GPS] time). It is possible that individuals who live in places that have more circuitous networks have fewer choices in routes, and thus fewer opportunities to reasonably deviate from the shortest route (thus lowering the ratio). Clearly, as with all results, additional research may provide additional answers.

Routes with more turns (T_{GPS}) have a significantly higher time ratio. The variable for turns on the shortest-path route (T_{sp}) was statistically insignificant, and dropped from the final regression. Other model variables (including variables for both network and household structure) were tested in various combinations and were not statistically significant.

Overall, the model explains approximately 15% of the variance in the time ratio. Though the variables are statistically significant, the small size of this sample, combined with the fact that it is for only one metropolitan area, necessitate additional research before strong conclusions can be drawn about the explanatory—much less causal—factors explaining why people do not choose the shortest path. The Discussion section offers some conjectures.

Table 4-6: Explaining τ, the Ratio of GPS Travel Time to Shortest-Path Travel Time

Independent Variables	Coef.	Std. Err.	t	P>|t|
Distancesp	-.0000185	6.67e-06	-2.78	0.006
CircuityGPS	-.6569722	.3180107	-2.07	0.041
Circuitysp	-.8381146	.4148644	-2.02	0.046
TurnsGPS	.0597149	.0232824	2.56	0.012
Agei	-.0096658	.0049401	-1.96	0.053
Constant	3.684621	.621362	5.93	0.000
Adjusted R2				0.1457
Sample Size (N)				124

4.10 Discussion

Using empirical data from a GPS-based study of approximately 250 households, and focusing on auto commuters in that dataset, this study tests a crucial assumption in transportation planning practice, embedded in the principle of user equilibrium due to Wardrop [23], that travelers choose the shortest-travel-time path. The research has found that most travelers do not choose the shortest-travel-time-path, and the overlap between their actual path and the analyst's best estimate of the shortest path is well below 100%. The following represents an attempt to understand why people are not taking the shortest path:

- **Selflessness**. Wardrop's principle [23] assumes that people are selfish, but perhaps they are selfless. It is assumed that individuals aim to minimize their own travel time rather than the travel time of society as a whole. However, people cannot know what decision will minimize society's travel time because of computational and informational issues, as discussed later. However, if individuals had that information, then perhaps they might selflessly choose a different route. In the absence of such information, individuals are (at best) only guessing whether what they are doing is in the best interest of society as a whole, even if their choice involves some self-sacrifice. (This assumes that individuals are still making their trips. In the case of roadway congestion, it would be better for everyone else—from a travel time perspective—to avoid the trip altogether.)
- **Rationality**. Wardrop's principle assumes that people are rational, but maybe people are not rational, or at least not rational all the time. This is true in the

sense that people react emotionally and intuitively, employing what Nobel Prize winner Daniel Kahneman [11] calls System 1 in *Thinking Fast and Slow*, based on heuristic rules. Individuals do not have time for rational assessment. In another sense, for a repeated decision (like commuting back and forth to work), it costs a significant amount of travel time—a scarce resource—to systematically behave irrationally. Therefore, it is assumed that people are behaving rationally (engaging Kahneman's System 2) when they can. The idea of bounded rationality, developed by Herbert Simon [20], has been applied to route choice problems by many researchers, see e.g. [7,8,14]. Models with bounded rationality can be built, with such models assuming or estimating the bounds to this rationality due to information, cognitive limits, and time available to make a decision.

- **Perception**. Individuals may select the shortest travel time on their route, but they may misperceive the travel time on the network due to perception or cognition limits. On a 24-minute trip, few travelers will know the travel time to the nearest 30 seconds or minute. In surveys, reported travel times are typically rounded to 5 minutes; on occasion, surveys round to the nearest 15 minutes. As a result, if people are only perceiving time in 5- or 15-minute chunks, saving 1 or 2 minutes will not be perceived as important [17].
- **Computation**. Travelers cannot accurately add travel times across different road segments, and they cannot systematically compare the travel times over alternative routes even if they had a complete dataset due to computational constraints.
- **Information**. People do not remember or store information related to complete maps of the roadway network. People often possess good mental maps of the local street network near their home, workplace, and frequently visited locations, but if they live far from where they work, then they tend not to know the detailed roadway network in-between. There are limits to people's ability to navigate. The cognitive or mental maps of most individuals are far from complete. Most people only have the experience of the routes they have actually used. Individuals can test other routes to gain experience/knowledge, but individuals (unlike GPS systems) do not innately possess this information.
- **Valuation**. Perhaps people are minimizing the weighted sum of travel time, where time spent in different conditions is valued differently. It is known from the transit literature, for instances, that time spent waiting for a bus is much more onerous than time spent aboard a vehicle in motion making progress toward its destination; this is effect is pronounced if the arrival time of the bus is uncertain [24,27].
- **Objective**. Wardrop's model assumes that people care only about minimizing travel time. People may be rational, but they may prioritize other trip characteristics instead of or in addition to travel time. There is evidence from other transportation choices that people, in general, are not minimizing travel

time [21]. When a person chooses a place to live, that individual is not choosing to minimize her commute time to work. In fact, some studies have considered a hypothetical relocation of everybody's place of residence, whereby each person's house was equivalent to the structure in which they currently reside, but was "moved" to be as close to their workplace as possible (given everyone else was similarly moved). As a result, average commute times fell from approximately 24 minutes to 8 or 10 minutes. There is a significant amount of "excess travel" from a strict travel time-minimizing perspective [21]. There are many reasons for excess travel, but the most obvious reason is that it is not "excess" from the point of view from people traveling. People are making home-location decisions for a variety of reasons: the journey to work is not the only consideration. (However, travel time must be a consideration for some individuals, otherwise cities would not exist.) It is possible that people, when choosing a home location, might underestimate the amount of time that will be spent traveling, and thus underestimate the frustration associated with long commutes, and are thus unhappier than expected [22]. A major source of time estimation error arises because most people search for homes on the weekend, but tend to commute on weekdays.

- **Search cost**. How long does it take to figure out the travel time on alternative routes? Is a traveler willing to spend 10 minutes exploring the network in order to save 30 seconds of travel time every day for the rest of her career? From a purely rational perspective, such a search may be worthwhile because the payback period is only 20 days. However, people often discount the possibility of saving time, worrying that a shortcut will be longer; individuals may also fear getting lost. Fear of the unfamiliar is a major deterrent to exploration [25].
- **Route quality**. Many factors describe the quality or condition of a route and its environment. Is it potholed or newly paved? Does it run through a pleasant or unpleasant neighborhood? Evidence from previous studies shows that some people prefer a longer route if it is an attractive boulevard or parkway rather than a drive through a freeway trench [26]. This study was unable to assess the aesthetics of alternative routes.
- **Reliability**. The likelihood of arriving on time, and not just the expected travel time, affects willingness to select a route. There is the old parable of the man who drowned in an average of one inch of water. Similarly, it might not matter to a traveler that the average travel time is 20 minutes if one day a week (but never knowing in advance which day) that traveler can expect a travel time of 60 minutes. Travelers do not want to leave 40 minutes earlier to avoid the occasional bad outcome, and they may be willing to take a slower but more reliable route. Travelers may even have a mixed strategy, or portfolio, combining different routes in order to achieve a personally satisfactory tradeoff between expected time and reliability [13]. In practice, this means that some people might take surface streets (which are generally

slower, but more reliable) instead of freeways (which are faster, but subject to more catastrophic breakdowns of traffic flow) [2]. In principle, with multiday GPS data, this question can be assessed more deeply. Unfortunately, the one-week timeframe and small sample size do not permit drawing conclusions about reliability from this dataset.

- **Pleasure of travel**. Maybe people are rational, but perhaps they prefer traveling to being at work or home, and so choose longer routes to prolong the experience. Many people want to commute; Redmond and Mokhtarian [19] find a positive value to some amount of commuting, and that the preferred commute length is not typically zero. However, it appears that many commutes are longer than the desired amount. For some people, the longer route, which provides some psychological buffer between the stresses of work and the stresses of home, is desired.

4.11 Conclusions and Future Research

This study analyzed multiday GPS data from the Minneapolis-St. Paul region. In general, it found that travelers do not take the shortest path, and many travelers use different routes between the same home and work location each day. Several reasons for these findings were explored, and multiple factors—such as network structure—affect people's choice of routes, perhaps through their perception of time, or through the cognitive burden (mental transaction costs) associated with route complexity.

This study also revealed many follow-on topics that could be the subject of a deeper investigation than could be provided with this study and its limited sample. Further investigation into whether sociodemographic and economic factors explain route choice is warranted. Intrahousehold models of travel demand may explain route choices, particularly to the extent that there are embedded pick-up and drop-off passenger trips that do not leave an obvious GPS signal (e.g., the engine is running, and the passenger exits the vehicle quickly). Such an analysis may require further validation against trip-diary data, or new survey techniques using better observational methods, in order to better understand people's activities. This research focused on work trips because of their regularity; however, most trips are not work trips, and further research should examine the factors affecting route choice for nonwork trips. More examination of preferences for route types is warranted, and this is easier to discover in more controlled experiments [3, 26]. In addition, the use of radio, Internet, in-vehicle GPS navigation systems, or other forms of traveler information is likely to continue to shape traveler route choices in coming years. Finally, the advent of automated vehicles may change this problem radically from one of traveler choice to one of understanding the logic of the embedded algorithms of the in-vehicle navigation systems.

In short, given that the shortest-path assumption explains so little of actual route choice, it is the authors' conclusion that applied route choice models should be reformulated with more realistic behavioral assumptions. The academic literature has discussed this problem for decades, but the magnitude of the problem has not been quantified until recently with the advent of GPS data. New multiday GPS data enable new approaches to this problem.

4.12 Acknowledgments

The Federal Highway Administration of the US Department of Transportation is acknowledged for funding the work under a grant to RSG. All analysis and errors are the responsibility of the authors.

The raw data used for the research are accessible through the Transportation Secure Data Center (TSDC) of the National Renewable Energy Laboratory http://www.nrel.gov/vehiclesandfuels/secure_transportation_data.html.

4.13 References

[1] S. Bekhor, M.E. Ben-Akiva, and M.S. Ramming. Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research,*, 144(1):235–247, 2006.

[2] Carlos Carrion and David Levinson. Value of travel time reliability: A review of current evidence. *Transportation research part A: policy and practice*, 46(4):720–741, 2012.

[3] Carlos Carrion and David Levinson. Valuation of travel time reliability from a GPS-based experimental design. *Transportation Research Part C: Emerging Technologies*, 35: 305–323, 2013.

[4] Carlos Carrion and David M Levinson. A model of bridge choice across the Mississippi River in Minneapolis. In *Network Reliability in Practice*, pages 115–129. Springer, 2012.

[5] Cynthia Chen, Hongmian Gong, Catherine Lawson, and Evan Bialostozky. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. *Transportation Research Part A: Policy and Practice*, 44(10):830–840, 2010.

[6] Nick Cohn. Real-time traffic information and navigation. *Transportation Research Record: Journal of the Transportation Research Board*, 2129(1):129–135, 2009.

[7] Xuan Di, Henry X Liu, Jong-Shi Pang, and Xuegang Jeff Ban. Boundedly rational user equilibria (BRUE): Mathematical formulation and solution sets. *Transportation Research Part B,*, (57):300—313, 2013.

[8] Xuan Di, Henry X Liu, and Xuegang Jeff Ban. Second best toll pricing within the framework of bounded rationality. 2014.

[9] David J Giacomin and David M Levinson. Road network circuity in metropolitan areas. In *Transportation Research Board 93rd Annual Meeting*, number 14-0955, 2014.

[10] Hongmian Gong, Cynthia Chen, Evan Bialostozky, and Catherine T Lawson. A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems*, 36(2):131–139, 2012.

[11] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[12] David Levinson and Ahmed El-Geneidy. The minimum circuity frontier and the journey to work. *Regional science and urban economics*, 39(6):732–738, 2009.

[13] David Levinson and Shanjiang Zhu. A portfolio theory of route choice. *Transportation Research Part C: Emerging Technologies*, 35:232–243, 2013.

[14] H.S. Mahmassani and G.L. Chang. On boundedly rational user equilibrium in trans portation systems. *Transportation Science,*, 21(2):89–99, 1987.

[15] T. Morikawa, T. Miwa, S. Kurauchi, T. Yamamoto, and K. Kobayashi. Driver's route choice behavior and its implications on network simulation and traffic assignment. *Simulation Approaches in Transportation Analysis*, pages 341–369, 2005.

[16] Metropolitan Council of the Twin Cities. Travel behavior inventory surveys.

[17] Pavithra Parthasarathi, David Levinson, and Hartwig Hochmair. Network structure and travel time perception. *PloS One*, 8(10):e77718, 2013.

[18] Carlo Giacomo Prato and Shlomo Bekhor. Applying branch-and-bound technique to route choice set generation. *Transportation Research Record,*, 1985(1):19–28, 2006.

[19] Lothlorien S Redmond and Patricia L Mokhtarian. The positive utility of the commute: modeling ideal commute time and relative desired commute amount. *Transportation*, 28(2):179–205, 2001.

[20] H.A. Simon. A behavioral model of rational choice. *The Quarterly Journal of Economics,*, 69(1):99–118, 1955. ISSN 0033-5533.

[21] Kenneth A Small and Shunfeng Song. 'Wasteful' commuting: a resolution. *Journal of Political Economy*, pages 888–898, 1992.

[22] Alois Stutzer and Bruno S Frey. Stress that doesn't pay: The commuting paradox*. *The Scandinavian Journal of Economics*, 110(2):339–366, 2008.

[23] J.G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers,*, 1:325–378, 1952.

[24] Xinkai Wu, David M Levinson, and Henry X Liu. Perception of waiting time at signalized intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 2135(1):52–59, 2009.

[25] Lei Zhang. Behavioral foundation of route choice and traffic assignment. *Transportation Research Record: Journal of the Transportation Research Board*, 2254(1):1–10, 2011.

[26] Lei Zhang and David Levinson. Determinants of route choice and value of traveler information: a field experiment. *Transportation Research Record: Journal of the Transportation Research Board*, 2086(1):81–92, 2008.

[27] Lei Zhang, Feng Xie, and David Levinson. Illusion of motion. *Transportation Research Record: Journal of the Transportation Research Board*, 2135(1):34–42, 2009.

[28] Yu Zheng, Like Liu, Longhao Wang, and Xing Xie. Learning transportation mode from raw GPS data for geographic applications on the web. In *Proceedings of the 17th international conference on World Wide Web*, pages 247–256. ACM, 2008.

[29] S. Zhu. The roads taken: Theory and evidence on route choice in the wake of the I-35W Mississippi River bridge collapse and reconstruction. PhD thesis, University of Minnesota, 2011.