Statistical Analysis of Activity-Travel Behavior Characteristics

CEE 598: Activity-Travel Behavior Modeling and Simulation

Raswanth Prasath S V

School of Sustainable Engineering and the Built Environment
Arizona State University

Spring 2025

Contents

List of Figures

List of Tables

1 Introduction

This project aims to conducted a statistical analysis of travel behavior characteristics, examining how different demographic and socioeconomic factors influence various aspects of travel. The project examined five key aspects of travel behavior: activity patterns, trip purposes, transportation mode choices, time-of-day travel patterns, and activity durations for different purposes. Using statistical analysis and hypothesis testing, the revealed important differences in how various groups - such as workers vs. non-workers, older adults vs. others, and households with different income levels and car availability - interact with transportation systems. These insights can help inform transportation planning and policy decisions to better serve diverse community needs.

2 Data

This project uses data from the 2017 National Household Travel Survey (NHTS), obtained from http://nhts.ornl.gov/. As the nation's authoritative source on daily travel patterns, the NHTS captures detailed information about how, why, when, and where people travel. The NHTS data structure consists of four interrelated files: The Household File, Person File, Trip File and the Vehicle File. The data was filtered to include only households/persons residing in Arizona, New Mexico, Colorado, and Utah (`HHSTATE`). The analysis utilized 5,645 adult respondents (AGE $\geq$ 18 years) from the NHTS dataset, filtered to include only weekday travel in the four-state region.

3 Analysis of Activity Frequency and Duration

3.1 Activity Frequency Analysis: Males vs. females

The analysis examined differences in the frequency of daily out-of-home activity between males and females. The sample included 5,645 adults (2,661 males and 2,984 females). Out-of-home activities were defined as trips with purposes coded as work (10), shopping/errands (40), or social/recreational (50). The `DWELTIME` variable quantified activity duration, excluding trips with missing or negative values.

Table 1: Out-of-home activity duration of the compared segments: Descriptive statistics with t-test results

DESCRIPTIVE STATISTICS			
Gender	**Segment Size**	**Mean**	**Std. Dev**
Male	2661	2.30	2.16
Female	2984	2.37	2.23
T-TEST: INDEPENDENT SAMPLE ($\alpha = 0.05$)			
t = -1.175; df = 5643; p-value = 0.2399			
95% confidence interval: [-0.184, 0.046]			

The results indicate that while females engage in slightly more daily out-of-home activities on average (0.07 more activities per day), this difference is not statistically

significant at the 0.05 level. The relatively small p-value (0.239) and narrow confidence interval suggest that any true population difference in activity frequency between genders is likely to be modest in magnitude.

The lack of significant difference in activity frequency between males and females suggests that, in this region, daily mobility patterns are relatively gender-neutral in terms of the number of activities undertaken.

3.2 Activity Duration Analysis: Auto-deficient vs. Auto-sufficient

The analysis compared the total duration of out-of-home daily activity between auto-deficient and auto-sufficient households. The sample included 5,645 adults (632 from auto-deficient households and 5,013 from auto-sufficient households). Auto deficiency was determined by comparing household vehicle count (`HHVEHCNT`) to licensed drivers (`DRVRCNT`).

Table 2: Out-of-home activity duration of the compared segments: Descriptive statistics with t-test results

DESCRIPTIVE STATISTICS			
Auto sufficiency/ deficiency	**Segment Size**	**Mean (min)**	**Std. Dev**
Auto-deficient	632	196.22	215.87
Auto-sufficient	5013	243.28	233.57
T-TEST: INDEPENDENT SAMPLE ($\alpha = 0.05$)			
t = -4.812; df = 5643; p-value = 1.53e-06			
95% confidence interval: [-66.23, -27.89]			

The results show a highly significant difference in activity duration between auto-deficient and auto-sufficient households. Members of auto-sufficient households spend on average 47.06 more minutes per day in out-of-home activities compared to those in auto-deficient households. The extremely small p-value ($p < 0.001$) and confidence interval that excludes zero provide strong evidence that this difference represents a genuine population effect.

The approximately 47-minute difference suggests that limited vehicle access may constrain not just the ability to reach destinations, but also the amount of time people can spend engaging in out-of-home activities.

4 Chi-Square Analysis of Trip Purpose Distributions

4.1 Workers vs. Non-Workers Analysis

The analysis revealed substantial differences in trip purpose distribution between workers and non-workers. Trip purpose distributions were analyzed using the `WHYTRP1S` variable. A chi-square independence test showed a highly significant relationship between employment status and trip purpose ($\chi^2 = 2959.4$, df = 7, $p < 0.001$). The distributions of trips by purpose for workers and non-workers are exhibited in Table 3.

Table 3: Trip Purpose Distribution by Employment Status

Trip Purpose	Workers		Non-Workers	
	Count	Percent	Count	Percent
Home	3415	33.06%	3621	35.43%
Work	2563	24.81%	67	0.66%
School/Religious	144	1.39%	224	2.19%
Medical/Dental	142	1.37%	382	3.74%
Shopping/Errands	1607	15.56%	3071	30.05%
Social/Recreational	852	8.25%	1244	12.17%
Transport Someone	674	6.53%	586	5.73%
Meals	697	6.75%	746	7.30%
Other	235	2.28%	280	2.74%

Table 4: Pearson's Chi-squared Test Results for Trip Distribution by Employment Status

Statistic	Value	Degrees of Freedom	p-value
χ^2	3044.9	8	<2.2e-16

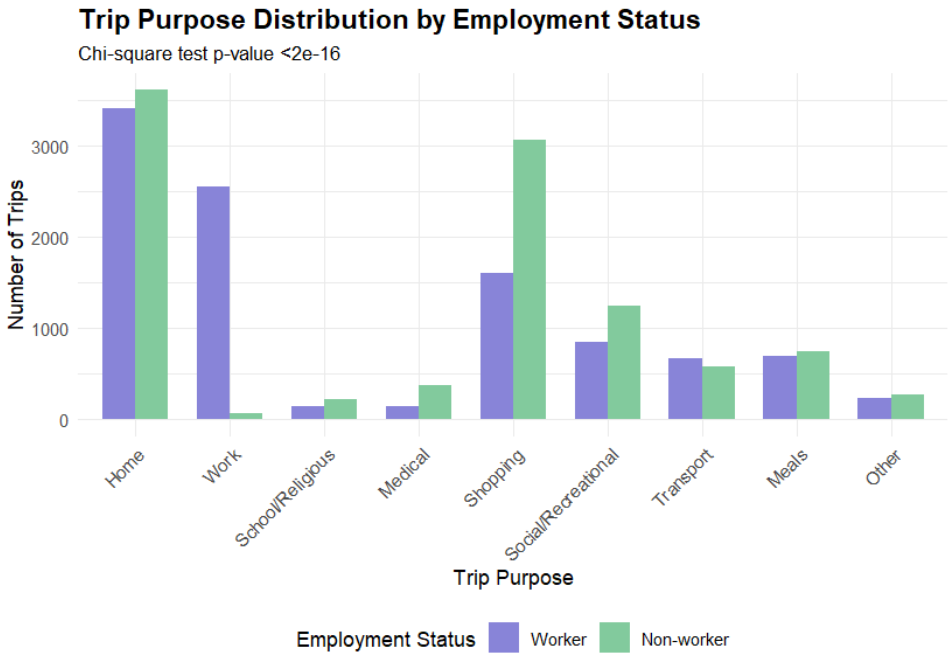

Figure 1: Trip Purpose Distribution by Employment Status

4.2 Age Group Analysis (65+ vs. Under 65)

The chi-square test showed significant differences in trip purpose distribution between age groups ($\chi^2 = 1226.2$, df = 7, p <0.001).

Table 5: Trip Purpose Distribution for 65+ Years Old vs. 18-64 Years Old

Trip Purpose	65+ Years Old		18-64 Years Old	
	Count	Percent	Count	Percent
Home	2699	35.62%	4337	33.43%
Work	273	3.60%	2357	18.17%
School/Religious	120	1.58%	248	1.91%
Medical/Dental	281	3.71%	243	1.87%
Shopping/Errands	2246	29.64%	2432	18.75%
Social/Recreational	887	11.70%	1209	9.32%
Transport Someone	297	3.92%	963	7.42%
Meals	555	7.32%	888	6.85%
Other	220	2.90%	295	2.27%

Table 6: Pearson's Chi-squared Test Results for Trip Distribution by Age

Statistic	Value	Degrees of Freedom	p-value
χ^2	1246.7	8	<2.2e-16

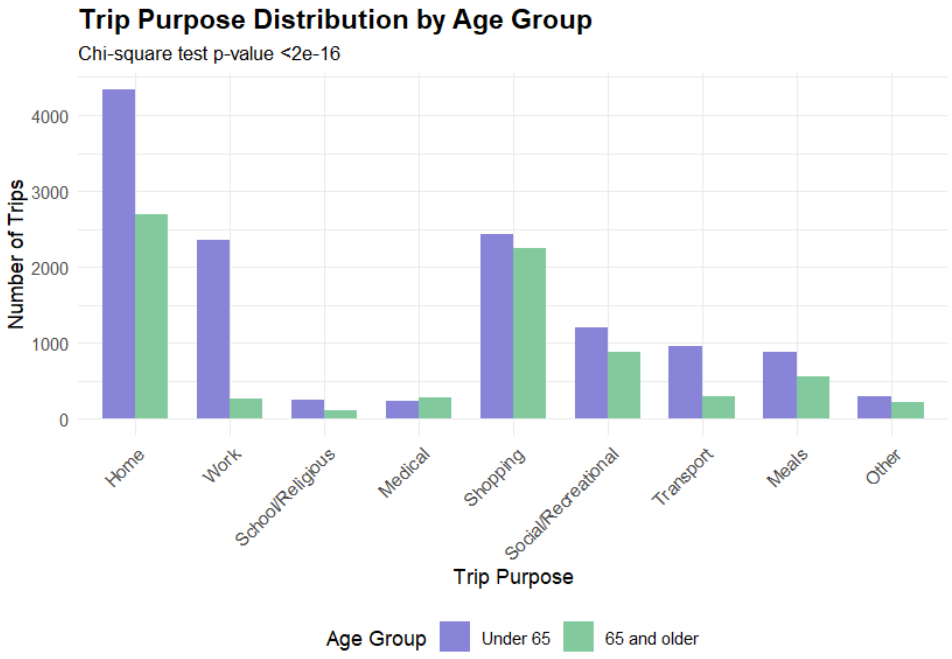

Figure 2: Trip Purpose Distribution by Age Group

5 Analysis of Travel Mode Choices

The National Household Travel Survey (NHTS) 2017 dataset provides detailed information about trip characteristics through various variables. Two key variables used in this analysis are `TRPTRANS` and `NUMONTRP`. The `TRPTRANS` variable indicates the mode of transportation used for each trip, while `NUMONTRP` represents the number of people on the trip, including the respondent. A new mode variable was created by combining `TRPTRANS`

(trip mode) and `NUMONTRP` (occupancy).

To align our analysis with commonly used categories in transportation, we rearranged the original NHTS trip mode categories. This rearrangement process involved consolidating some categories and creating new ones based on both the mode of transportation (`TRPTRANS`) and the number of occupants (`NUMONTRP`).

Table 7 presents the outcome of this rearrangement exercise. The new categories include:

Table 7: Rearranged Trip Mode Categories

New Trip Mode	**TRPTRANS**	**NUMONTRP**
SOV	03=Car 04=SUV 05=Van 06=Pickup truck	1
HOV2	03=Car 04=SUV 05=Van 06=Pickup truck	2
HOV3+	03=Car 04=SUV 05=Van 06=Pickup truck	> 2
Bus	10=School bus 11=Public or commuter bus 13=Private / Charter / Tour / Shuttle bus 14=City-to-city bus (Greyhound, Megabus)	-
Rail	15=Amtrak / Commuter rail 16=Subway / elevated / light rail / streetcar	-
Bike	02=Bicycle	-
Walk	01=Walk	-
Ride/carsharing	17=Taxi / limo (including Uber / Lyft) 18=Rental car (Including Zipcar / Car2Go)	-
Other	07=Golf cart / Segway 08=Motorcycle / Moped 09=RV (motor home, ATV, snowmobile) 12=Paratransit / Dial-a-ride 19=Airplane 20=Boat / ferry / water taxi 97=Something Else	-

5.1 Home-Based Work Trips: Auto-Deficient vs. Auto-Sufficient

This section examines the modal split for home-based work (HBW) trips, comparing individuals from auto-deficient and auto-sufficient households. An auto-deficient household is defined as having fewer vehicles than licensed drivers, while an auto-sufficient household has an equal or greater number of vehicles compared to licensed drivers.

5.1.1 Modal Split Distribution

Figure 1 presents the modal split distribution for HBW trips by auto availability. Single-occupancy vehicle (SOV) is the most prevalent mode for both groups. However, the percentage of SOV trips is significantly higher for individuals in auto-sufficient households (82.0%) compared to auto-deficient households (61.5%). A higher percentage of individuals in auto-deficient households use high-occupancy vehicle (HOV2) modes (21.1%) compared to those in auto-sufficient households (9.1%). Transit modes (bus and rail) and active modes (bike and walk) are used at a higher rate by auto-deficient households.

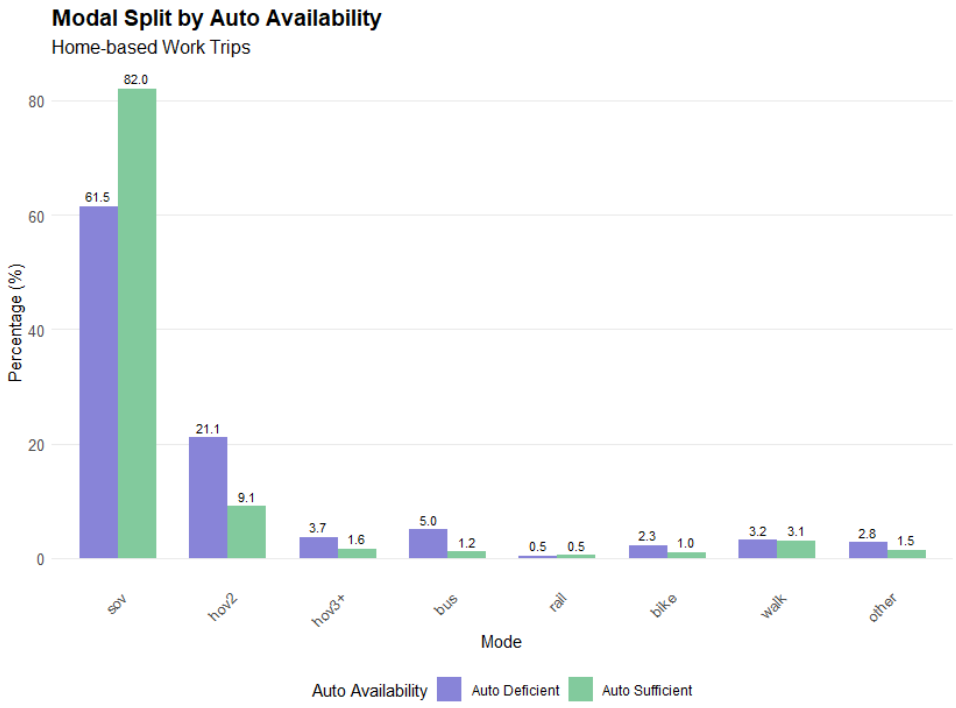

Figure 3: Modal Split by Auto Availability

Figure 3 visualizes the modal split distribution for HBW trips by auto availability, clearly illustrating the differences in mode choice between auto-deficient and auto-sufficient households.

5.1.2 Statistical Analysis

A two-sample t-test was conducted to determine if there is a significant difference in SOV usage between auto-deficient and auto-sufficient households. The results are presented in Table 8.

The t-test results show that the two segments are statistically different from one another (t = -7.403; df = 2720; p <0.001). This result supports the idea that individuals

from auto-deficient households use single-occupancy vehicles to a lesser extent than those from auto-sufficient households. The 95% confidence interval for the difference in means is [-0.2601, -0.1511], suggesting that auto-deficient households use SOVs between 15.11% and 26.01% less often than auto-sufficient households for home-based work trips.

Table 8: T-Test of SOV Usage for Auto-Deficient vs. Auto-Sufficient Households

Statistic	Value	Degrees of Freedom	p-value
t-statistic	-7.403	2720	<0.001

5.2 All Trips: Low Income vs. High Income

This section examines the modal split for all trips, comparing individuals from low-income households (income $29,999 or less) with those from high-income households (income $75,000 or more). Income-based mode splits used `HHFAMINC` to define low-income ($\leq$$29,999) and high-income ($\geq$$75,000) groups.

5.2.1 Modal Split Distribution

Single-occupancy vehicle (SOV) trips are more prevalent among high-income households (52.48%) compared to low-income households (45.00%). Low-income individuals use active transportation modes more frequently, with higher percentages for walking (11.88% vs. 8.25%) and biking (1.70% vs. 0.84%). Low-income households show higher usage of public transit, with bus trips at 2.53% (vs. 0.50% for high-income) and rail trips at 0.43% (vs. 0.13% for high-income).

High-occupancy vehicle usage is relatively similar between the two income groups, with a slightly higher percentage of HOV2 trips for low-income households. Table 9 presents the modal split distribution for all trips by income level.

Table 9: Modal Split: Low Income vs. High Income

Trip Mode	Income $29,999 or less		Income $75,000 or more	
	Count	Percent	Count	Percent
SOV	1459	45.00%	4173	52.48%
HOV2	832	25.66%	1949	24.51%
HOV3+	356	10.98%	907	11.41%
Walk	385	11.88%	656	8.25%
Bike	55	1.70%	67	0.84%
Bus	82	2.53%	40	0.50%
Rail	14	0.43%	10	0.13%
Other	59	1.82%	149	1.87%
Total	3242	100.00%	7951	99.99%

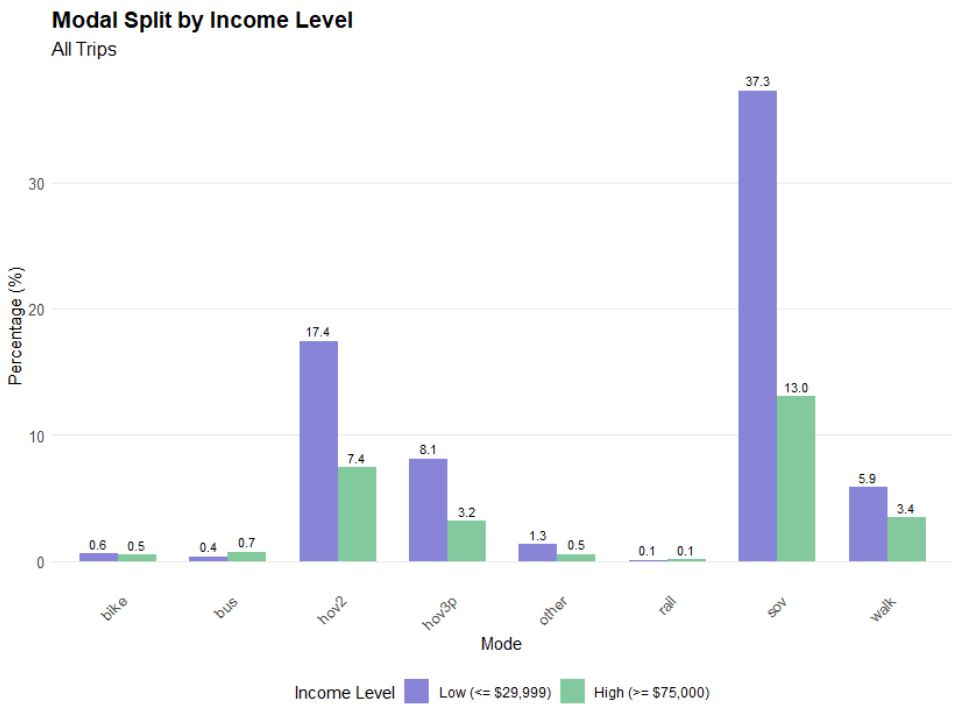

Figure 4: Modal Split by Income Level

5.2.2 Statistical Analysis

A two-sample t-test was conducted to determine if there is a significant difference in SOV usage between low-income and high-income groups. The results are presented in Table 10.

Table 10: T-Test Results: SOV Usage by Income Level

Statistic	Value	Degrees of Freedom	p-value
t-statistic	7.1961	11191	<0.001

The t-test results indicate a statistically significant difference in SOV usage between low-income and high-income groups ($t = 7.1961$, $df = 11191$, $p < 0.001$). The mean SOV usage for low-income households (0.4500) is significantly lower than that of high-income households (0.5248). The 95% confidence interval for the difference in means is [0.0544, 0.0952], suggesting that high-income households use SOVs between 5.44% and 9.52% more often than low-income households.

These findings suggest that income level has a substantial impact on mode choice, particularly in the use of single-occupancy vehicles, public transit, and active transportation modes.

6 Time of Day Distributions

The temporal distribution of trips was analyzed with particular focus on peak period travel patterns and the differences between work and non-work trips. The analysis examined trips during the traditional peak periods: AM peak (6:00-8:59 AM) and PM peak

(4:00-6:59 PM). Time-of-day distributions used `STRTTIME` (start time) for home-based work trips which are filtered by `TRIPPURP="HBW"` and binned into 30- minute intervals. Home-based shopping trips are filtered by `TRIPPURP="HBSHOP`. AM (6:00–8:59 AM) and PM (4:00–6:59 PM) peaks were analyzed to calculate work-related trip percentages using `WHYFROM/WHYTO` variables (codes 3–4 for work).

6.1 Peak Period Analysis

The analysis of peak period travel revealed distinct differences between morning and evening travel patterns. During the AM peak period (6:00-8:59 AM), work-related trips constitute 38.7% of all trips, while during the PM peak period (4:00-6:59 PM), they make up 26.6% of trips. This asymmetry suggests that morning travel is more concentrated around work purposes compared to evening travel.

Table 11: Percent of trips with work end purposes and other purposes during the peak times

Trip Purpose	AM Peak	PM Peak
Other	2,044 (61.3%)	3,048 (73.4%)
Work-related	1,287 (38.7%)	1,104 (26.6%)
Total	3,331 (100%)	4,152 (100%)

The PM peak period has a higher total number of trips (4,152) compared to the AM peak (3,331), suggesting more diverse travel purposes in the evening. Non-work trips dominate both peak periods, but their share is notably higher during the PM peak (73.4% vs 61.3%). Work-related trips show stronger concentration in the morning peak, with nearly 40% of AM peak trips being work-related. The lower share of work trips during the PM peak (26.6%)

The temporal distribution of home-based work trips (Figure 5) shows distinct morning and afternoon peak periods.

The temporal distribution of home-based Shopping trips (Figure 6) shows distinct morning and afternoon peak periods.

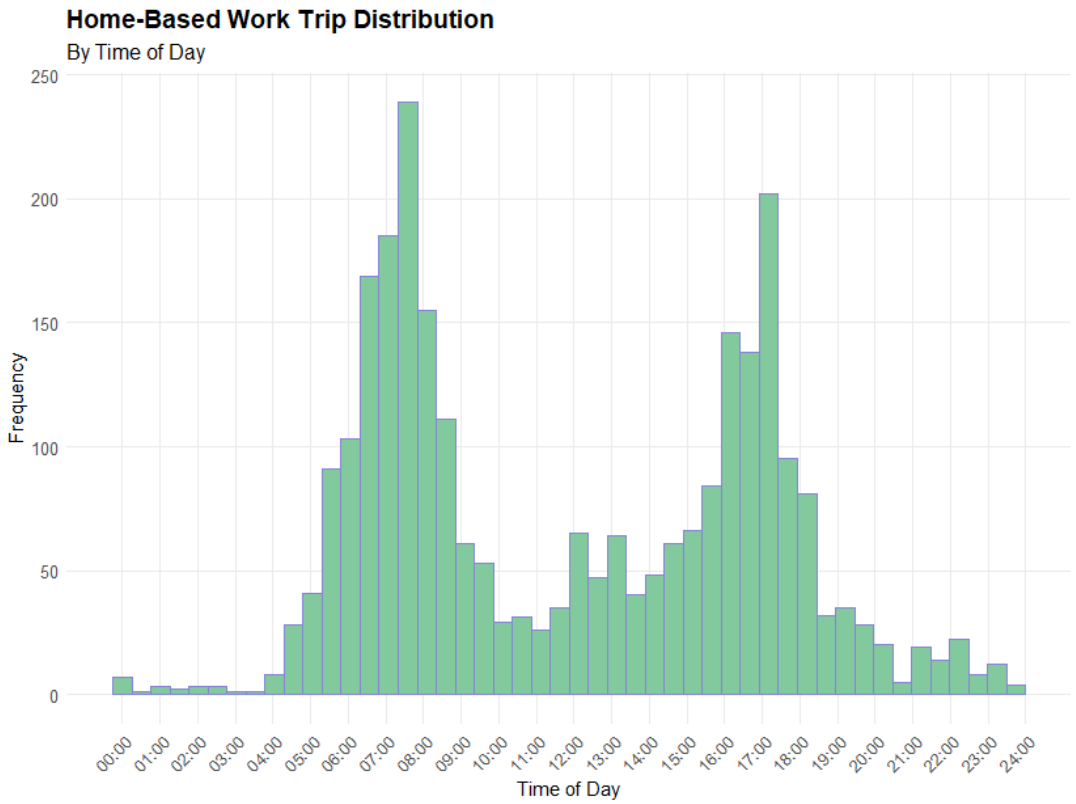

Figure 5: Home-Based Work Trip Distribution

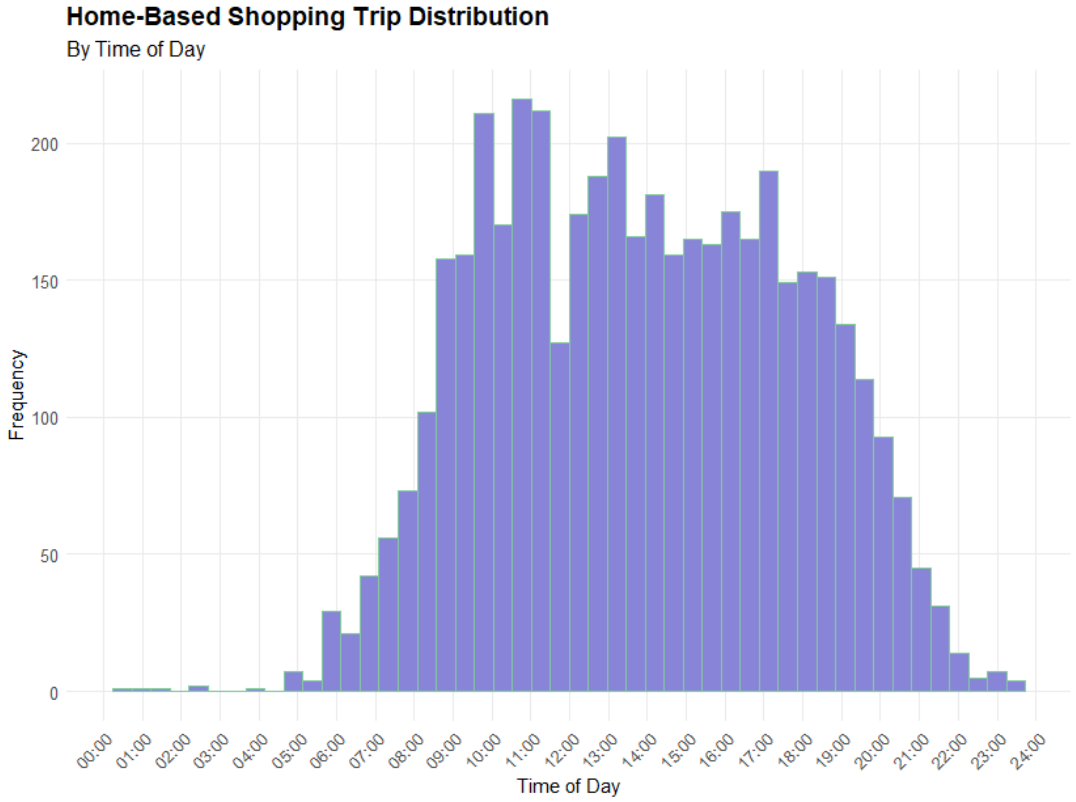

Figure 6: Home-Based Shopping Trip Distribution

7 Activity Duration Distributions

This section analyzes the daily activity durations for three key activity types: Work, Shopping/Errands, and Social/Recreational activities. Work (`WHYTRP1S=10`), shopping (`WHYTRP1S=40`), and social/recreational (`WHYTRP1S=50`) activities were aggregated per person. Zero-duration participants were excluded for this analysis and the rationale was to focus only on meaningful activities with measurable durations.

7.1 Work Activity Duration

The distribution of work activity durations is shown in Figure 7. The participation rate for work activities is 30.3%, with 1,711 participants out of 5,647 individuals. The duration statistics are summarized in Table 12.As the distribution is getting close to a bell-shaped distribution, the sample respondents' work duration appears to be about normal.

Table 12: Work Activity Duration Statistics (in minutes)

Min	1st Quartile	Median	Mean	3rd Quartile	Max
1	365	495	452.8	555	1035

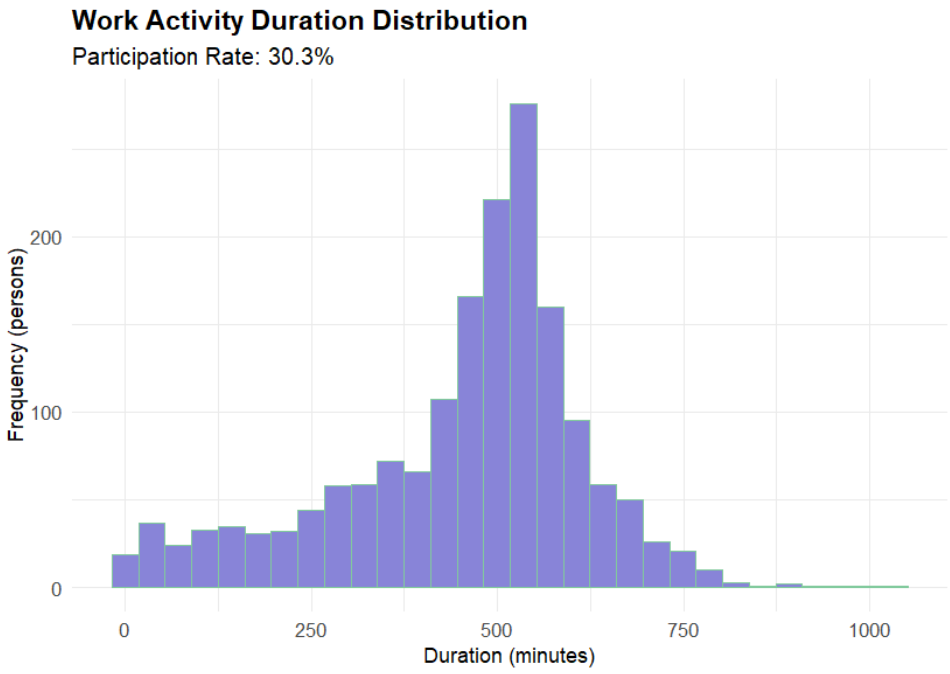

Figure 7: Work Activity Duration Distribution

7.2 Shopping/Errands Activity Duration

The distribution of shopping/errands activity durations is presented in Figure 8. The participation rate is 43.9%, with 2,479 participants out of 5,647 individuals. Table 13 provides a summary of the duration statistics.

Table 13: Shopping/Errands Activity Duration Statistics (in minutes)

Min	1st Quartile	Median	Mean	3rd Quartile	Max
1	20	43	58.73	75	720

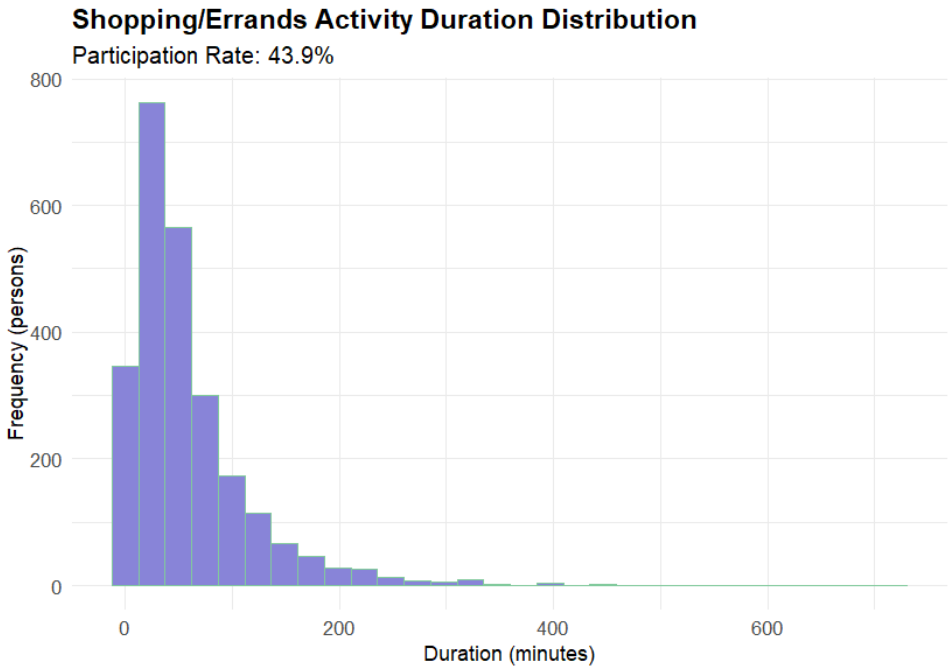

Figure 8: Shopping/Errands Activity Duration Distribution

7.3 Social/Recreational Activity Duration

The distribution of social/recreational activity durations is shown in Figure 9. The participation rate is 25.7%, with 1,451 participants out of 5,647 individuals. Table 14 summarizes the duration statistics.

Table 14: Social/Recreational Activity Duration Statistics (in minutes)

Min	1st Quartile	Median	Mean	3rd Quartile	Max
1	59.5	107	138.5	188	712

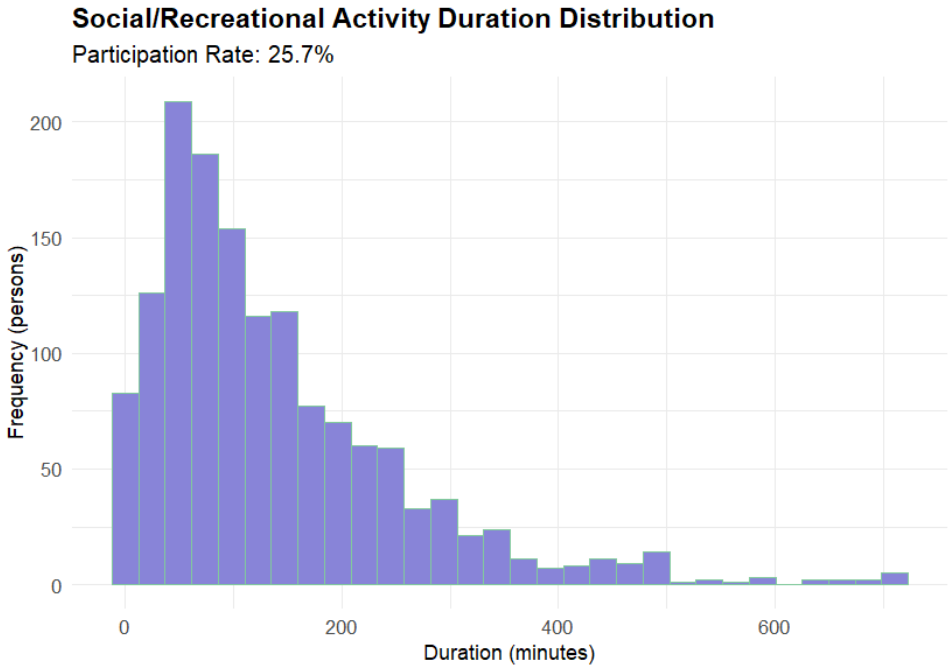

Figure 9: Social/Recreational Activity Duration Distribution

The analysis highlights that work activities have the longest average duration (452.8 minutes), followed by social/recreational activities (138.5 minutes) and shopping/errands activities (58.73 minutes). These findings provide insights into how individuals allocate their time across different activity types.