Project
STA 2023: Statistics I
Spring 2019

Descriptive Analysis and Hypothesis Testing

Submitted to
Dr. Sravani Vadlamani
Florida Polytechnic University

Submitted by
Luiz Gustavo Fagundes Malpele
Department of Data Science
Florida Polytechnic University

April 24, 2019

Table of Contents

List of Figures

List of Tables

1. Introduction

The purpose of this project is to investigate the travel behavior characteristics of adults from Orlando-Kissimmee-Sanford region, this project also test hypothesis regarding the behavior of the adult in the subsample. The data comes from the 2017 National Household Travel Survey (NHTS) and it includes all trips made by each adult on the sample and the details regarding this trip.

2. Data

The data of the project is below, there were used bar charts to represent categorical variables and summary statistic for continuous variables.

Variable	Mean	Std Dev	Minimum	Maximum
trpmiles_sum	35.9173620	29.8865271	1.5000000	184.7800000
trvlcmin_sum	97.8527607	66.3146478	10.0000000	415.0000000
trpcnt_sum	4.3558282	2.4911720	1.0000000	15.0000000

TABLE 1. CONTINUOUS VARIABLES FOR AGGREGATED DATA SET

Table 1. Summary statistics for the Aggregated trip's data set, it contains data such as mean, standard deviation, minimum, and maximum for three different variables: trpmiles_sum (Total trip distance by a person), trvlcmin_sum (Tota minutes travelled by a person), and trpcnt_sum (Total number of trips made by a person).

Variable	Mean	Std Dev	Minimum	Maximum
trpmiles	8.2458451	11.2402293	0.0400000	92.3300000
trvlcmin	22.4647887	21.3315099	1.0000000	270.0000000
hhvehcnt	2.2338028	1.1186260	0	6.0000000

TABLE 2. CONTINUOUS VARIABLES FOR CONDENSED DATA SET

Table 2. Summary statistics for the Condensed trip's data set, it contains data such as mean, standard deviation, minimum, and maximum for three different variables: trpmiles (Distance in miles for each trip made by a person), trvlcmin (Trip duration in minutes), and hhvehcnt (Amount of household vehicles).

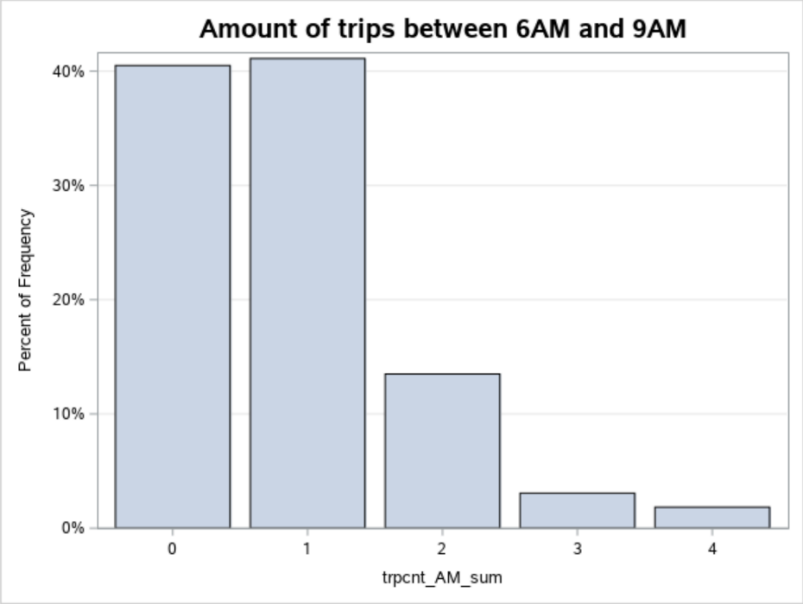

FIGURE 1 NUMBER OF TRIPS BETWEEN 6AM AND 9AM

Figure 1. Bar chart for number of trips that happened between 6AM and 9AM during week days (from Monday to Friday).

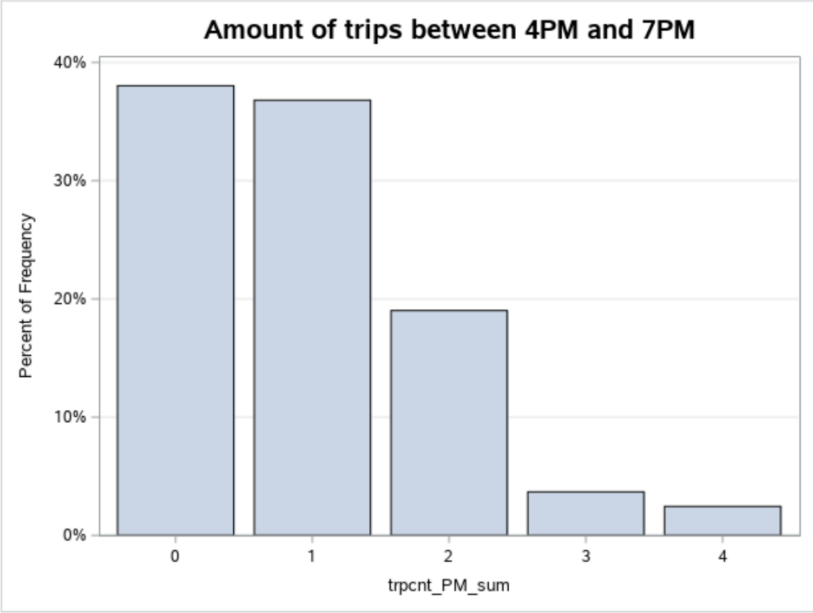

FIGURE 2. NUMBER OF TRIPS BETWEEN 4PM AND 7 PM

Figure 2. Bar chart for number of trips that happened between 4PM and 7PM during week days (from Monday to Friday).

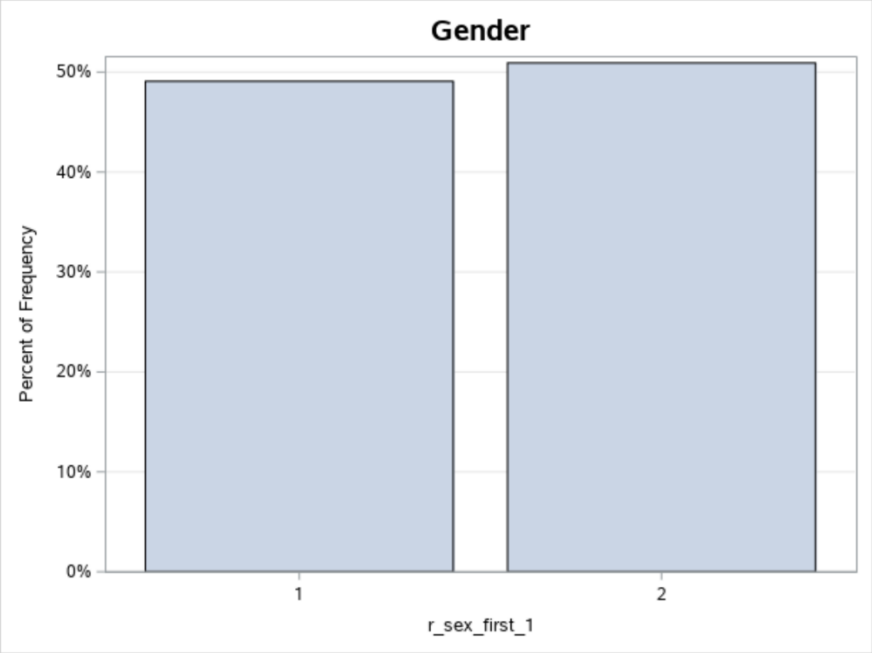

FIGURE 3. GENDER

Figure 3. Bar chart for gender category: 1 (male), 2 (female).

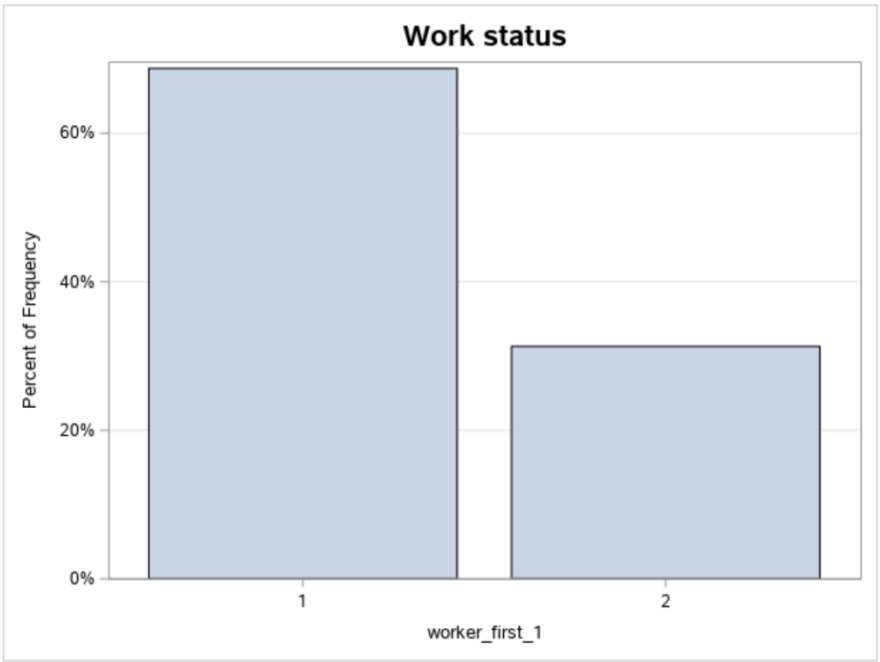

FIGURE 4. WORK STATUS

Figure 4. Bar chart for work status: 1 (Worker), 2 (Non-worker).

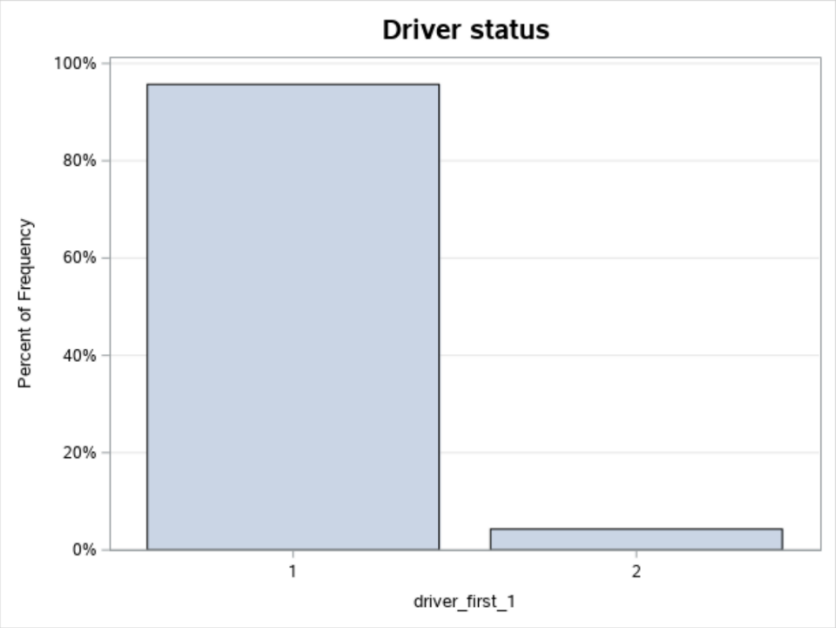

FIGURE 5. DRIVER STATUS

Figure 5. Bar chart for driver status: 1 (Driver), 2 (Not driver).

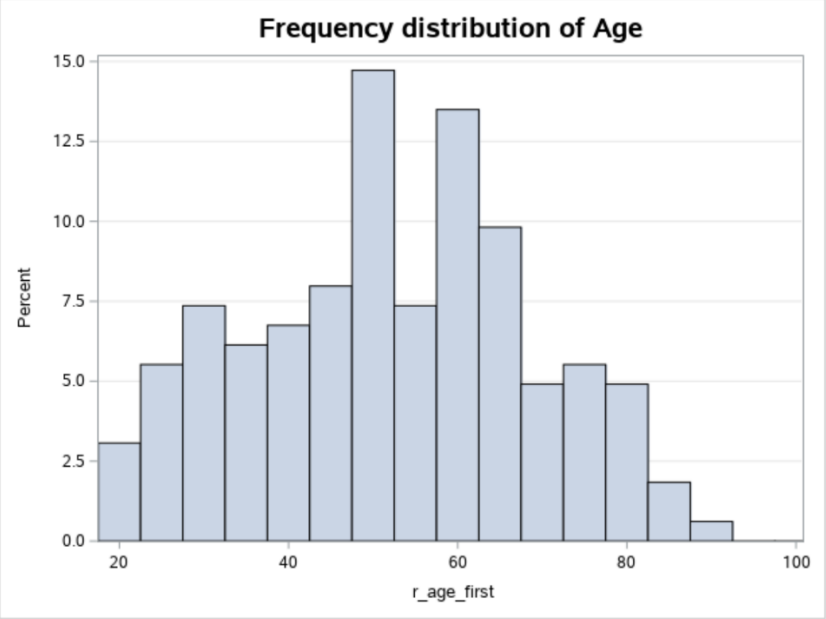

FIGURE 6. FREQUENCY DISTRIBUTION OF AGE

Figure 6. Histogram describing the frequency distribution of respondent's ages. The decision to show it in a histogram was to better visibility of data.

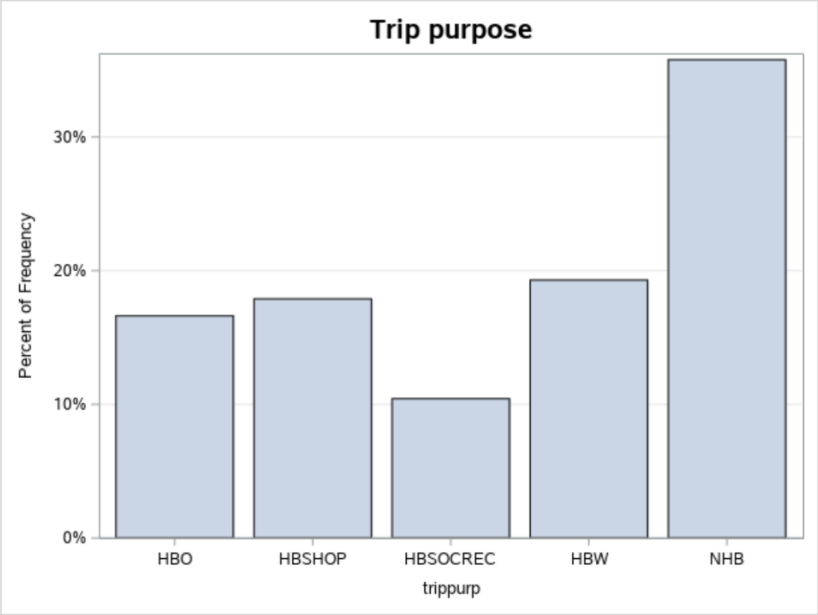

FIGURE 7. TRIP PURPOSE

Figure 7. Bar chart for purpose of trips, it was divided into five different categories which are: HBO (Home Base Other), HBSHOP (Home Based Shopping), HBSOCREC (Home Based Social Recreation), HBW (Home Based Work), and NHB (Non-Home Based).

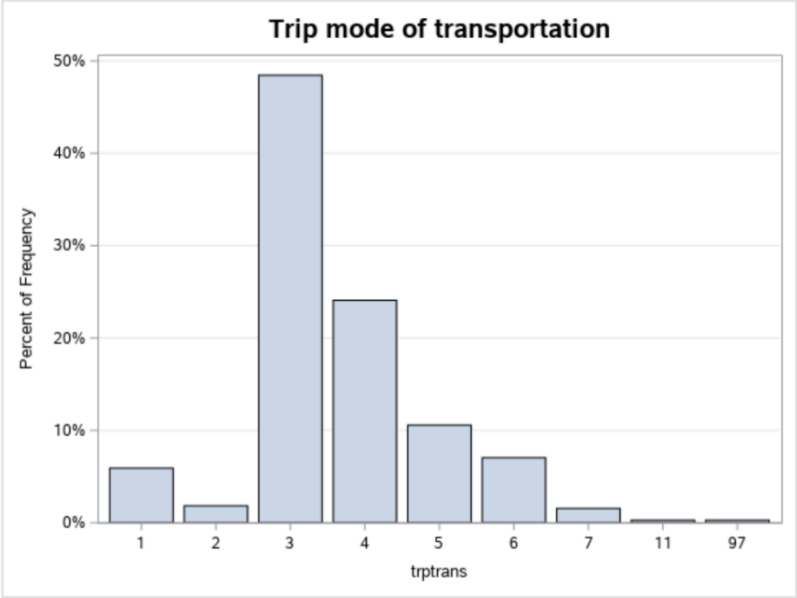

FIGURE 8. TRIP MODE OF TRANSPORTATION

Figure 8. Bar chart for purpose of trip's mode of transportation, it was divided into nine different categories which are: 1 (Walk), 2 (Bicycle), 3 (Car), 4 (SUV), 5 (Van), 6 (Pick Up Truck), 7 (Gold Cart/Segway), 11 (Public/Commuter Bus), and 97 (Other).

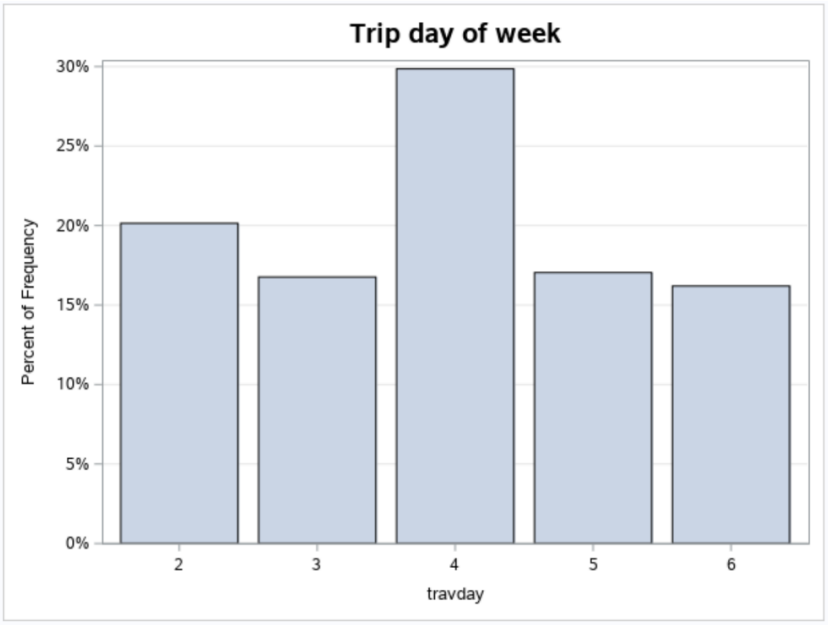

FIGURE 9. TRIP DAY OF WEEK

Figure 9. Bar chart which shows frequency of trips for each day of the week, it was divided into five different categories which are: 2 (Monday), 3 (Tuesday), 4 (Wednesday), 5 (Thursday), and 6 (Friday).

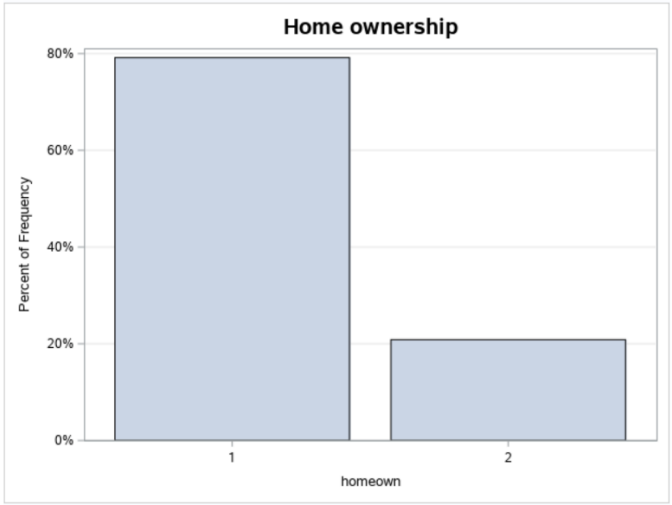

FIGURE 10. HOME OWNERSHIP

Figure 10. Bar chart which shows the home ownership related to respondents, the two categories stand for 1 (Own) and 2 (Rent).

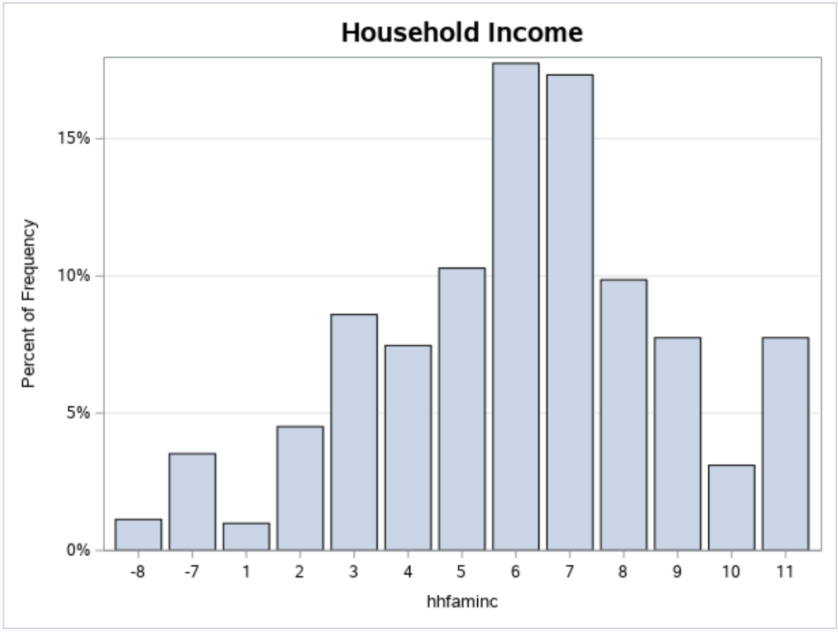

FIGURE 11. HOUSEHOLD INCOME

Figure 11. Bar chart for household income of respondents, it was divided into thirteen different categories which are: -8 (I don't know), -7 (I prefer not to answer), 1 (less than or equal to $10,000), 2 (from $10,000 to $14,999), 3 (from $15,000 to $24,999), 4 (from $25,000 to $34,999), 5 (from $35,000 to 49,999), 6 (from $50,000 to $74,999), 7 (from $75,000 to $99,999), 8 (from $100,000 to 124,999), 9 (from $125,000 to $149,999), 10 (from $150,000 to $199,999), and 11 (More than or equal to $200,000).

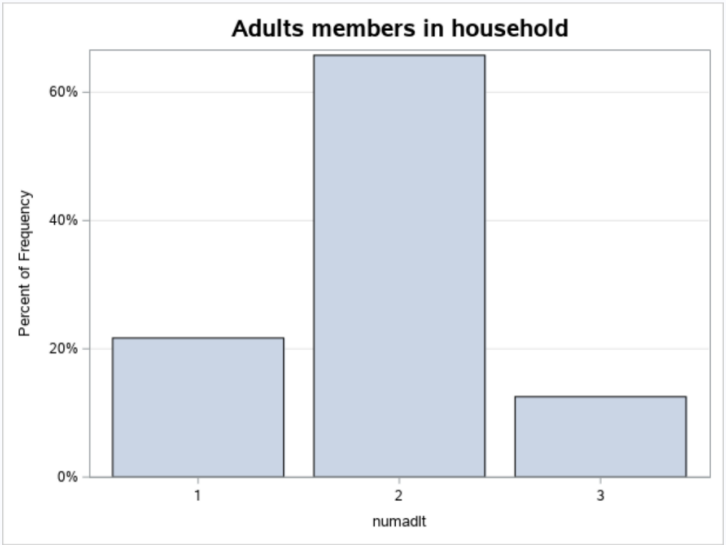

FIGURE 12. ADULT MEMBERS IN HOUSEHOLD

Figure 12. Bar chart for number of adults, at least 18 years old, who are household members for the same house, the numerical variables are in a range from 1 to 3.

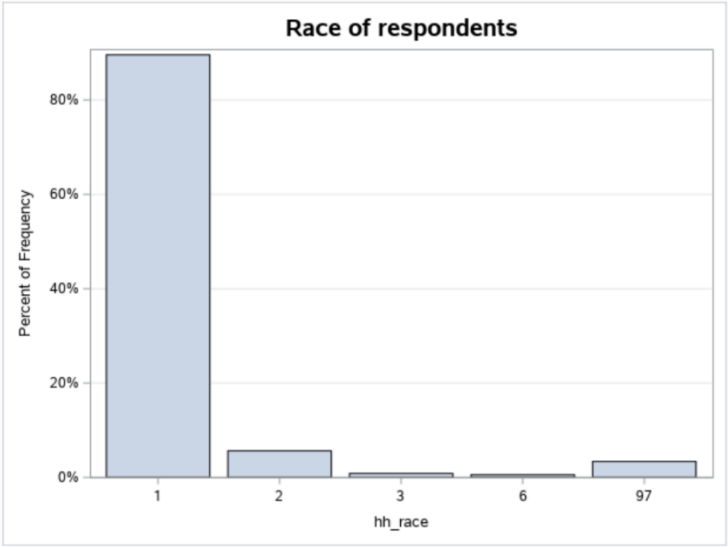

FIGURE 13. RACE OF RESPONDENTS

Figure 13. Bar chart for race of household respondents, it was divided into five different categories which are: 1 (White), 2 (Black or African American), 3 (Asian), 6 (Mixed Race), and 97 (Other).

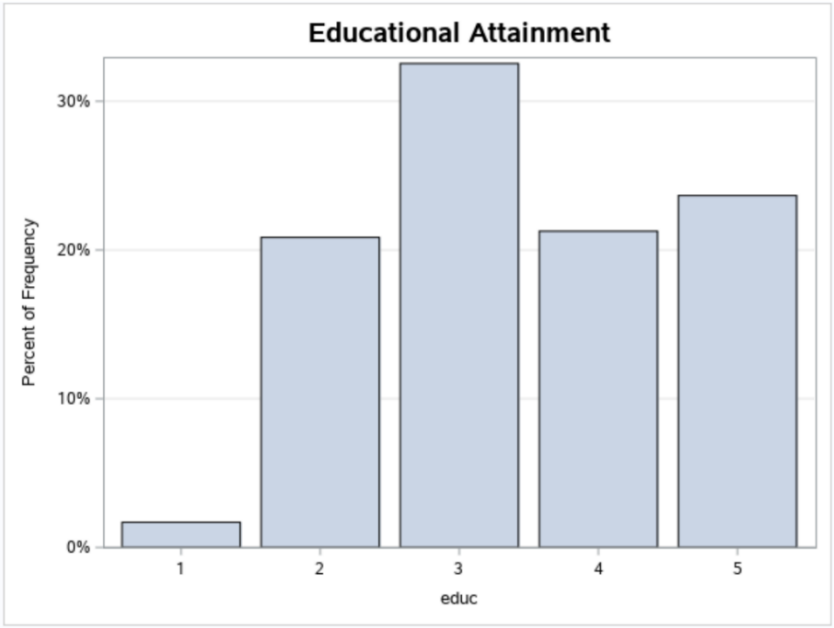

FIGURE 14. EDUCATIONAL ATTAINMENT

Figure 14. Bar chart for educational attainment of respondents, it was divided into five different categories which are: 1 (Less than high school), 2 (High School graduate or GED), 3 (Associate degree), 4 (Bachelor's Degree), 5 (Graduate or Professional Degree).

3. Distribution of Trip by Purpose

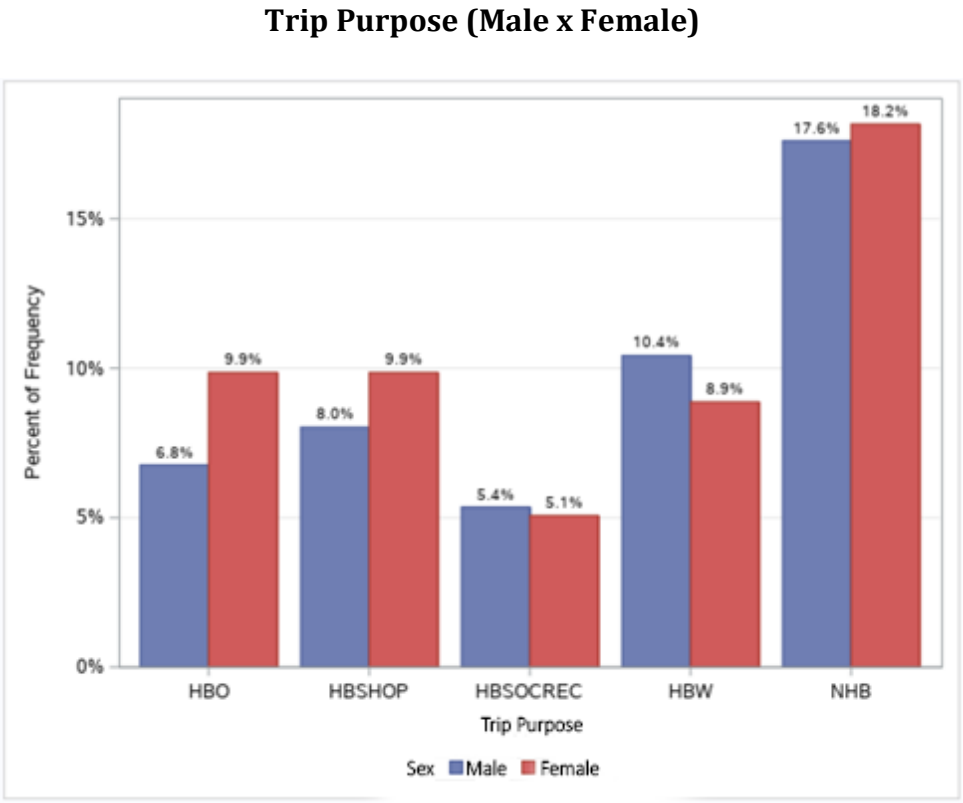

FIGURE 15. TRIP PURPOSE (MALE X FEMALE)

Figure 15. The data was divided into men and women and the percentage frequency of trip purpose was investigated for both. The trip purpose category was divided into HBO (Home Based Other), HBSHOP (Home Based Shopping), HBSOCREC (Home Based Social Recreation), HBW (Home Based Work), NHB (Non-Home Based).

The frequency distribution of trip purpose shows a contrast on home-based work and non-home-based work for men and women. It is noticeable that for home-based work the percentage of men (10.4%) is superior to women (8.9%) and the inverse also applies for non-home-based work, which the percentage of women (18.2%) surpasses the men's (17.6%). Last but not least, the biggest differences are observed on home based others and home-based shop, which indicate a trend of women travelling for duties such as shopping (9.9%) and for others (9.9%).

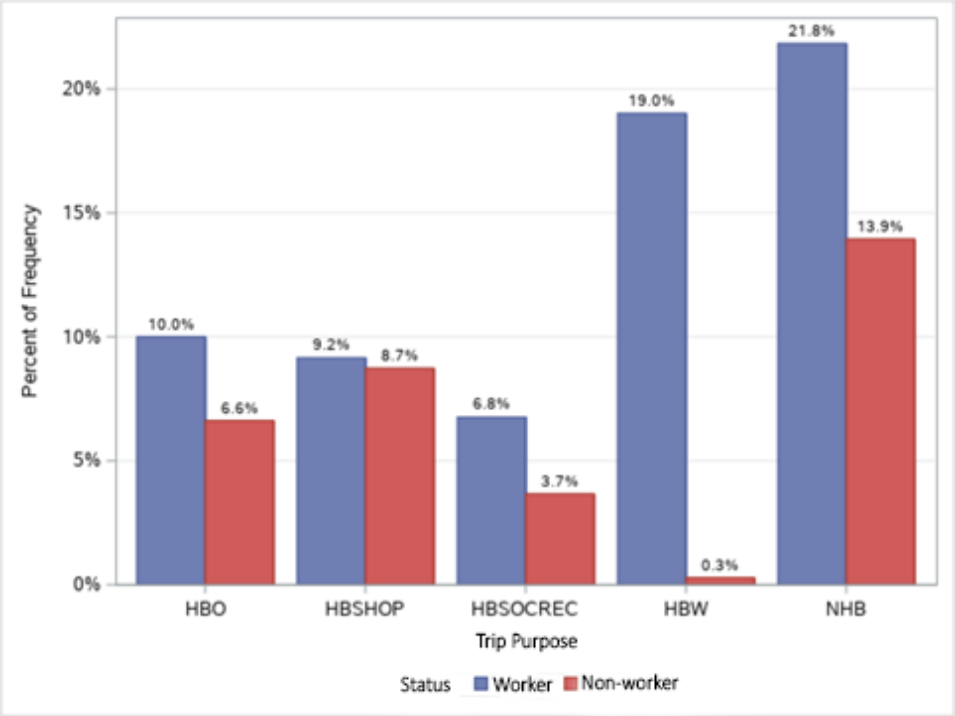

FIGURE 16. TRIP PURPOSE (WORKERS X NON-WORKERS)

Figure 16. The data was divided into workers and non-workers and the percentage frequency of trip purpose was investigated for both. The trip purpose category was divided into HBO (Home Based Other), HBSHOP (Home Based Shopping), HBSOCREC (Home Based Social Recreation), HBW (Home Based Work), NHB (Non-Home Based).

When the frequency distribution is divided into workers and non-workers the biggest differences are noticeable on the Home-Based Work and Non-Home-Based Work and the reason is the fact that if a person is not working, there is no reason to travel for this reason. On all categories the workers surpass the non-workers, the reason for that is probably the fact that workers have more purchase power in relation to non-workers, it is also important to point out that non-workers and workers have a similar percentage when it comes to home-based shopping and it is reasonable as both need to buy at least basic products for everyday life.

4. Distribution of Trips by Mode of Transportation

Trip transportation for all trips (Male x Female)

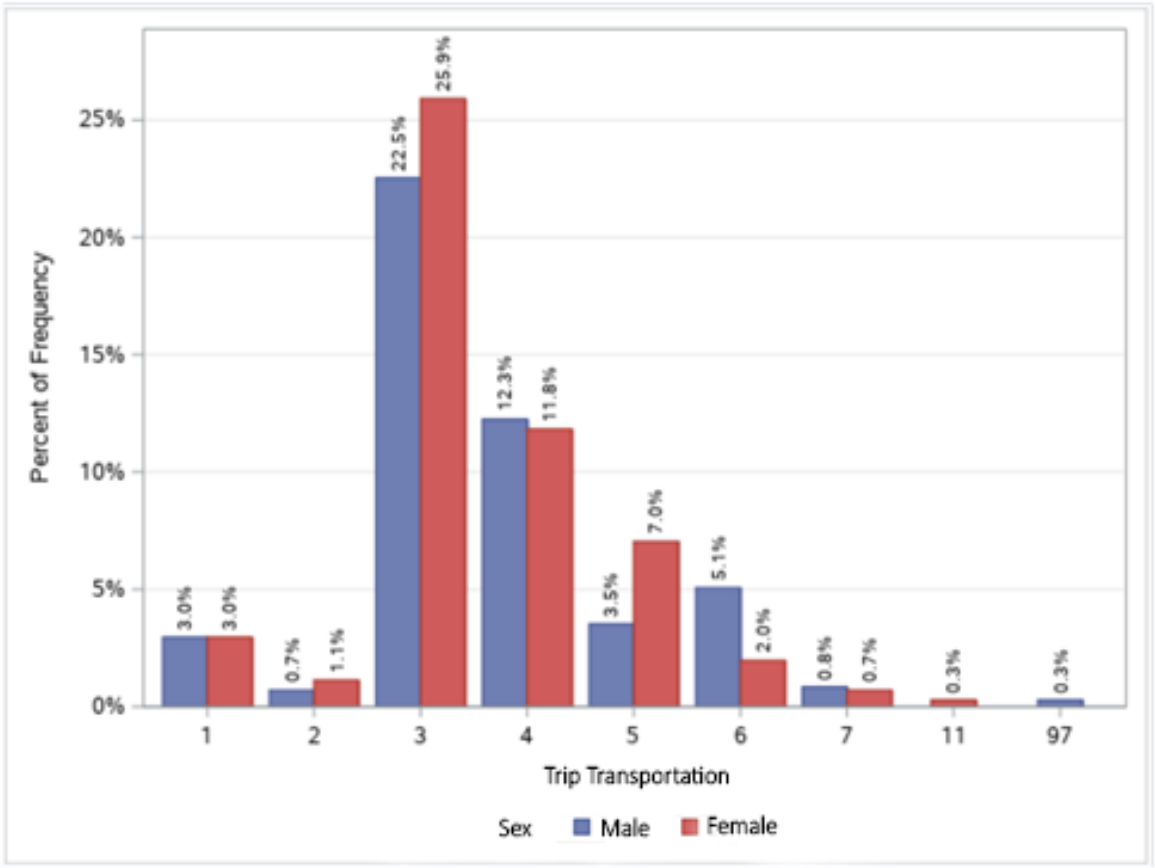

FIGURE 17. TRIP TRANSPORTATION FOR ALL TRIPS (MALE X FEMALE)

Figure 17. The data was divided into female and male and the percentage frequency of transportation methods was investigated for all trips. The trip transportation category was divided into 1 (Walk), 2 (Bicycle), 3 (Car), 4 (SUV), 5 (Van), 6 (Pick Up Truck), 7 (Golf Cart/Segway), 11(Public/Commuter Bus), 97 (Other).

The chart shows three major differences when it is divided upon sex, the first is the usage of car as transportation women (25.9%) surpass men (22.5%), in this category the difference of the percentage means 13.1% more for women. The second one occurs on the Van category, which women (7.0%) surpass men (3.5% by relatively 50%. The last noticeable trend is on pickup truck, men are 5.1% and women 2.0%, which is 58.8% more.

Trip transportation for Home Based Work (Male x Female)

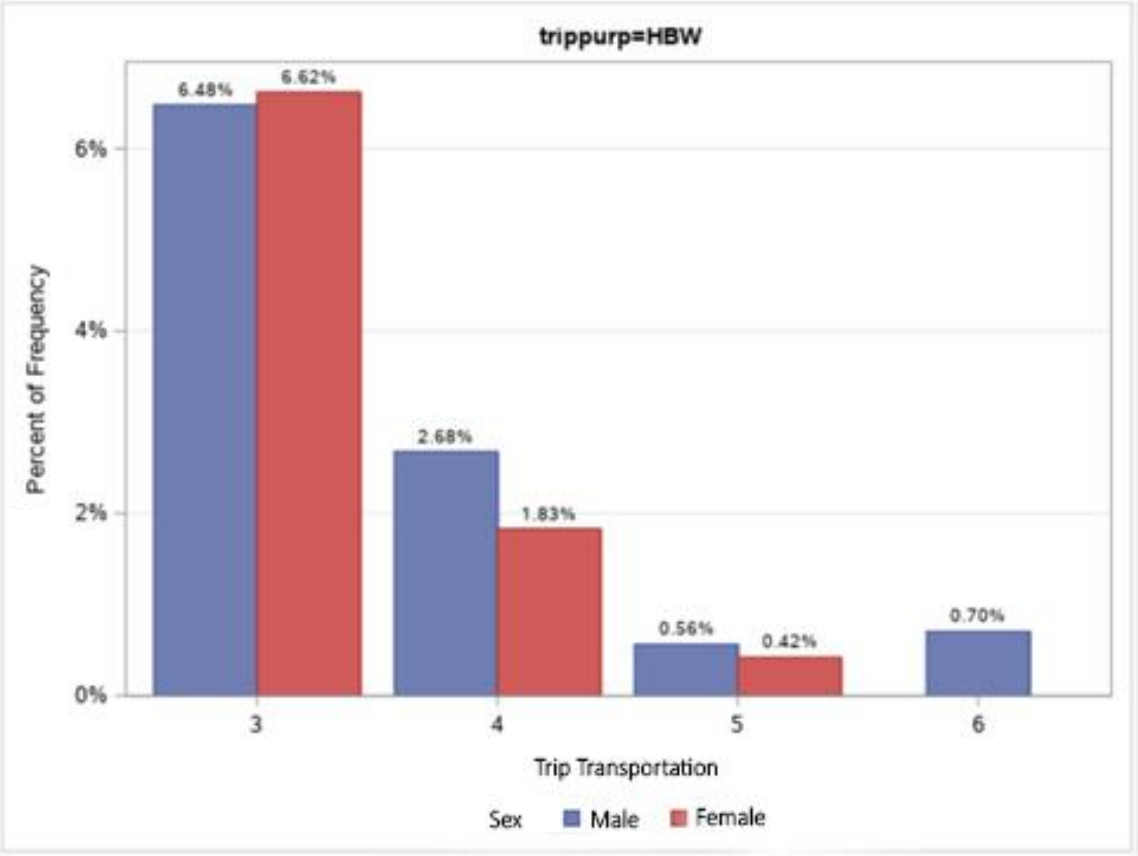

FIGURE 18. TRIP TRANSPORTATION FOR HBW (MALE X FEMALE)

Figure 18. The data was divided into female and male and the percentage frequency of transportation methods was investigated for home-based-work. The trip transportation category was divided into 3 (Car), 4 (SUV), 5 (Van), 6 (Pick Up Truck).

There are noticeable differences for home-based work graph, there transportation methods were reduced to four, in the all trips category there were nine categories, the existing ones for this frequency distribution are car, SUV, Van, and Pick Up Truck. The data for car usage is similar for both male and female, only 0.14% of difference, while for SUVs men usage surpass women in 0.85%, which is relatively 31.7% more. Lastly, only male respondents appear on the pick-up truck category.

Trip transportation for Home Based Shopping (Male x Female)

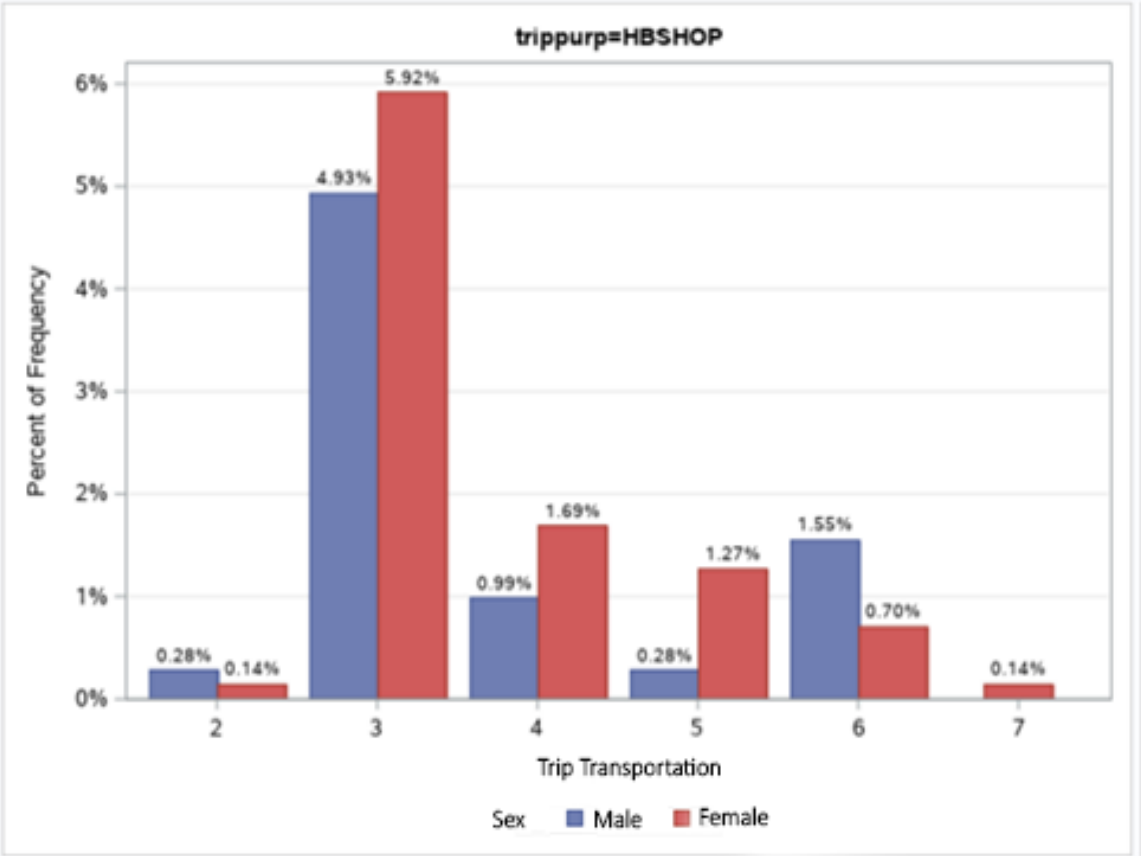

FIGURE 19. TRIP TRANSPORTATION FOR HBS (MALE X FEMALE)

Figure 19. The data was divided into female and male and the percentage frequency of transportation methods was investigated for home-based-work. The trip transportation category was divided into 1 (Walk), 2 (Bicycle), 3 (Car), 4 (SUV), 5 (Van), 6 (Pick Up Truck), 7 (Golf Cart/Segway), 11(Public/Commuter Bus), 97 (Other).

This chart presents the same characteristic as the last, it does not include some methods of transportation such as walk, public/commuter bus, and the category other. In the home-based shopping distribution, women are more likely than men to use SUV, Van, and car as a mean of transportation. The trend of men using pickup trucks also repeats in this chart, and the golf cart/Segway category appears with few women (0.14%) using it for shopping.

5. Time of Day Distribution

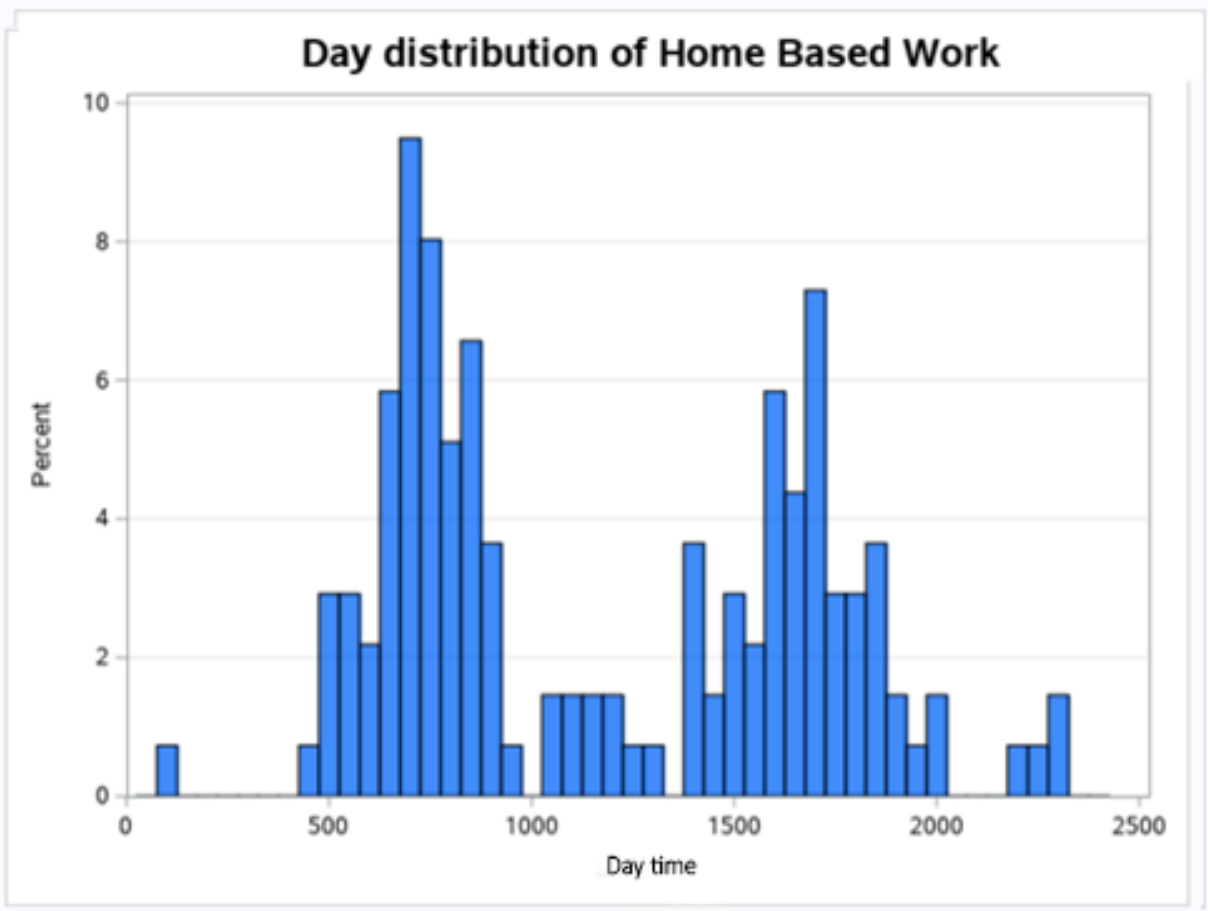

FIGURE 20. DAY DISTRIBUTION OF HBW

Figure 20. This figure refers to day time distribution of trips related to home-based work. The data was divided into 48 equal 30 minutes intervals represented on the x-axis, each 30 minute-interval correspond to 50 on the x-axis, 1050 is 10:00, while the y-axis contains the frequency distribution.

This histogram shows that people start leaving for work at 5 am and there is a huge peak at 7 am, which is the time that is considered the rush hour as most people are on their cars in direction to work. This histogram is bimodal as there is also another peak at 7 pm which is the time that people left work to go back home.

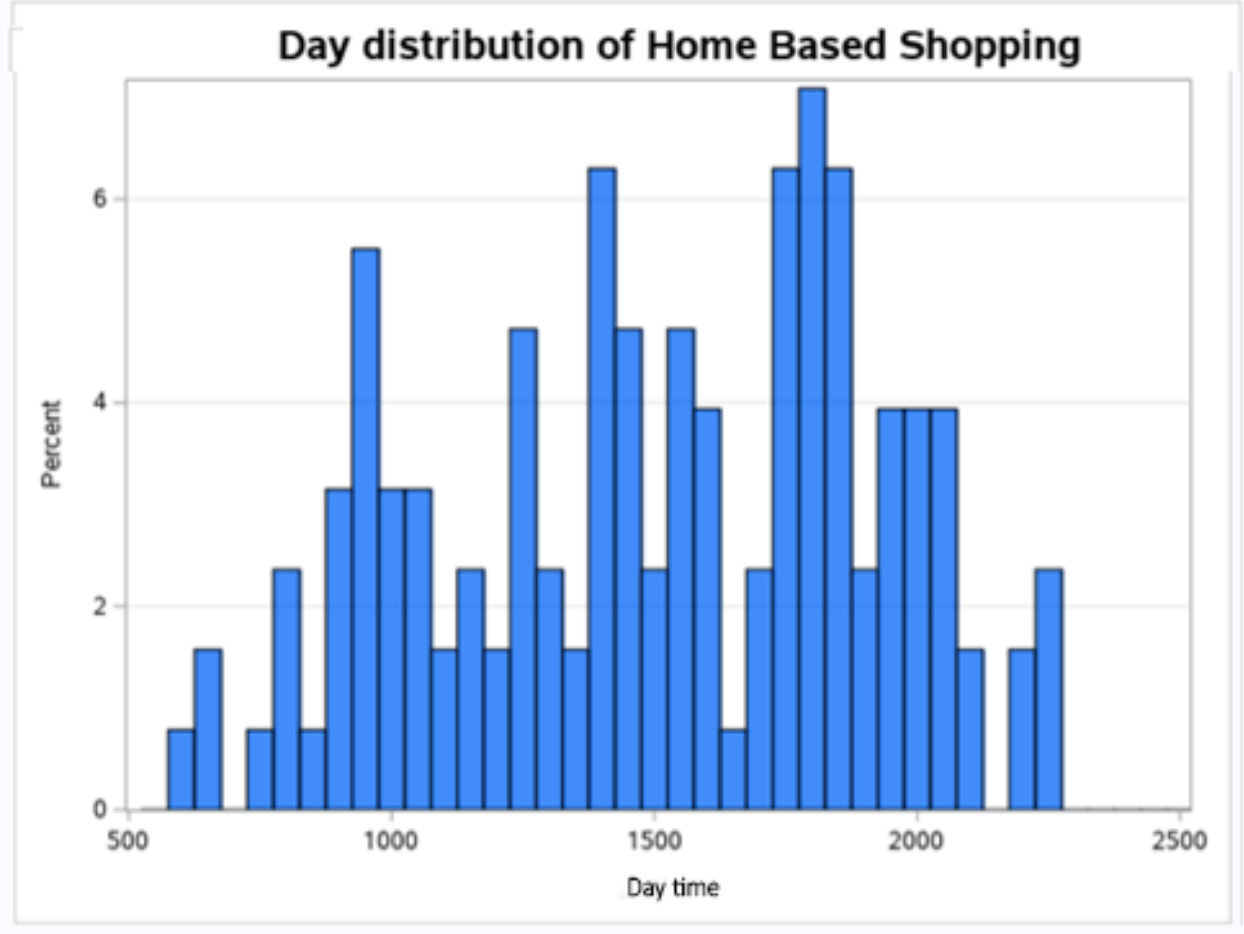

FIGURE 21. DAY DISTRIBUTION OF HBSHOP

Figure 21. This figure refers to day time distribution of trips related to home-based shopping. The data was divided into 48 equal 30 minutes intervals represented on the x-axis, each 30 minute-interval correspond to 50 on the x-axis.

This histogram shows a trend for home-based shopping related trips happening at 9:30 am during the morning period, the biggest peak happens around 17:30 and 18:30, which is the time that people usually leave work and get home, and it is important to point out that there is an important peak between 19:30 and 20:30 which is probably also related to the time people get home after work.

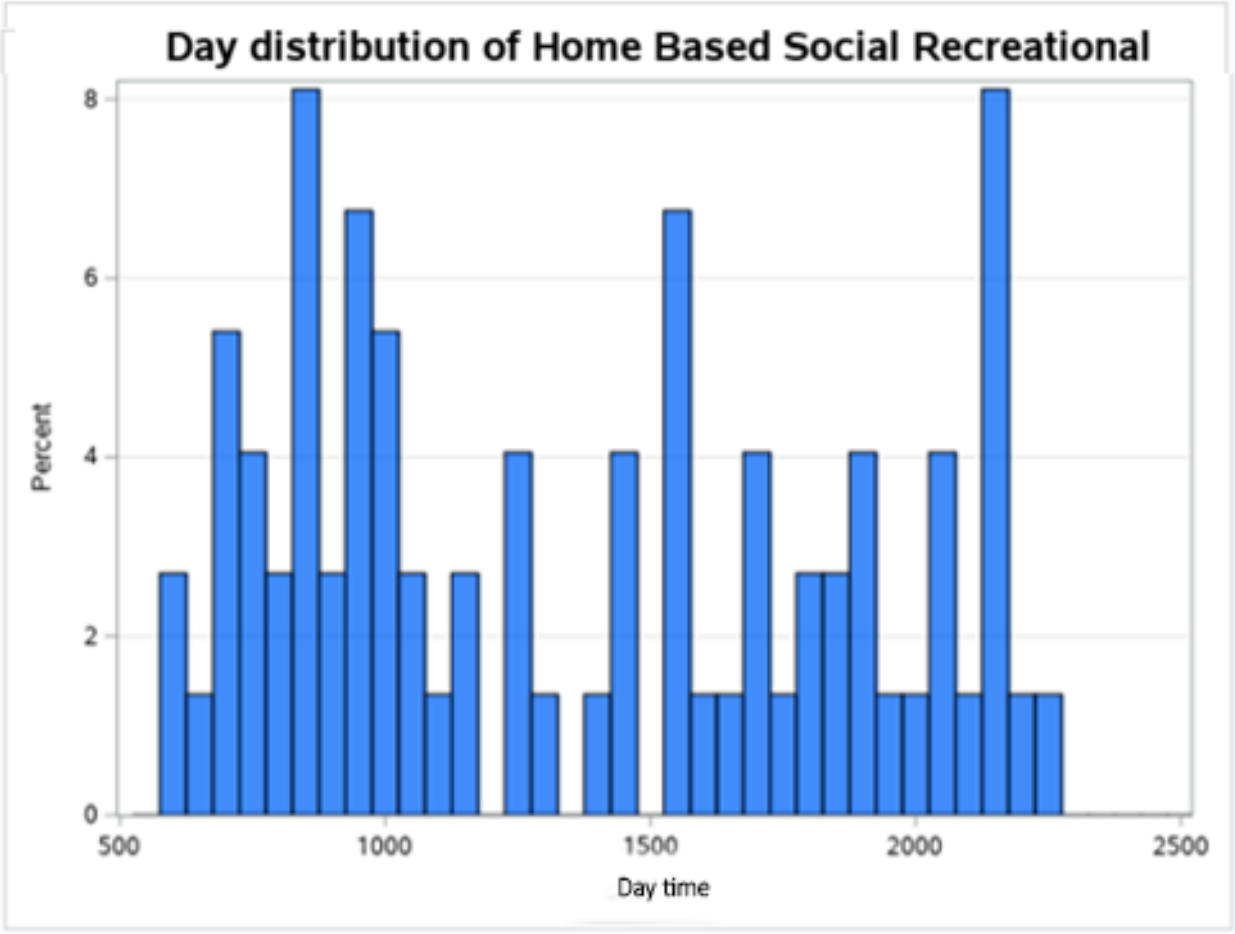

FIGURE 22. DAY DISTRIBUTION OF HBSOCREC

Figure 22. This figure refers to day time distribution of trips related to home-based social recreational. The data was divided into 48 equal 30 minutes intervals represented on the x-axis, each 30 minute-interval correspond to 50 on the x-axis.

This histogram shows a trend for home-based social recreational trips happening between 8:30 am and 9:30 am, which is probably related to non-workers as it is just after the rush hour and there is a peak at 10:00 pm which is probably involves people who arrived from work.

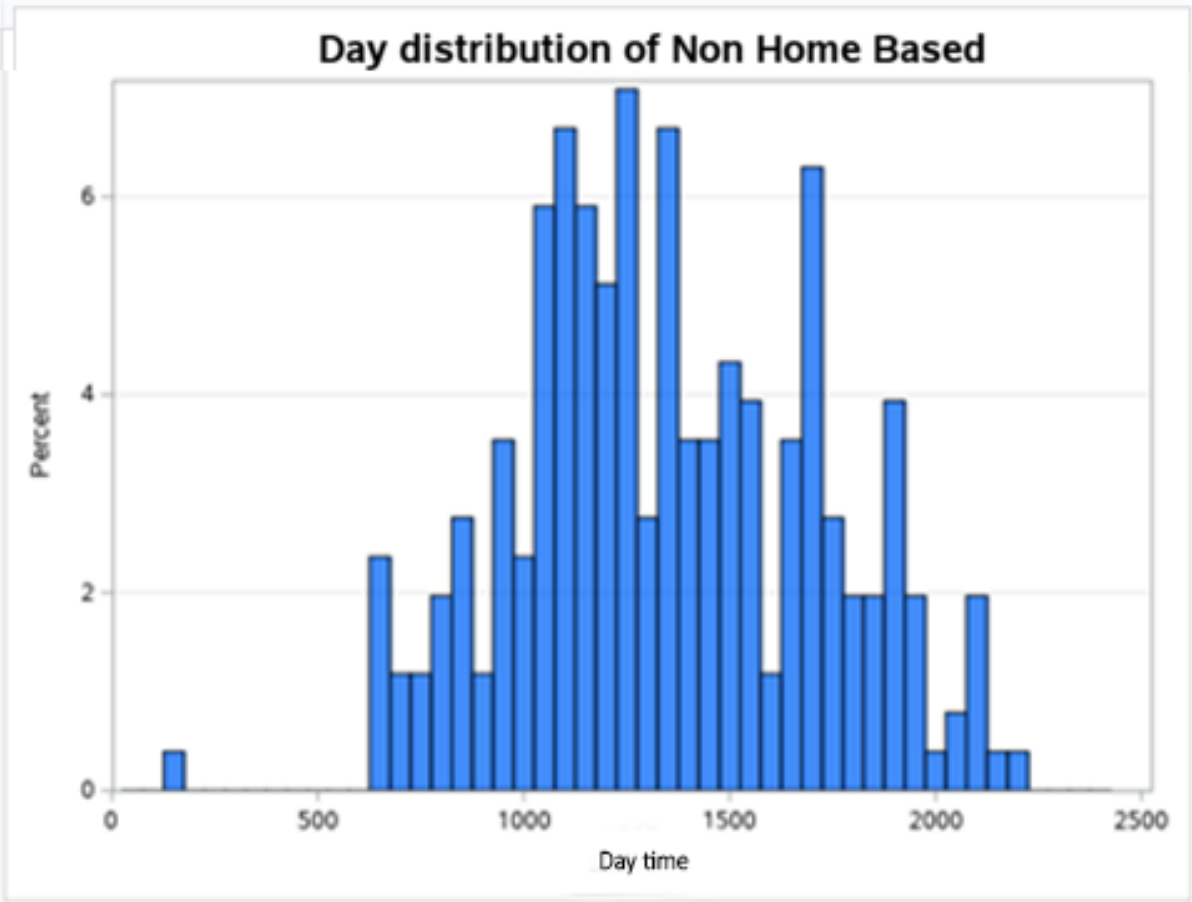

FIGURE 23. DAY DISTRIBUTION NHB

Figure 23. This figure refers to day time distribution of trips related to non-home-based activities. The data was divided into 48 equal 30 minutes intervals represented on the x-axis, each 30 minute-interval correspond to 50 on the x-axis.

This histogram shows a trend for non-home-based activities trips happening between 10:30 am and 12:30 am, which could be related lunch time, the biggest peak happens at 12:00 which is the time most people have lunch. Lastly, there is another considerable peak between 5:30 pm and 6:00 pm which could also be related to dinner or another activity non-home-based.

6. Trip Length Duration (Distance and Time)

Trip distance distribution for Home Based Work

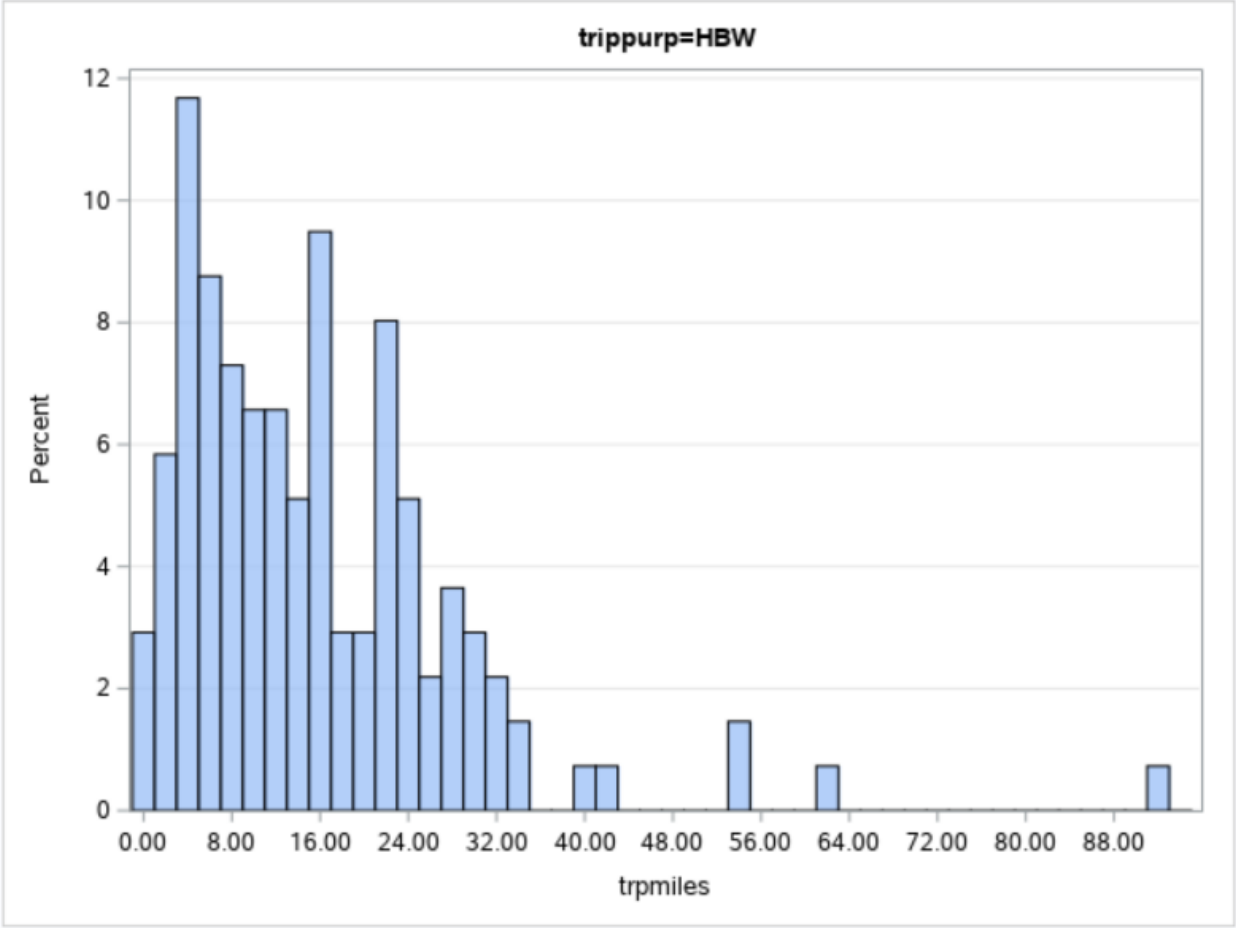

FIGURE 24. TRIP DISTANCE DISTRIBUTION (HBW)

Figure 24. This figure refers to trip distance distribution of trips related to home-based work. The data was divided into 2 miles-distance intervals represented on the x-axis and the y-axis there is the percentage frequency for each interval.

This histogram shows a trend for distance related to home-based work trips which most noticeable peaks happen at 6 miles (11.7%), 16 miles (9.4%), and 22 miles (8%). Short trips are predominant in this frequency distribution histogram, so it can infer that most people leave near their work places; the histograms' shape can be classified as a normal distribution as it is bell shaped.

Trip time distribution for Home Based Work

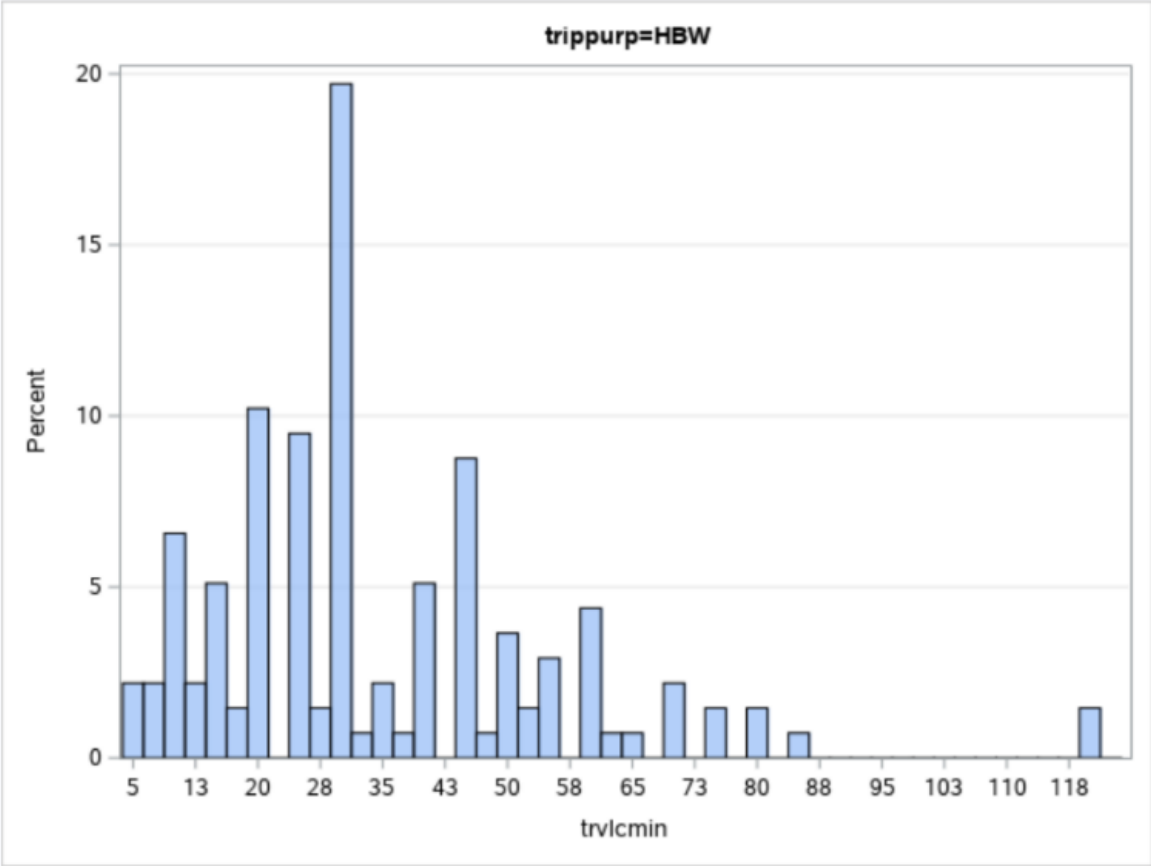

FIGURE 25. TRIP DURATION DISTRIBUTION (HBW)

Figure 25. This figure refers to trip duration distribution of trips related to home-based work. The data was divided into 2.5 minutes-duration intervals represented on the x-axis and the y-axis there is the percentage frequency for each interval.

This histogram shows a trend for trip duration related to home-based work trips which most noticeable peaks happens at 30 minutes. Although short trips were predominant on the trip duration distribution, if the previous data is compared to this histogram, it can infer that they take still take a considerable time, which is probably due to the rush hour and traffic. The histograms' shape can be classified as a normal distribution as it is bell shaped and it is unimodal.

Trip distance distribution for All trips

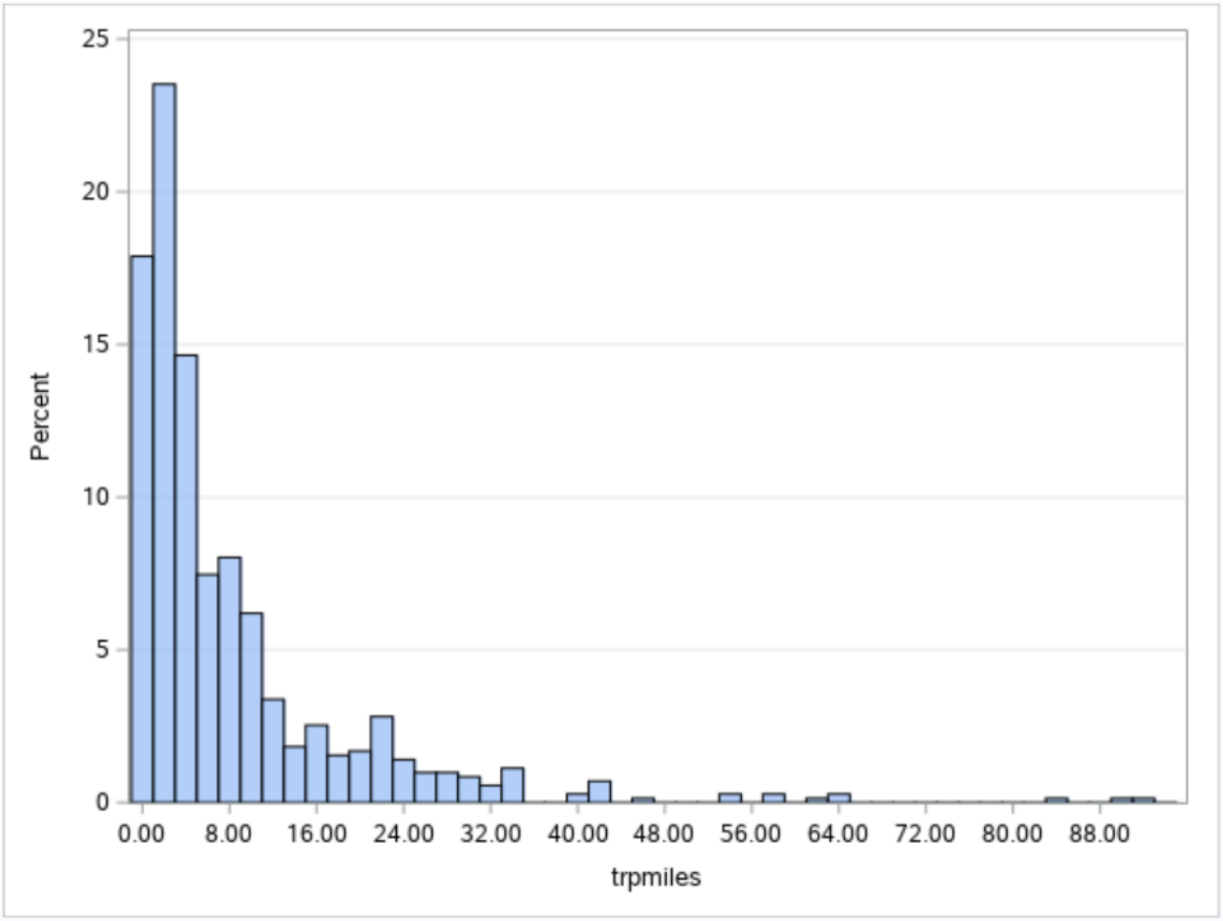

FIGURE 26. TRIP DISTANCE DISTRIBUTION FOR ALL TRIPS

Figure 26. This figure refers to trip distance distribution of all trips. The data was divided into 2 miles-distance intervals represented on the x-axis and the y-axis there is the percentage frequency for each interval.

This histogram shows a trend for trip distance which most noticeable peaks happens between 2 and 4 miles. Short trips were predominant on the trip distance distribution, it can be inferred that with the inclusion of non-vehicle trips such as walk and bicycle the short trips relative frequency has grown. Furthermore, this histogram is strongly positive-skewed so it is not bell-shaped and the reason for it is the growth of small trips as already stated above.

Trip duration distribution for All trips

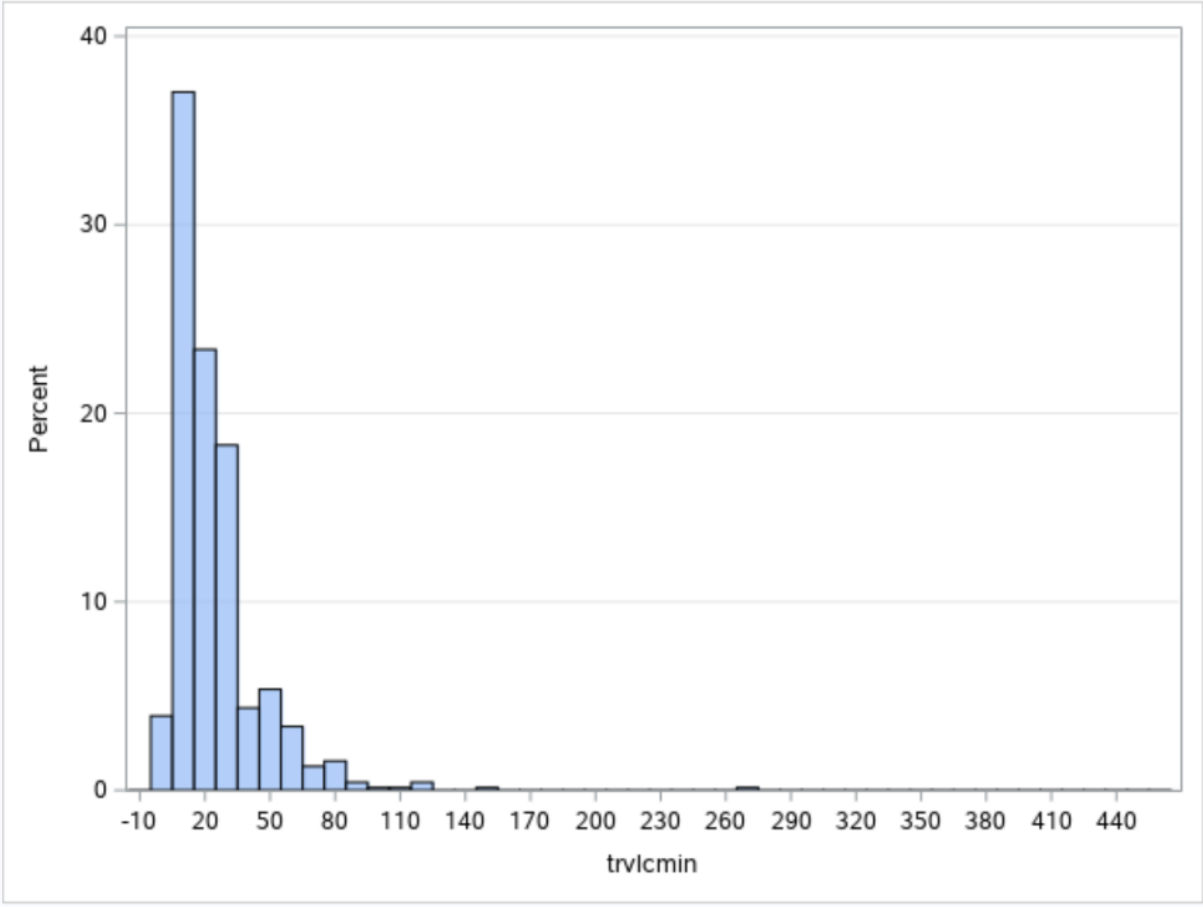

FIGURE 27. TRIP DURATION DISTRIBUTION FOR ALL TRIPS

Figure 27. This figure refers to trip duration distribution of all trips. The data was divided into 10 minutes-duration intervals represented on the x-axis and the y-axis there is the percentage frequency for each interval.

This histogram shows a trend for trip duration which most noticeable peaks happens between 0 and 10 minutes. Short trips were predominant on the trip duration distribution, the reason is the same as for trip distance distribution for All Trips, the inclusion of non-vehicle trips such as walk and bicycle the short trips relative frequency has grown. Lastly, this histogram is strongly positive-skewed so it is not bell-shaped and the reason for it is the growth of small trips as already stated above.

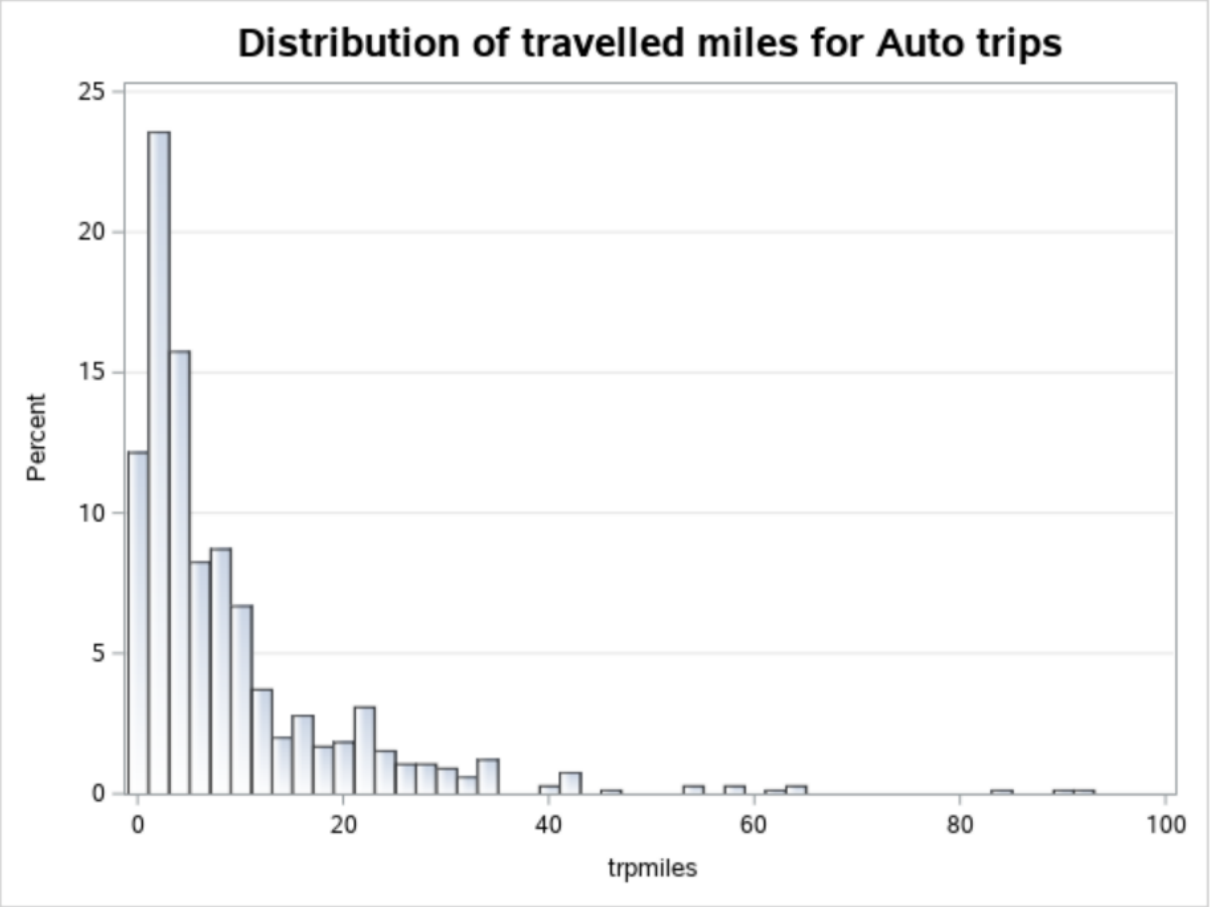

FIGURE 28. TRIP DISTANCE DISTRIBUTION FOR AUTO TRIPS

Figure 28. This figure refers to trip distance distribution of auto trips. The data was divided into 2 miles-distance intervals represented on the x-axis and the y-axis there is the percentage frequency for each interval.

This histogram shows a trend for trip distance which most noticeable peaks happens between 2 and 4 miles and it is also noticeable that this histogram is really like the All trips distribution, the most noticeable difference are the slightly smaller relative frequencies for shorter trips, the reason is the exclusion of bicycle and walk trips, which are usually short. Nevertheless, short trips were still predominant on the trip distance distribution and it can be inferred that even though the distance of the trip is small, vehicles are. Furthermore, this histogram is strongly positive-skewed so it is not bell-shaped and the reason for it is the growth of small trips as already stated above.

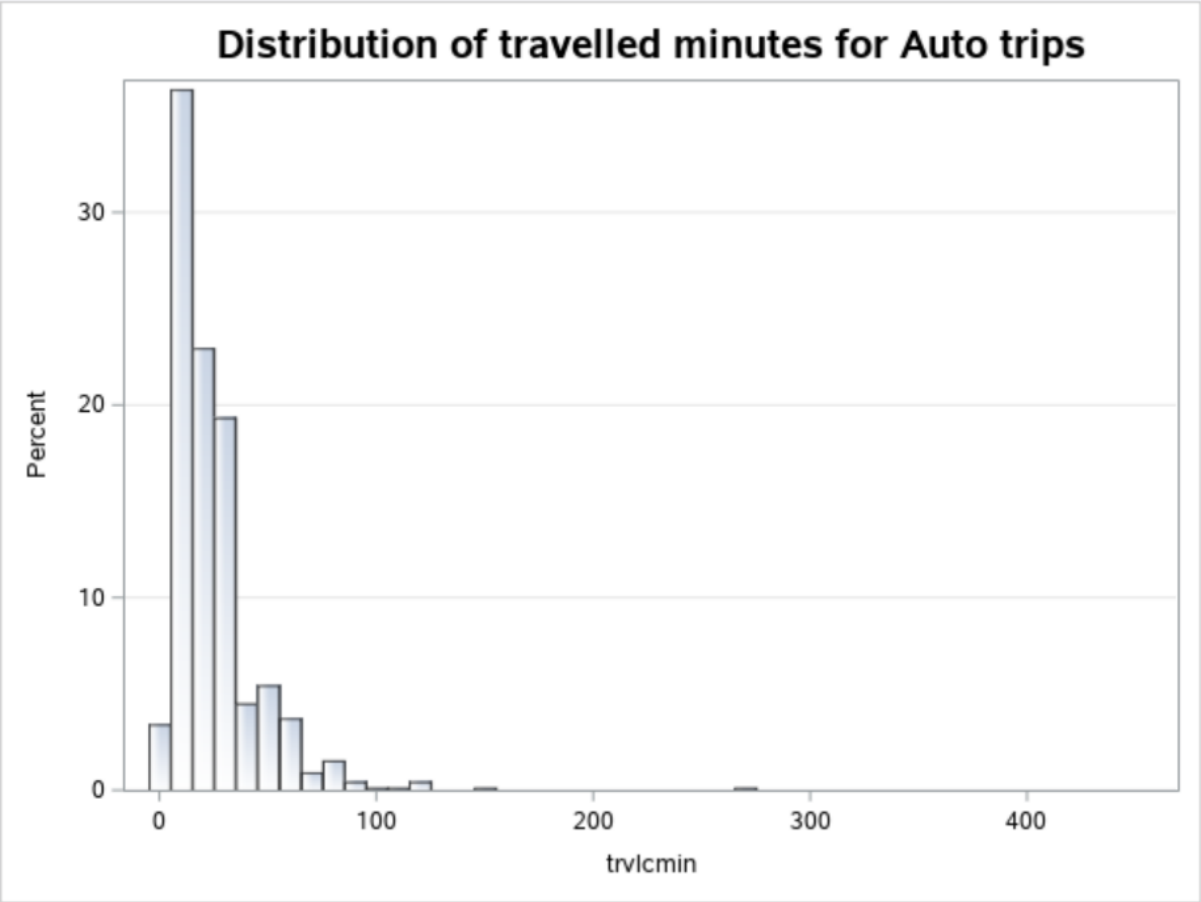

FIGURE 29. TRIP DURATION DISTRIBUTION FOR AUTO TRIPS

Figure 29. This figure refers to trip duration distribution of auto trips. The data was divided into 10 minutes-duration intervals represented on the x-axis and the y-axis there is the percentage frequency for each interval.

This histogram shows a trend for trip duration which most noticeable peaks happens between 5 and 15 minutes and it is also noticeable that this histogram is really like the All trips distribution, the most noticeable difference are the smaller relative frequencies for shorter trips, the 0 to 20 minutes interval is the region which this small decrease in relative frequency happens, the reason is the exclusion of bicycle and walk trips, which are usually short. Nevertheless, short trips were still predominant on the trip duration distribution and it can be inferred that even though the distance of the trip is small, vehicles are preferable. This histogram is strongly positive-skewed so it is not bell-shaped and the reason for it is the growth of small trips as already stated above.

Trip distance distribution for Transit trips (bus and rail)

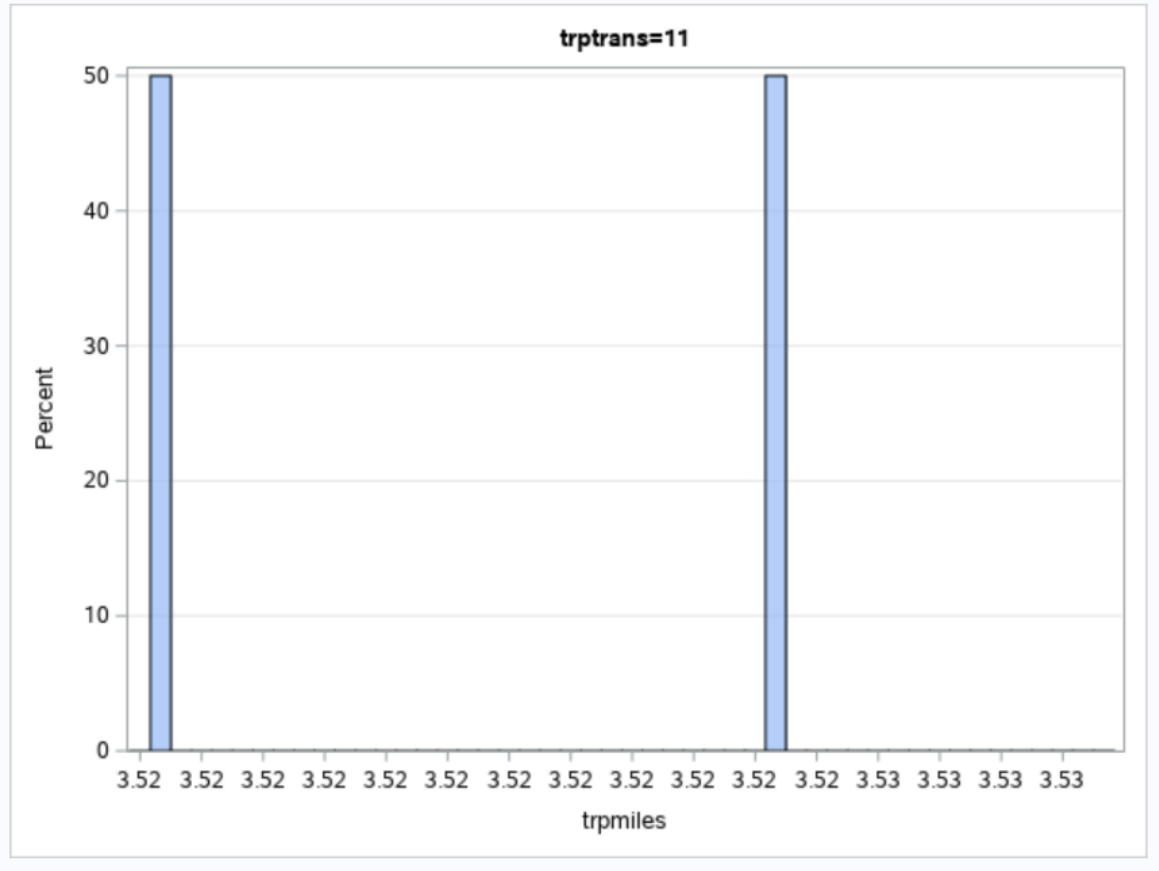

FIGURE 30. TRIP DISTANCE DISTRIBUTION FOR TRANSIT TRIPS

Figure 30. This figure refers to trip distance distribution for Transit trips. The data was divided into 0.1 miles-distance intervals represented on the x-axis and the y-axis there is the percentage frequency for each interval.

This histogram shows a different pattern which can be described by two different peaks, both with 50% of relative frequency, and the reason for these two peaks is that transit trips occur in a fixed path, so the trips will always have the same distance. Lastly, the reason for two different peaks is explained by the necessity of a path to go and another to return.

Trip duration distribution for Transit trips (bus and rail)

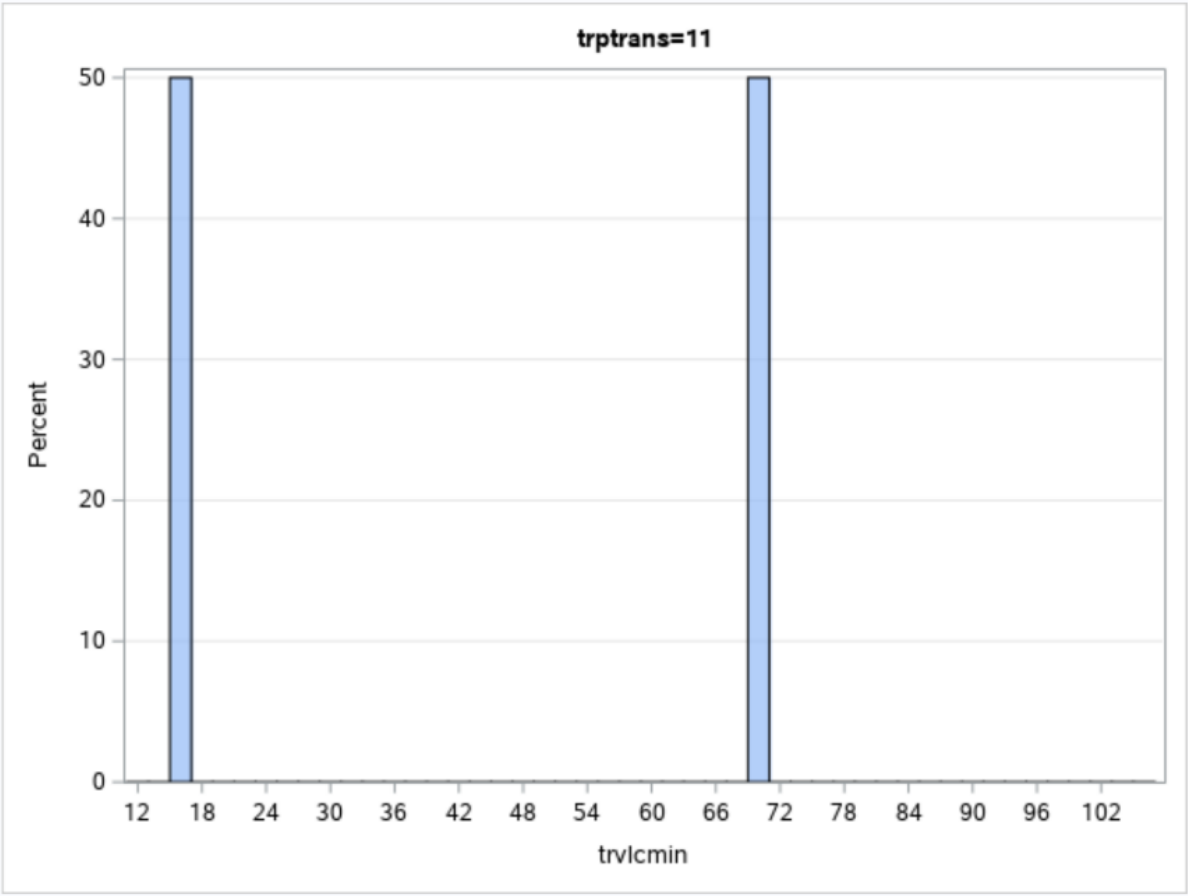

FIGURE 31. TRIP DURATION DISTRIBUTION FOR TRANSIT TRIPS

Figure 31. This figure refers to trip duration distribution for Transit trips. The data was divided into 3 minutes-duration intervals represented on the x-axis and the y-axis there is the percentage frequency for each interval.

This histogram shows a similar pattern to the last histogram which can be described by two different peaks, one at 15 minutes and other at 68 minutes, both with 50% of relative frequency, and the reason for these two peaks is that transit trips occur in a fixed path and the different times of the day that people went to work and returned home, in one probably there was not traffic as there is a peak at 15 minutes and on the other probably did as it took 69 minutes to travel the same distance.

7. Statistical Significance or Hypothesis Testing

Two-Sample T-Test and CI: Amount of Trips, Sex

Method

μ_1: mean of Amount of Trips when Sex = Male (1)
μ_2: mean of Amount of Trips when Sex = Female (2)
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: Amount of Trips

Sex	N	Mean	StDev	SE Mean
1	80	4.28	2.39	0.27
2	83	4.43	2.60	0.29

Estimation for Difference

Difference	95% CI for Difference
-0.159	(-0.930, 0.613)

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-0.41	160	0.685

TABLE 3. T-TEST (NUMBER OF TRIPS, SEX)

T-test with unequal variances to test if statistically there is a difference on the amount trips done by men and women.

As the P-value equals 0.685 we retain the null hypothesis, which states that the difference of mean number of trips for men and women is zero. The p-value is more than 0.05 which is the significance level for 95% confidence, this way, there is not enough evidence to support that there is a difference on the number of trips done for men and women.

Two-Sample T-Test and CI: Travel Time, Work Status

Method

μ_1: mean of Travel Time when Work Status = Worker (1)
μ_2: mean of Travel Time when Work Status = Non-worker (2)
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: Travel Time

Work Status	N	Mean	StDev	SE Mean
1	112	96.2	62.1	5.9
2	51	101.5	75.2	11

Estimation for Difference

Difference	95% CI for Difference
-5.2	(-29.2, 18.8)

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-0.43	82	0.665

TABLE 4. T-TEST (TRAVEL TIME, WORK STATUS)

T-test with unequal variances to test if statistically there is a difference on the mean travelled time of workers and non-workers.

As the P-value equals 0.665 we retain the null hypothesis, which states that the difference of mean on the travelled time of workers and non-workers is zero. The p-value is more than 0.05 which is the significance level for 95% confidence, this way, there is not enough evidence to support that there is a difference on the travelled time of workers and non-workers.

Two-Sample T-Test and CI: Travelled Distance, Age

Method

μ_1: mean of Travelled Distance when Age = More or equal to 50 (1)
μ_2: mean of Travelled Distance when Age = Less than 50 (2)
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: Travelled Distance

Age	N	Mean	StDev	SE Mean
1	92	36.8	32.0	3.3
2	71	34.8	27.0	3.2

Estimation for Difference

Difference	95% CI for Difference
2.02	(-7.12, 11.17)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
0.44	159	0.663

TABLE 5. T-TEST (TRAVELLED DISTANCE, AGE)

Table 5. T-test with unequal variances to test if statistically there is a difference on the mean travelled distance when age is 50 and less than 50.

As the P-value equals 0.663 we retain the null hypothesis, which states that the difference of mean on the travelled time of those who are 50 or older and younger than it is zero. The p-value is more than 0.05 which is the significance level for 95% confidence, this way, there is not enough evidence to support that there is a difference on the mean travelled distance when age is 50 and less than 50.

8. Conclusion

The objective of this study is to understand the travelling behavior of respondents from Orlando-Kissimmee-Sanford region and to test hypotheses regarding their travelling behavior. In short, travelling behavior mostly vary according trip purpose, trip mode of transportation, and time of the day.

Characteristics such as sex, work status, and age were tested using paired T-Test with unequal variances to check if there was any difference regarding number of trips, travelled time, and travelled distance. For all the three tests mentioned above, there was not enough evidence to support the hypotheses that number of trips varied on sex, travelled time varied on work status, and that travelled distance varied on age.

In conclusion, the three more important variables to consider while analyzing the travelling behavior of respondents were trip purpose, trip mode of transportation, and time of the day.

9. Codes:

Question 1:

Bar chart codes:

```
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.IMPORT;
        title height=14pt "Trip purpose";
        vbar trippurp / stat=percent;
        yaxis grid;
run;

ods graphics / reset;
title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Adults members in household";
19         vbar numadlt / stat=percent;
20         yaxis grid;
21 run;
22
23 ods graphics / reset;
24 title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Trip day of week";
19         vbar travday / stat=percent;
20         yaxis grid;
21 run;
22
23 ods graphics / reset;
24 title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Home ownership";
19         vbar homeown / stat=percent;
20         yaxis grid;
21 run;
22
23 ods graphics / reset;
24 title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Household Income";
19         vbar hhfaminc / stat=percent;
20         yaxis grid;
21 run;
22
23 ods graphics / reset;
24 title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Educational Attainment";
19         vbar educ / stat=percent;
20         yaxis grid;
21 run;
22
23 ods graphics / reset;
24 title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Trip mean of transportation";
19         vbar trptrans / stat=percent;
20         yaxis grid;
21 run;
22
23 ods graphics / reset;
24 title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Frequency distribution of Age";
19         histogram r_age_first / nbins=20;
20         xaxis min=18 max=100;
21         yaxis grid;
22 run;
23
24 ods graphics / reset;
25 title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Race of respondents";
19         vbar hh_race / stat=percent;
20         yaxis grid;
21 run;
22
23 ods graphics / reset;
24 title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Amount of trips between 6AM and 9AM";
19         vbar trpcnt_AM_sum / stat=percent;
20         yaxis grid;
21 run;
22
23 ods graphics / reset;
24 title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Amount of trips between 4PM and 7PM";
19         vbar trpcnt_PM_sum / stat=percent;
20         yaxis grid;
21 run;
22
23 ods graphics / reset;
24 title;
```

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sgplot data=WORK.IMPORT;
18         title height=14pt "Gender";
19         vbar r_sex_first_1 / stat=percent;
20         yaxis grid;
21 run;
22
23 ods graphics / reset;
24 title;
```

```
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.IMPORT;
        title height=14pt "Work status";
        vbar worker_first_1 / stat=percent;
        yaxis grid;
run;

ods graphics / reset;
title;
```

```
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.IMPORT;
        title height=14pt "Driver status";
        vbar driver_first_1 / stat=percent;
        yaxis grid;
run;

ods graphics / reset;
title;
```

Question 2. Trip purpose (Male x Female)

```
proc sgplot data=WORK.CONDENSED;
    vbar trippurp / group=r_sex groupdisplay=cluster datalabel stat=percent;
    yaxis grid;
run;

ods graphics / reset;
```

Worker vs Non-worker bar graph:

```
proc sgplot data=WORK.CONDENSED;
    vbar trippurp / group=worker groupdisplay=cluster datalabel stat=percent;
    yaxis grid;
run;

ods graphics / reset;
```

Question 3:

Transportation mode (Men Vs Women) graph:

```
proc sgplot data=WORK.CONDENSED;
    vbar trptrans / group=r_sex groupdisplay=cluster datalabel stat=percent;
    yaxis grid;
run;

ods graphics / reset;
```

Transportation mode for HBW and HBSHOP(Men Vs Women) graph:

```
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sort data=WORK.CONDENSED out=_BarChartTaskData;
    by trippurp;
run;

proc sgplot data=_BarChartTaskData;
    by trippurp;
    vbar trptrans / group=r_sex groupdisplay=cluster datalabel stat=percent;
    yaxis grid;
run;

ods graphics / reset;

proc datasets library=WORK noprint;
    delete _BarChartTaskData;
    run;
```

Question 4:

Time day (30 min interval) histogram:

```
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sort data=WORK.CONDENSED out=_HistogramTaskData;
    by trippurp;
run;

proc sgplot data=_HistogramTaskData;
    by trippurp;
    histogram strttime / showbins nbins=48 fillattrs=(color=CX9abff7
        transparency=0.25);
    xaxis max=2400;
    yaxis grid;
run;

ods graphics / reset;

proc datasets library=WORK noprint;
    delete _HistogramTaskData;
    run;
```

Question 5:

Code HBW (Time)

```
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sort data=WORK.CONDENSED out=_HistogramTaskData;
    by trippurp;
run;

proc sgplot data=_HistogramTaskData;
    by trippurp;
    histogram trvlcmin / showbins nbins=48 fillattrs=(color=CX9abff7
        transparency=0.25);
    yaxis grid;
run;

ods graphics / reset;

proc datasets library=WORK noprint;
    delete _HistogramTaskData;
    run;
```

Code HWB (Distance)

```
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sort data=WORK.CONDENSED out=_HistogramTaskData;
    by trippurp;
run;

proc sgplot data=_HistogramTaskData;
    by trippurp;
    histogram trpmiles / showbins nbins=48 fillattrs=(color=CX9abff7
        transparency=0.25);
    yaxis grid;
run;

ods graphics / reset;

proc datasets library=WORK noprint;
    delete _HistogramTaskData;
    run;
```

Code All (Time)

```
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.CONDENSED;
    histogram trvlcmin / showbins nbins=48 fillattrs=(color=CX9abff7
        transparency=0.25);
    yaxis grid;
run;

ods graphics / reset;
```

Code All trips (Miles)

```
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.CONDENSED;
    histogram trpmiles / showbins nbins=48 fillattrs=(color=CX9abff7
        transparency=0.25);
    yaxis grid;
run;

ods graphics / reset;
```

Code auto trips (Time)

```
ds graphics / reset width=6.4in height=4.8in imagemap;

roc sgplot data=WORK.IMPORT (where=(trptrans=3 or trptrans=4 or trptrans=5 or
              trptrans=6));
       title height=14pt "Distribution of travelled minutes for Auto trips";
       histogram trvlcmin / nbins=48 filltype=gradient dataskin=gloss;
       yaxis grid;
un;

ds graphics / reset;
itle;
```

Code auto trips (Miles)

```
ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.IMPORT (where=(trptrans=3 or trptrans=4 o
               trptrans=6));
        title height=14pt "Distribution of travelled miles for
        histogram trpmiles / nbins=48 filltype=gradient dataski
        yaxis grid;
run;

ods graphics / reset;
title;
```

Question 6. Was done on Minitab, so it was not possible to provide a code describing the t-test for unequal variances.