Public Transportation Use by Gender in Kenya

Daniel DeFoe
ddefoe@calpoly.edu

Brooke Hanna
blhanna@calpoly.edu

Kyle Reavis
kreavis@calpoly.edu

Abstract: *Gender inequality is a vital consideration for Kenyan national transportation policy due to the different needs of men and women. Our team was contracted by the World Bank to work with data from three surveys regarding the current landscape of public transportation and attitudes towards existing policies in Kenya. Understanding the patterns in usage and experience for travelers in Kenya with respect to gender can help us understand how public transportation systems can be designed to better serve everyone in the community. Our team then investigated the differences in travel usage and experience for members of other social groups, showing that specific income groups, ages, and demographics have specific needs for their public transportation system. Based on these findings, it is clear that the Kenyan national transportation policy should take into account the specific needs of each of these groups in order to better serve their population.*

Key words: *Gender, Public Transportation, Safety, Income*

Introduction

Investigations carried out in the past have stated there is reason to believe that the use of public transportation in Kenya is different between men and women. It has been posed that there are a disproportionate number of problems facing women that prevent them from benefiting from public transportation equally to men. There are many possible explanations for these differences, such as access to education and childcare responsibilities. Traditional to the culture in Kenya, one piece of literature states, "The presence of children reduces the likelihood that a woman will work outside the home settlement, but does not have this effect for men," and "women's relative lack of education holds them back from earning higher wages, and therefore their earning potential does not justify spending money to travel to work" (Salon, Gulyani). It is apparent through cultural norms in Kenya that women and men will have different needs regarding public transportation.

Public transportation is designed from a gender neutral perspective to appeal to everyone. However, background literature suggests that in the sphere of public transportation women consistently face more problems than men. Our team wanted to respond to this hypothesis with our own analysis.

Methodology

We used a multi-step approach to explore the possible gender differences in public transportation experiences in Kenya. First, data was considered by gender to see if there were differences in usage or experience. To do this, we first made visualizations to show the possible difference between genders.

Then, we ran the proper statistical tests; either a two sample t-test or a chi-squared test. We used two sample t-tests when the variable of interest only had two possible outcomes, such as a yes or no response to a question. We ran Chi-squared tests when the variable of interest had more than two outcomes. The statistical tests showed which variables were statistically significantly different between genders.

In order for a test to be statistically significant, the p-value must be below a certain threshold value. We used a significance level of 0.05 for all analysis because that is standard practice for hypothesis testing. Then, we adjusted the significance level by dividing 0.05 by the number of tests performed on that particular dataset to account for multiple testing. This adjusted value was what the tests' p-values were compared to, in order to decide if the results were statistically significant. If a result is statistically significant it implies that the observed differences were too large to happen by random chance alone and that there may be an underlying factor associated with these differences.

After discovering evidence of these differences, which are established later in the report, we investigated the causes of said differences, such as income level or car ownership. If we know what factors are associated with a difference in transportation experience between genders, we can help the Kenyan public transportation system best accommodate the needs of both men and women.

Data Description

JICA

The Integrated Urban Development Master Plan Household Survey provided data for describing the citizens of Nairobi's travel habits. This survey was conducted from 2013 to 2014 to aid a broad urban development plan. The questions spanned several forms detailing a respondent's household information, trip information, and travel preferences. Questions were often formatted with a category name, for example "MONTHLY INCOME", followed by a list of potential responses or a blank space to be filled in. 16,797 respondents were interviewed, with 8,459 of the respondents being female and the remaining 8,338 respondents being male. Of the 985,016 households estimated to be in Nairobi city, this survey managed to interview 10,000 of the households. Another point of interest in this study were individual trips made by respondents. This survey details 18,798 of such trips made by female respondents and 19,835 trips from male respondents.

Cleaning the JICA data mostly consisted of renaming columns and filling in missing data. We altered the column names to contain underscores rather than spaces. Then there were three columns with missing values that we addressed. We dropped a field labeled "Total" because it consisted only of missing values. There were also missing values in a field labeled "OCCUPATION Others," which was a free response section where respondents would list their occupation if it was not already included in the survey. We replaced these missing values with empty strings. Finally, a field labeled "Travel Time (HH)" contained missing values that we replaced with zeroes. However, none of our analysis utilized this specific field, since the "Travel Time" field provided similar information without the issue of missing values.

BRT

The World Bank's BRT Feasibility Study was conducted in 2016 and 2017 for the Kenyan National Highway Authority to collect information from travelers to appraise the possibility of a Bus Rapid Transit (BRT) system in Nairobi. There were 6 different surveys included in the BRT study, but the only one we used was the traveller interview survey. This was in large part because it took note of the respondent's gender. The traveller interview survey collected samples at 4 main sites along Mombasa Road and Waiyaki Way. Each site was surveyed for

one weekday from 7 A.M to 9 P.M. The traveller interview survey asked questions to try to build a general idea of travel characteristics and quality of a respondent's current trip. It also aimed to give insight into why a respondent would choose to travel the way they do. The questions in the survey tracked individual trips, providing details regarding the modes of transportation used and duration of their travel. There were also multiple sections of questions concerning respondents' opinions about their travel experience. Questions included prompts to rate levels of agreement with negative statements regarding the experience, as well as asking how much extra a respondent would pay for safer or quicker travel. The BRT survey had 556 respondents, 325 males and 231 females. The fact that the BRT traveller interview data included information on gender is why it is included in our analysis. However, these results are limited for our analysis because of the study's smaller sample size and lack of focus on gender and gender differences. For these reasons, results from the BRT Feasibility Study will not be considered as strongly as results from the other data sources.

To clean the BRT data, we changed most of the column names from their original form and filled in missing data appropriately. Numeric columns with NA's were filled with 0 where a value of 0 was implied by the non-response, but many others were left as NA to not skew analysis. There were missing data for the columns where respondents were asked to rank the most important factors of travel to them, but missing data here just meant that the factor did not rank in the respondent's top 3. We just re-coded these as "Not Top Important".

PT Users

The public transport user's survey was conducted as a primary source of data for the purpose of this analysis. The survey was conducted on behalf of both the World Bank Transportation Department as well as the State Department of Transport. The goal of this study was twofold. First, to better understand the mobility and travel needs of public transit users. Second, to recommend solutions to any barriers and gaps identified, especially with regards to gender. Although the previous two studies had a large amount of data, neither survey was conducted with the sole purpose of understanding the differences that women and men face when using public transport in Kenya. This survey aimed to explore the difference between women and men's experiences and perspectives regarding public transportation, including issues regarding sexual harassment and overall safety. Four hundred randomly selected private minibus (matatu) users were interviewed along seven roadways in Kenya. Of the 399 respondents, 298 were female, 100 were male and 1 was unknown. We removed the respondent of unknown gender since this analysis was performed from a gender-based perspective.

The data cleaning for the PT-Users survey was also relatively simple. Mainly, we again renamed columns and filled in missing values. We renamed columns to be more concise, as the original names were full sentences detailing the exact question that was asked. Missing values were filled with empty strings if the columns contained text responses and zero if the response was numeric. Filling in the numeric columns with zero did not skew the data since we filtered out the zero values in the analysis.

Question 1: How do travel patterns differ between genders?

Feature Engineering and Exploratory Analysis

We started our work by developing a list of several variables that described the travel habits of residents of Nairobi. The final variables of interest for this portion of our analysis were travel duration, travel start time, travel purpose, travel frequency, preferred travel mode, and travel cost. Specifically, what we were interested in was the characteristics of survey

respondents' everyday trips: how long they were, when they were, how many there were, etc. We intended to see how these attributes differed between genders, as this could provide insight into how travel experiences differed for males and females. For example, evidence that one gender preferred a certain travel mode over the other could drive future questions about what's behind this gap in preferences. The JICA data was especially helpful in addressing such questions because one of its sections was specifically geared towards getting more detailed trip information from each respondent. The larger sample size from this survey also meant that our hypothesis tests would be able to have greater statistical power: the ability to correctly identify a difference among populations. Our findings and visualizations of these relationships are shown below.

Sub Question 1: Does the duration of travel differ by gender in Kenya ?

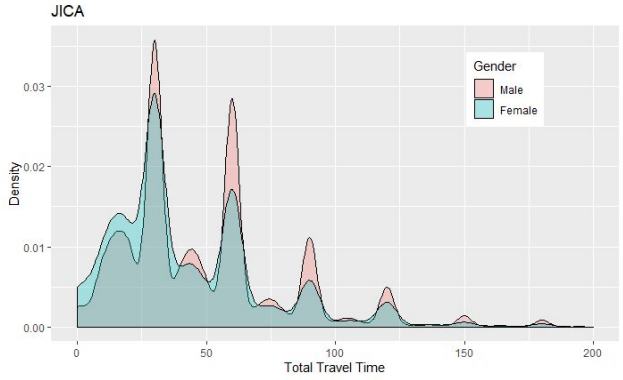

Fig. 1.1.1 *Travel Duration by Gender(JICA)*

Figure 1.1.1 shows that female respondents were more likely than male respondents to take trips that were 30 minutes or less. However, for trips that were longer than 30 minutes, male respondents consistently outnumbered female respondents.

Sub Question 2: Do travel times differ by gender in Kenya?

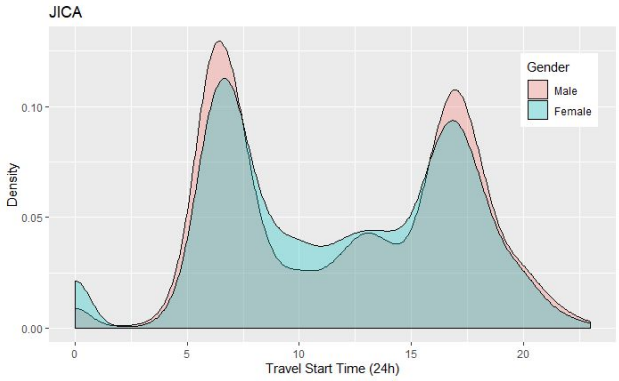

Fig. 1.2.1 *Travel Start Time by Gender(JICA)*

Figure 1.2.1 shows that males take more trips during the hours before and after a typical work day, around hours 7 and 17. Additionally, females make more trips during the day, between the hours of 8 and 16. This difference could be attributed to male respondents reporting they commute mainly to and from work while female respondents report running errands during typical work hours more often.

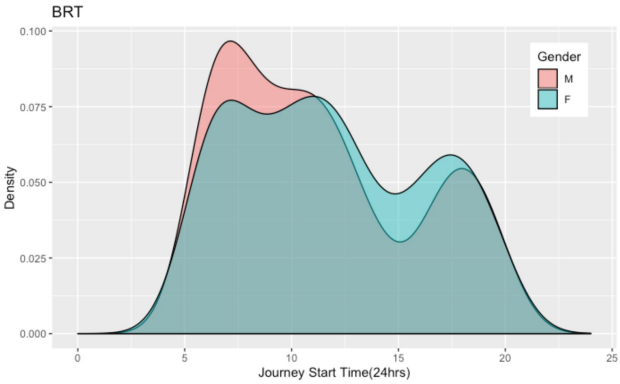

Fig. 1.2.2 *Travel Start Time by Gender(BRT)*

Results from the BRT survey in figure 1.2.2 support the findings from the JICA survey that males take more of their trips around morning commute hours, and females travel more during the afternoon.

Sub Question 3: Are there differences in travel purpose by gender in Kenya?

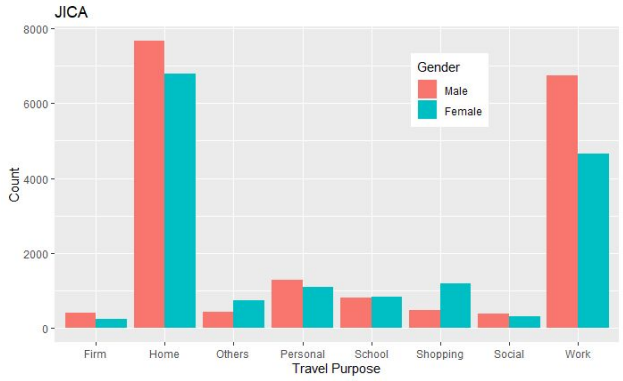

Fig. 1.3.1 *Travel Purpose by Gender(JICA)*

Figure 1.3.1 shows that more male respondents went on work-related trips than female respondents. However, female respondents went on more shopping-related trips and trips classified as 'other' on the survey than male respondents. The results for the alternate travel purposes did not show a notable difference by gender.

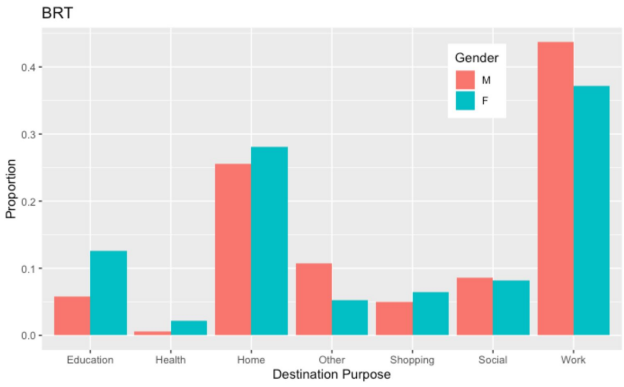

Fig. 1.3.2 *Travel Purpose by Gender(BRT)*

Similar to what figure 1.3.1 displays, figure 1.3.2 shows more males reported traveling to work than females, and that more females reported traveling for the purpose of shopping than males. However, figure 1.3.2 shows a larger difference in more females traveling for education and more males traveling for "other".

Sub Question 4: Does frequency of travel differ by gender in Kenya?

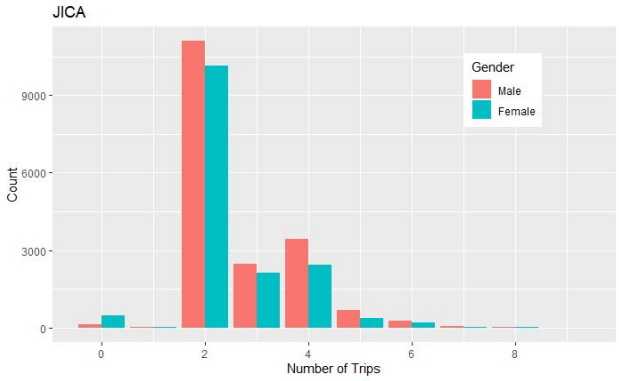

Fig. 1.4.1 *Travel Frequency by Gender(JICA)*

Figure 1.4.1 shows that the majority of both males and females report taking two trips per day. There were not large differences in the number of trips taken by males and females; however, the most notable differences are females reporting taking zero trips more often than males.

Sub Question 5: Does preferred travel mode differ by gender in Kenya?

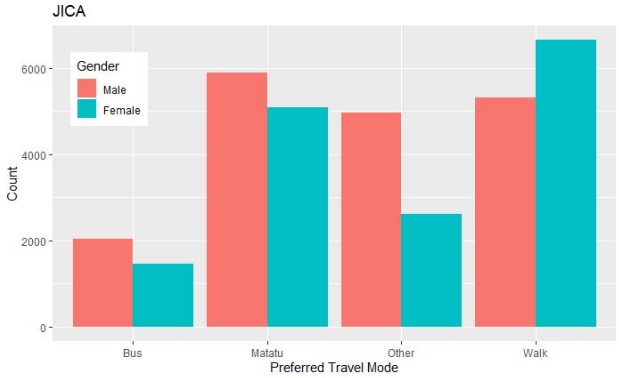

Fig. 1.5.1. *Preferred Travel Mode by Gender(JICA)*

Figure 1.5.1 shows that more female respondents preferred to travel by walking than male respondents. The opposite was true for the bus, matutu, and "other" categories, wherein more male respondents preferred these travel modes than female respondents. Male respondents significantly preferred travel methods in the "other" category more often than female respondents.

Sub Question 6: Does travel cost differ by gender in Kenya?

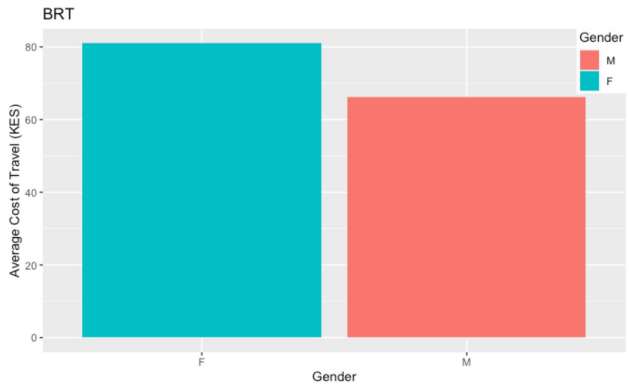

Fig. 1.6.1. *Travel Cost by Gender(BRT)*

Figure 1.6.1 shows that on average, for all recorded travel journeys, including walking trips, female respondents in the BRT survey paid an average of 15 Kenyan Shillings more than male respondents. This is especially noteworthy considering that Figure 1.6.1 showed more female respondents were walking than male respondents in the JICA survey. The number of male respondents who reported walking in in the BRT survey was actually greater than the number of female respondents, but when walking trips were excluded, a difference of about 15 Kenyan Shillings between male respondents and female respondents's average travel costs still existed.

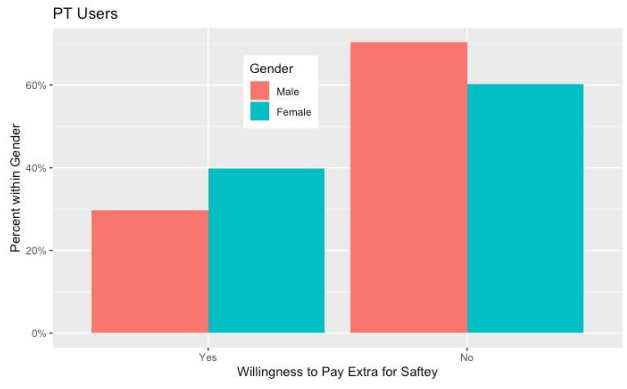

Fig. 1.6.2 *Willingness to Pay Extra by Gender(PT)*

Figure 1.6.2 addresses the question of respondents willingness to pay an extra fare for a safer public transportation experience. Females respondents were more likely to respond favorably to this hypothetical proposition than male respondents.

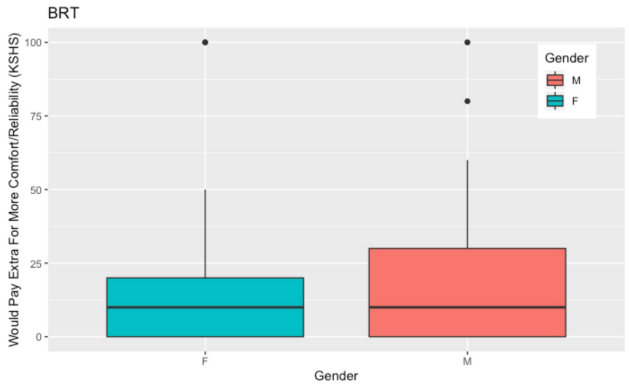

Fig. 1.6.3 *Willingness to Pay Extra For Comfort/Reliability by Gender(BRT)*

Figure 1.6.3 portrays that male respondents are seemingly willing to pay slightly more for increased comfort and reliability than female respondents are. However, the medians are roughly equal, so in general the graph is communicating that male respondents who are willing to pay more will respond with greater amounts that they would pay, as compared to female respondents.

Sub Question 7: Does concern over sexual harassment on public transportation diiffer by gender in Kenya?

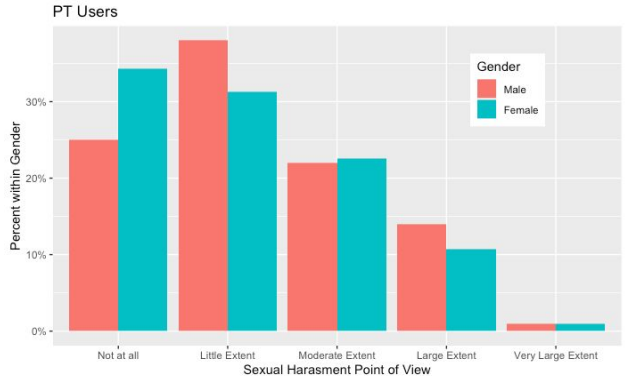

Fig. 1.7.1. *Sexual Harassment Point of View by Gender(PT)*

Figure 1.7.1 shows how respondents answered a question asking if they were satisfied with the government's efforts to address incidents of sexual harassment on public transportation. When compared to male respondents, female respondents

were more likely to be unsatisfied with the government's efforts.

Hypothesis Testing

To verify that the differences we were observing held statistical significance, it was important to our team that we ran proper tests. When we tested for the differences we observed with the data from the JICA survey, the results were very promising. Differences between genders in travel duration, start time, and frequency were assessed using a two sample t-test. We used a chi-squared test to examine the differences in trip purpose and preferred travel mode across genders. We used an adjusted p-value of .01 as our threshold, to adjust for performing multiple tests on the same data, and we found very strong evidence that travel duration, start time, purpose, preferred mode, and frequency of use were different between genders at the .05 significance level.

Variable	**P-Value**
Travel Duration	< 0.001
Start Times of Travel	< 0.001
Frequency of Use	< 0.001
Trip Purpose	< 0.001
Preferred Travel Mode	< 0.001

The following results come from the PT User survey. Again, we used two sample t-tests and chi-squared tests to test if there were differences in the following variables between males and females. Here we used an adjusted p-value of 0.0083 to correct for multiple testing. At the 0.05 significance level, the variables that were statistically significant were those that asked respondents if they had witnessed sexual harassment, indecent remarks, or unwanted touching. The remaining variables did not result in a statistically significant difference between genders. It is important to note that the PT Users survey had a sample size of 399 so it has less statistical power in detecting differences as compared to the JICA survey, which had a sample size of 16,797. This may be an explanation of why there were fewer variables detected as significant than the JICA survey.

Variable	**P-Value**
Travel Cost	0.7626
Willingness to Pay Extra for Safety	0.1661
Sexual Harassment Addressment Satisfaction	0.4427
Road Safety Satisfaction	0.9334
Security Satisfaction	0.7563
Witness Sexual Harassment	0.006073
Indecent Remarks Observed	< 0.001
Unwanted Touching Observed	< 0.001
Rape Observed	0.08849

Summary

This section provided strong evidence of differences in transportation use between genders in Kenya. There is an especially large difference in the experience of men and women when looking at which travel mode is preferred, with women preferring walking more than any other option and men preferring transportation methods other than walking, such as the bus or matutu. There are also some large differences in what time these trips

would start and how long the trips would last, which tended to revolve more heavily around standard working hours for men. This aligns with the fact that men would report traveling for work-related purposes more often than women. Now that we believe there are differences in transportation usage between genders in Kenya, we will explore if differences exist when we consider the data by income group instead of gender.

Question 2: How do the Travel Patterns of Low-Income and High-Income Respondents Differ?

Feature Engineering and Exploratory Analysis

Our methodology here strongly mirrors our previous analysis of travel patterns by gender. We again used the JICA data here due to its large sample size and travel-oriented questions. For this portion, the trip-oriented variables in this dataset we wanted to utilize were almost the same exact ones as before: travel duration, travel start time, travel purpose, travel frequency, and preferred travel mode. Evidence of differences in travel patterns among low income and high income respondents indicate that income would be crucial to investigate in addition to gender in order to draw the most informed conclusion possible. We used information on the respondents' monthly income to distinguish low income respondents from high income respondents. To accomplish this, our team classified respondents with a monthly income of 20,000 Ksh or lower as low income, whereas high income respondents needed a monthly income exceeding 20,000 Kshs per month. We chose this threshold of 20,000 Kshs per month since in the PT Users survey the information regarding income was provided in ranges of 20,000 Kshs. Thus, respondents with the lowest income reported having a monthly income between 0 and 20,000 Kshs. 13,383 respondents were classified as high income respondents and 20,603 respondents were classified as low income respondents.

Sub Question 1: Does travel duration differ by income group in Kenya?

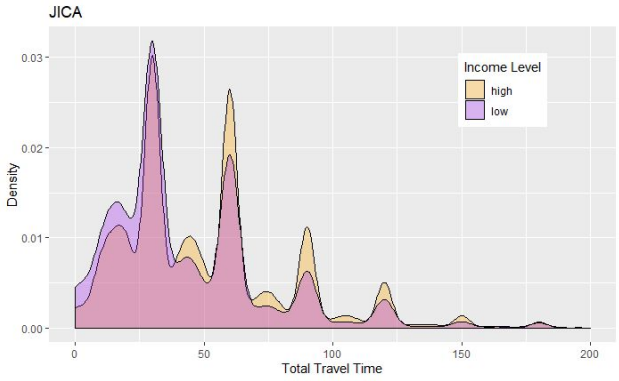

Fig. 2.1.1 *Travel Duration by Income Group*

Figure 2.1.1 shows that respondents with low incomes were more likely than respondents with high income to take trips that were 40 minutes or less. However, for trips that were longer than 40 minutes, high income respondents consistently outnumbered low income respondents.

Sub Question 2: Does travel start time differ by income group in Kenya?

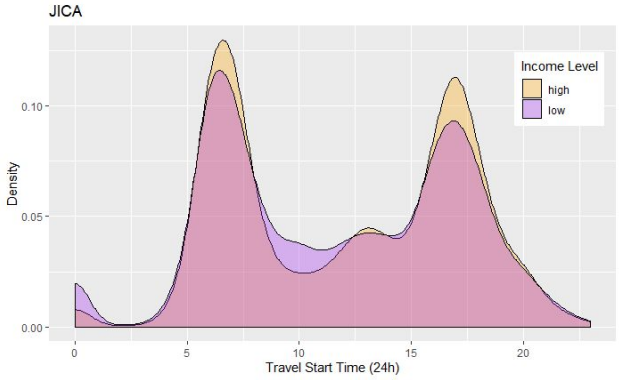

Fig. 2.2.1. *Travel Start Time by Income Group*

Figure 2.2.1 shows that high income respondents peak in travel start time around hours 8 and 17, typical start and end points of a work day. Additionally, low income respondents make more trips in between these specific hours. This resembles what we saw with male and female respondents and likely relates to high income respondents commuting to and from work more often than low income respondents.

Sub Question 3: Does travel purpose differ by income group in Kenya?

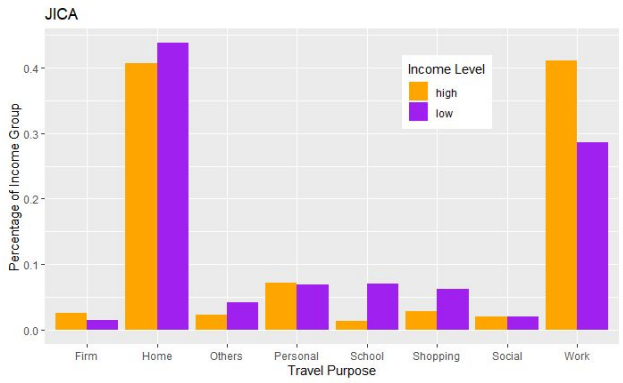

Fig. 2.3.1. *Travel Purpose by Income Group*

Figure 2.3.1 shows that low income respondents went on less work-related trips than high income respondents. However, low income respondents went on more shopping-related trips, school-related trips, and trips classified as 'other' on the survey than high income respondents. The results for the alternate travel purposes did not show a notable difference by income group.

Sub Question 4: Does use of public transportation differ by income group in Kenya?

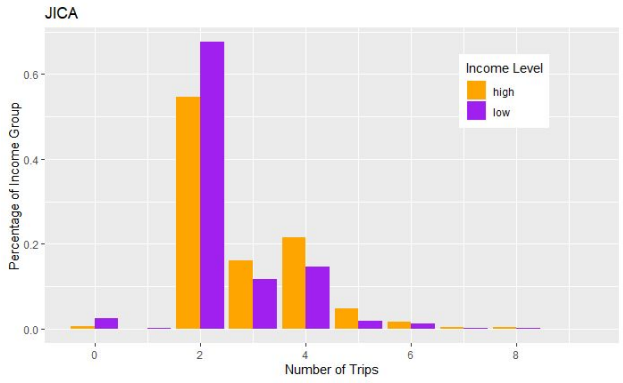

Fig. 2.4.1 *Travel Frequency by Income Group*

Figure 2.4.1 shows that the majority of both high and low income respondents take two trips per day. However, the low income group had a notably larger proportion of respondents taking two or less trips per day. It's also worth noting that, when compared to the low income respondents, high income respondents were more likely to report taking over 2 trips a day.

Sub Question 5: Does preferred travel mode differ by income group in Kenya?

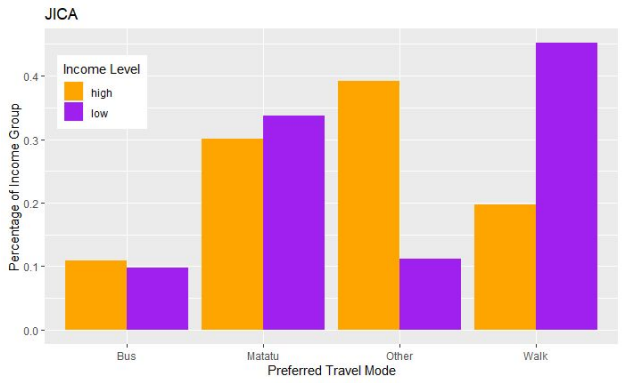

Fig. 2.5.1. *Preferred Travel Mode by Income Group*

Figure 2.5.1 shows that low income respondents overwhelmingly preferred to travel by walking in comparison to high income respondents. On the other end, high income respondents preferred the "other" category much more often than low income respondents. Both income groups preferred to travel by bus or matutu at very similar rates.

Hypothesis Testing

Similar to the analysis done for Question 1, we ran hypothesis tests to validate the differences we observed. We used an overall significance level of .05, meaning each test used an individual significance level of .002 after accounting for multiple testing. We used a chi-squared test to examine the differences in trip purpose and preferred travel mode across income groups. We assessed differences between income groups in travel duration, start time, and frequency using t-tests. Again, we used an adjusted p-value of .002 and found very strong evidence that travel duration, start time, purpose, preferred mode, and frequency was different between income groups at the .05 significance level.

Variable	P-Value
Travel Duration	< 0.001
Start Times of Travel	< 0.001
Frequency of Use	< 0.001
Trip Purpose	< 0.001
Preferred Travel Mode	< 0.001

Summary

This section provided us with strong evidence that there are existing differences in transportation use between high and low income groups in Kenya. Low income respondents were more likely to take shorter trips that were not during typical commuter hours and preferred walking as their main mode of travel, compared to high income respondents. High income respondents were much more likely to travel for longer periods of time during the typical commuter hours for the purpose of work. High income respondents in the JICA survey far preferred the "other" option for method of transportation over the bus, matatu, or walking, which could very well mean transportation by car.

Question 3: How do the Travel Patterns of Low-Income and High-Income Females Differ?

Feature Engineering and Exploratory Analysis

The goal of our next analysis was to understand how travel patterns differ between females with low incomes and high incomes. We again classified low-income females as having an income of 20,000 Kshs or less per month, while high income females had a monthly income of more than 20,000 Kshs. For this analysis, we used both the JICA survey and PT Users survey. We used both surveys when the answers to the desired research questions were available from both JICA and PT Users; such as exploring differences between low and high income women in length of trip and frequency of use. We conducted the remaining analyses with only the PT Users survey due to its inclusion of questions regarding cost, safety and sexual harassment; which are not present in the JICA survey. There were 214 females in the survey that had information on their income; 76 in the low-income group, and 138 in the high-income group.

Sub Question 1: Does travel duration differ between low-income and high-income females Kenya?

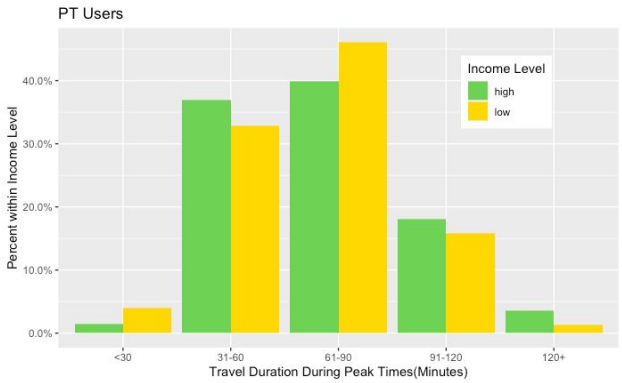

Fig. 3.1.1. *Trip Length by Income Level, Peak Times (Female Respondents)*

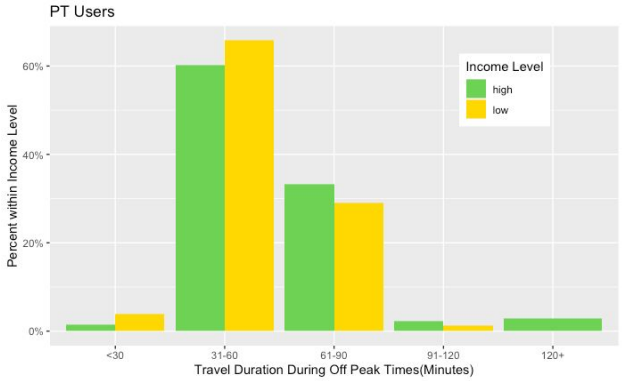

Fig. 3.1.2. *Trip Length by Income Level, Off Peak Times (Female Respondents)*

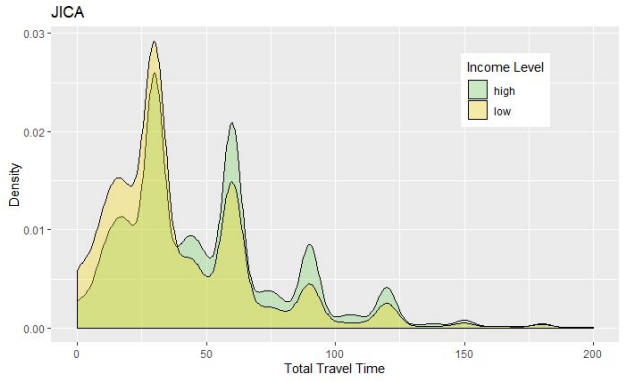

Fig. 3.1.3. *Trip Length by Income Level (Female Respondents)*

In the above figures we can see slight differences in trip duration between low and high income level females, both for peak hours and off-peak hours of transportation. Overall, the majority of trips taken by all females have a duration of 31-90 minutes. During off peak hours, a greater proportion of low income females seem to be taking trips lasting an hour or less, while greater proportion of high income females report taking longer trips. During peak hours, low income females only outnumber high income females for trips of length 61-90 minutes. High income females have a greater proportion for all other trip lengths. This could mean that if low income females travel during peak hours, they generally have to travel further for work.

Sub Question 2: Does frequency of public transportation use differ between low-income and high-income females in Kenya?

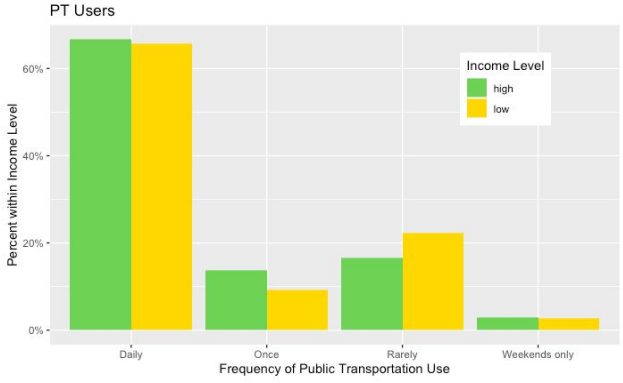

Fig. 3.2.1. *Frequency of Public Transportation Use by Income Level (Female Respondents)*

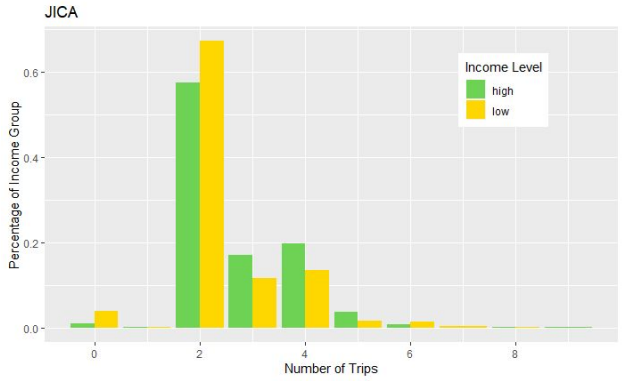

Fig. 3.2.2 *Frequency of Public Transportation Use by Income Level (Female Respondents)*

Figure 3.2.1 shows that there are only small differences in how often females use public transportation, dependent on income level. Additionally, these differences are small and do not show a consistent pattern in which group uses transportation more often. Figure 3.2.2 shows that low income females are more likely than their high income counterparts to go on 2 or less trips. However; high income females are more likely to go on 3 or more trips.

Sub Question 3: Does travel cost differ between low-income and high-income females Kenya?

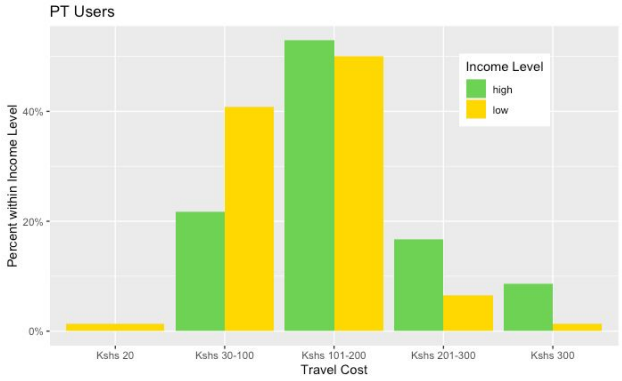

Fig. 3.3.1. *Travel Cost by Income Level (Female Respondents)*

The above figure shows a difference in cost of travel by income level of females. Low income females greatly outnumber high income females in trips with a cost of 100 Kshs or less, while high income females outnumber low income females in trips with a cost of more than 100 Kshs. It is a possibility that

low income females are not able to take trips which cost more than 100 Kshs and therefore have limited accessibility as compared to high income females.

Sub Question 4: Does travel purpose differ between low-income and high-income females Kenya?

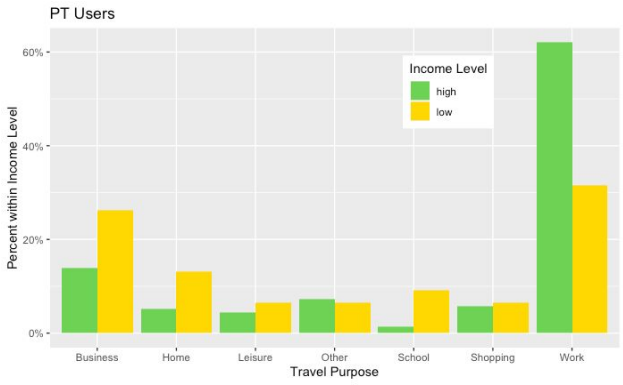

Fig. 3.4.1. *Travel Purpose by Income Level (Female Respondents)*

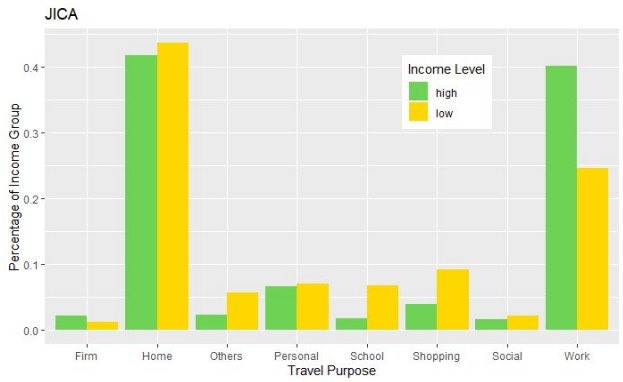

Fig. 3.4.2. *Travel Purpose by Income Level (Female Respondents)*

Figure 3.4.1 displays a large difference in travel purpose between low income and high income females. About 60% of trips taken by high income females are for work, while only about 30% of trips taken by low income females are for work. For all remaining travel purposes, low income females have a greater proportion of respondents than high income females. Figure 3.4.2 tells a similar story in which high income females, compared to low income females, are more likely to be traveling for work and less likely to be traveling for school or shopping.

Sub Question 5: Does security satisfaction differ between low-income and high-income females Kenya?

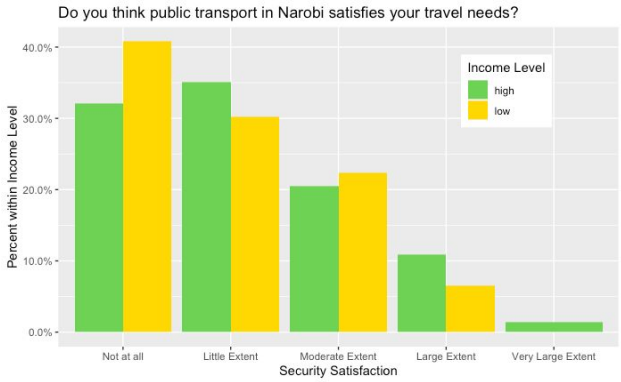

Fig. 3.5.1 *Security Satisfaction by Income Level (Female Respondents)*

Survey respondents were asked if the public transport sector satisfied their travel needs in from the point of view of road safety, security, and sexual harassment. Figure 3.5.1 shows the response to this question in regards to security. In general, low income females were less satisfied with the public transit sector. A similar trend was apparent for road safety and sexual harassment.

Sub Question 6: Does report rate of sexual harassment differ between low-income and high-income females Kenya?

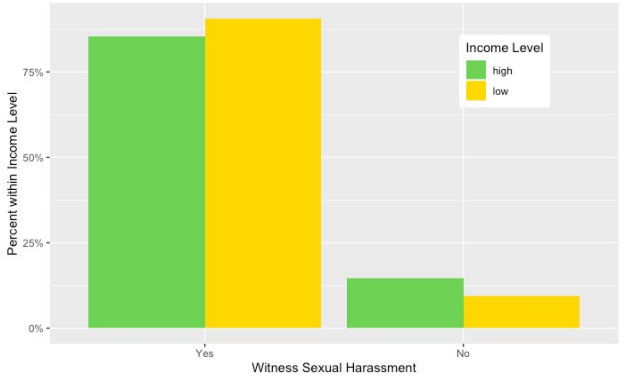

Fig. 3.6.1. *Witnessed Sexual Harassment by Income Level (Female Respondents)*

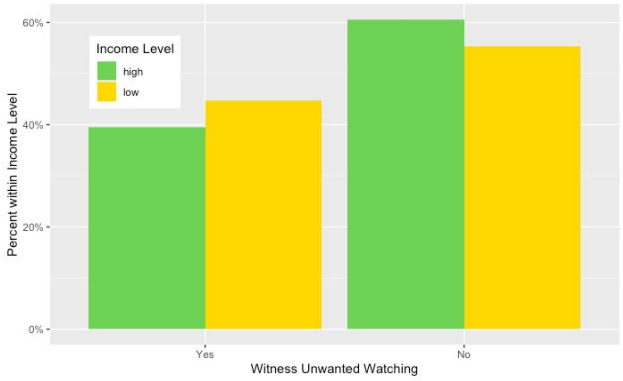

Fig. 3.6.2. *Witnessed Unwanted Touching by Income Level (Female Respondents)*

The above two figures came from a set of questions asking respondents if they had witnessed sexual harassment in general and specifically by the type of harassment. For all of these questions, low income females responded yes more often than the high income females while high income females responded no more often.

Hypothesis Testing

We carried out chi-squared tests in order to formalize the differences between low-income and high-income females. With a significance level of 0.05, we adjusted the p-value to be 0.0045 for the PT Users survey and 0.0167 for the JICA survey to account for multiple testing. Both travel duration and frequency of use were statistically significant in the JICA survey, but not for the PT users survey. This is due to the much larger sample size of the JICA survey, and thus its larger statistical power that we discussed earlier. Travel purpose was statistically significant in both surveys and travel cost was also statistically significant at the 0.05 significance level. No other variables were statistically significant in regards to the differences between low-income and high-income females.

Variable	**P-Value**
Travel Duration(Peak Times)	0.5465
Travel Duration(Off Peak Times)	0.3729
Travel Duration (JICA)	<0.001
Frequency of Use	0.6351
Frequency of Use (JICA)	<0.001
Travel Cost	0.0025
Travel Purpose	<0.001
Travel Purpose (JICA)	<0.001
Willingness to Pay Extra for Safety	0.1208
Sexual Harassment Addressment Satisfaction	0.6636
Road Safety Satisfaction	0.8047
Security Satisfaction	0.4767
Witness Sexual Harassment	0.3666
Unwanted Touching Observed	0.5417

**Data is from the PT Users Survey unless otherwise specified*

Summary

In this section we found significant differences in how the travel patterns of low income and high income females differed. Low income females typically take two trips or less that cost less than 100 Kshs; while high income females are able to take more frequent and more expensive trips. High income females are more likely to be traveling for work and low income females were more likely to be traveling for all other purposes. Low income females were less satisfied with security and road saftey while also reported witnessing every type of sexual harassment more often than high income

females. Overall, there is a large difference between low income and high income female respondents. Low income females appear to be at a disadvantage, feeling less safe on public transit. Low income females also appear to have different needs regarding transit due to their different trip purposes and their pattern of taking shorter, cheaper trips. Due to these apparent differences, the needs of both low income and high income females must be taken into account so that public transportation in Kenya satisfies the needs of all groups.

Question 4: How do the travel patterns of respondents with and without a car differ?

Feature Engineering and Exploratory Analysis

After providing evidence that travel patterns and usage on public transportation significantly differs between males and females, low-income and high-income respondents, as well as between low-income and high-income females; we wanted to investigate if these differences also existed between respondents with and without a car. All three surveys included a question asking the respondent if they owned a car, so this variable was used to create the two groups. Similar to the previous sections, we will first examine visualizations to see how specific variables differ between both males with and without a car as well as females with and without a car. Then, we will run the proper statistical test to see if the variable is statistically significantly different between those with and without a car.

Sub Question 1: Does the duration of travel differ between those with and without a car?

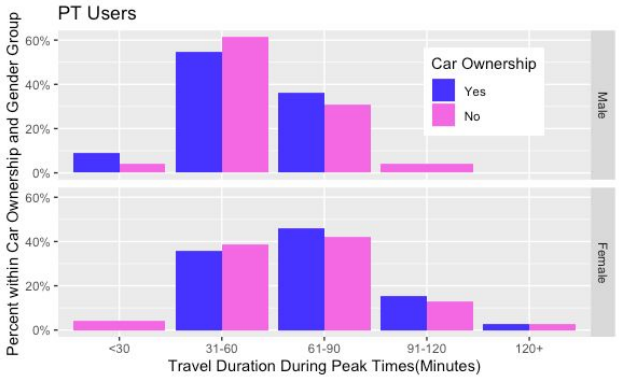

Fig. 4.1.1 *Travel Duration During Peak Times by Car Ownership (PT Users)*

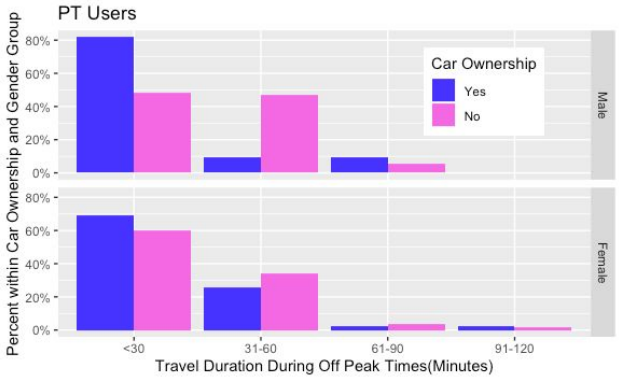

Fig. 4.1.2 *Travel Duration During Off-Peak Times by Car Ownership (PT Users)*

The graphs above show the duration of travel during off peak times on the left-hand side and duration of travel during peak times on the right-hand side. The distributions for males (shown on the top row) and females (shows on the bottom row) are almost identical. During off peak times, those with a car take a larger proportion of trips that are less than 30 minutes while those without a car take longer trips. During peak times, the duration of travel does not differ greatly between those with and without a car.

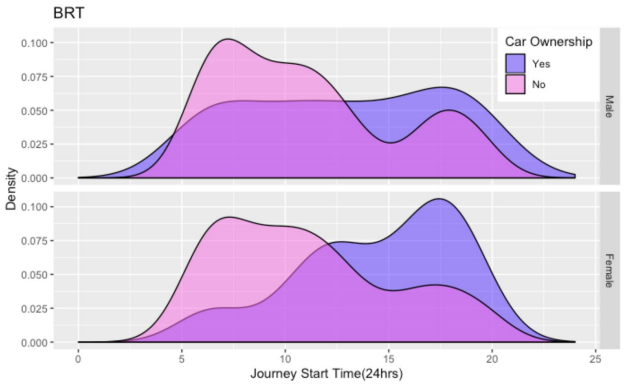

Fig. 4.1.3 *Start Time by Car Ownership (BRT)*

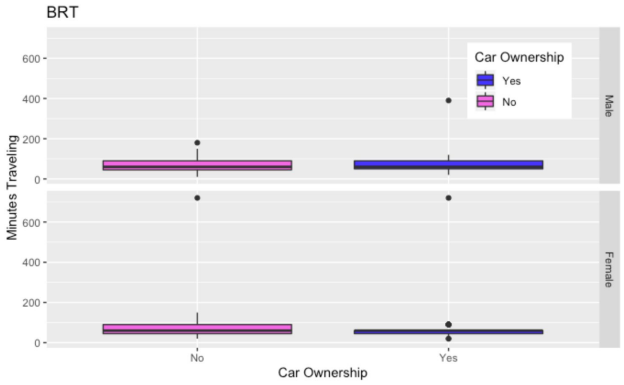

Fig. 4.1.4 *Travel Duration by Car Ownership (BRT)*

The figures above are generated from the BRT data. The second graph shows that the distributions of travel times are almost identical for males and females either with or without a car. In the first graph, journey start time is shown. Both males and females in the BRT survey with at least one car take more trips later in the day, and that respondents without cars report taking trips earlier in the day.

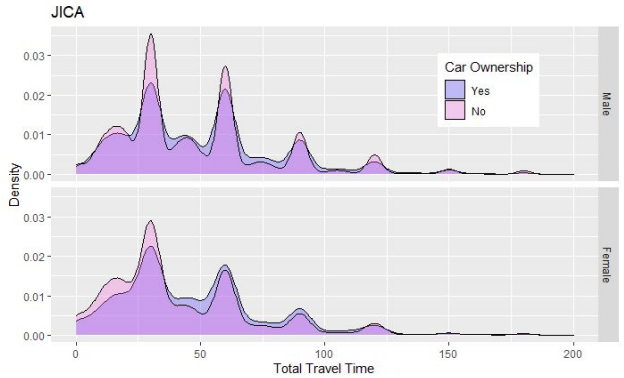

Fig. 4.1.5 *Travel Start Time by Car Ownership (JICA)*

Individuals without a car tend to take trips from 0-30 minutes more often than car owners, and the reverse is true for trips exceeding 30 minutes. When comparing distributions across genders, we see that the difference is most dramatic in the first 50 minute range. In this period male respondents without cars tend to take many trips around the 30 minute mark, whereas female respondents without cars take trips that are 30 minutes or less more consistently.

Sub Question 2: Does frequency of public transportation use differ between those with and without a car?

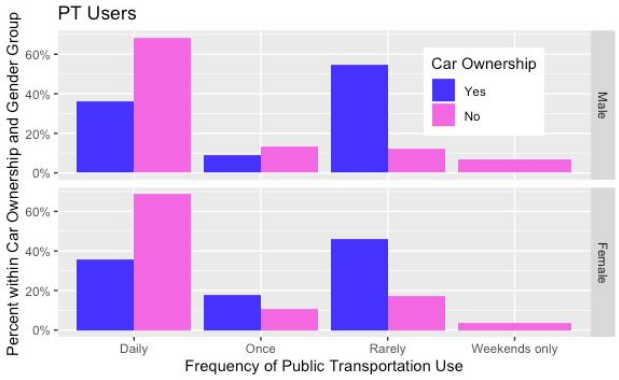

Fig. 4.2.1 *Frequency of Public Transportation Use by Car Ownership (PT Users)*

Frequency of public transportation use differs greatly between those who do and do not own a car; however that distribution is almost identical for males and females. Roughly 50% of those with a car use public transportation rarely, while only about 18% of those without a car use it rarely. Amlost 75% of those without a car use public transportation daily, while only 35% of those with a car use public transportation daily.

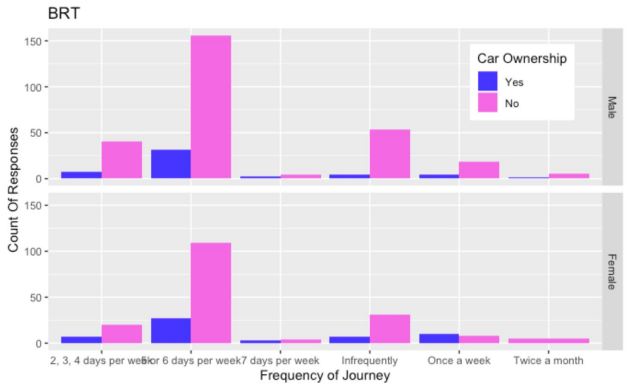

Fig. 4.2.2 *Frequency of Public Transportation Use by Car Ownership (BRT)*

Most respondents make trips 5 or 6 days a week, especially those without a car. The graph above also shows that most of the respondents of the BRT survey do not have a car.

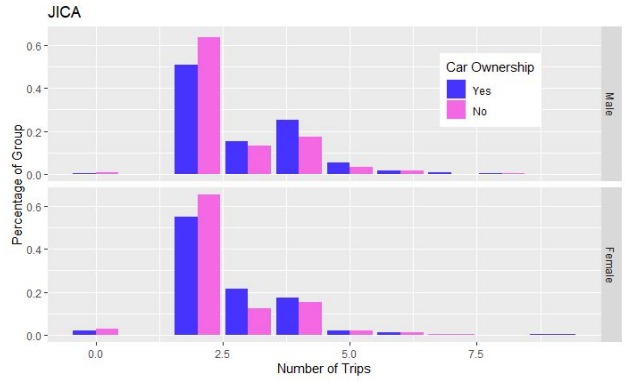

Fig. 4.2.3 *Frequency of Public Transportation Use by Car Ownership (JICA)*

For both males and females, respondents without a car were more likely than car owners to take two or less trips per day while those with a car were more likely to be taking 3 trips or more.

Sub Question 3: Does travel cost differ between those with and without a car?

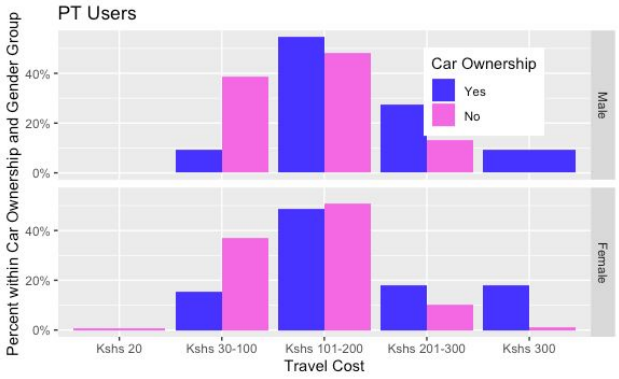

Fig. 4.3.1 *Travel Cost by Car Ownership (PT Users)*

The graph above shows a difference in average travel cost for both males and females in the PT Users survey; those without a car pay less on average for trips, while those with a car pay more for trips made.

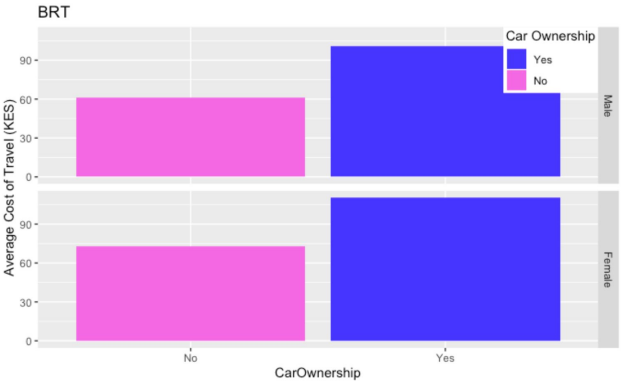

Fig. 4.3.2 *Travel Cost by Car Ownership (BRT)*

BRT respondents with cars report paying more on average for their trips than respondents without cars. Females report paying more than males, regardless of their car ownership.

Sub Question 4: Does willingness to pay extra for improved public transit differ between those with and without a car?

Fig. 4.4.1 *Willingness to Pay Extra for Safety by Car Ownership (PT Users)*

Those with a car are much more likely to pay extra for safer public transport compared to those without a car. This difference is much more prominent for females, potentially meaning that females with access to cars have disposable income and thus are able to pay for increased safety measures.

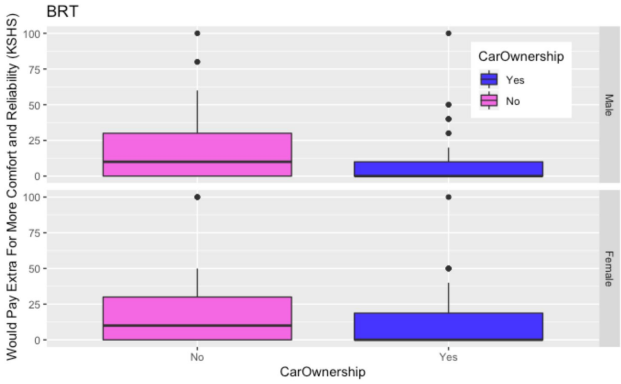

Fig. 4.4.2 *Willingness to Pay Extra for Comfort and Reliability by Car Ownership (BRT)*

Male and female BRT respondents with cars generally said they would pay less for increased comfort and reliability than respondents without cars.

Sub Question 5: Does travel purpose differ between those with and without a car?

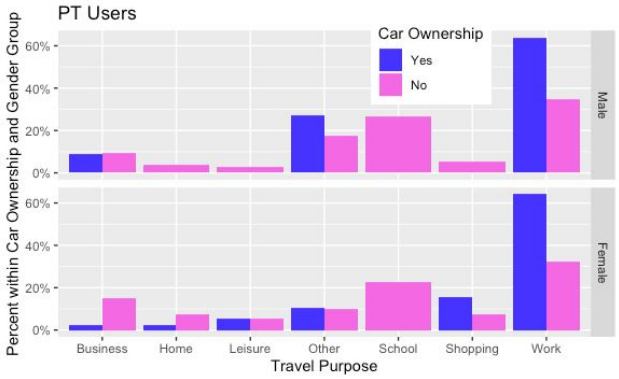

Fig. 4.5.1 *Travel Purpose by Car Ownership (PT Users)*

Respondents with a car make about 65% of their trips for work, while those without a car have more varied trip purposes. For example, those without a car make more trips for business and school than those with a car. This is true for both males and females.

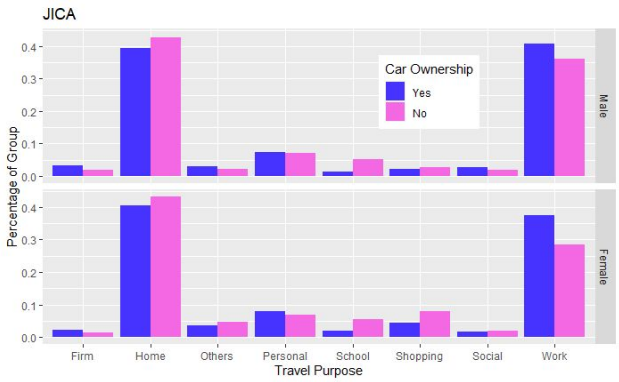

Fig. 4.5.2 *Travel Purpose by Car Ownership (JICA)*

JICA respondents who are car owners are more likely than non-car owners to be traveling for work, whereas individuals without a car are more likely to be travelling for home, school, or shopping purposes.The distributions across genders are very similar to one another, though the gap between car owners vs non-car owners traveling for work purposes is noticeably bigger for women.

Sub Question 6: Does security satisfaction differ between those with and without a car?

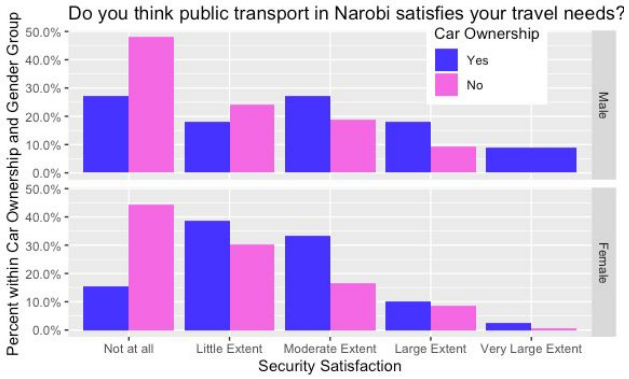

Fig. 4.6.1 *Security Satisfaction by Car Ownership (PT Users)*

The figure above shows that those without a car are less satisfied with the security of public transportation compared to those with a car. Additionally, all females are generally less satisfied with security compared to males.

Sub Question 7: Does report rate of sexual harassment differ between those with and without a car?

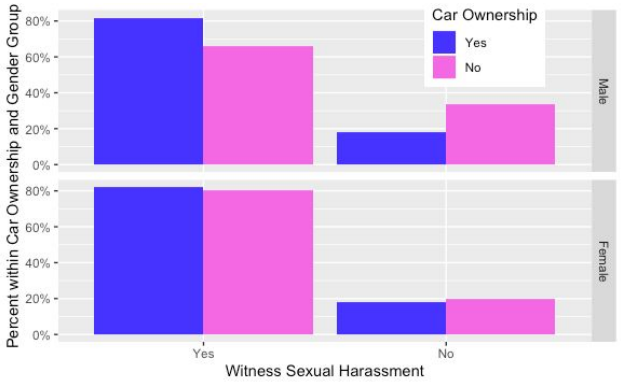

Fig. 4.7.1 *Report Rate of Sexual Harassment by Car Ownership (PT Users)*

There is not a large difference in the report rate of sexual harassment between those with a car and those without. This roughly equal distribution holds for the questions regarding if respondents witnessed specific forms of sexual harassment.

Hypothesis Testing

BRT Variable	**P-Value**
Travel Cost	0.01836
Extra Cost For Comfort/Reliability	0.00073*
Travel Duration	0.9321
Trip Purpose	0.002181*
Frequency of Use	0.01337

PT Users Variable	**P-value**
Travel Time (peak hours)	0.8802
Travel Time (off peak hours)	0.2025
Frequency of Use	2.571e-07*
Travel Cost	1.184e-08*
Willingness to Pay Extra for Safety	3.609e-05*
Travel Purpose	2.516e-05*
Security Satisfaction	0.0002707*
Witness Sexual Harassment	0.6845

JICA Variable	**P-Value**
Preferred Travel Mode	<.001*
Travel Time	<.001*
Travel Purpose	<.001*
Number of Trips	<.001*

The above tables show the significance tests run on the BRT, PT Users and JICA survey, respectively. Either a two sample t-test or a chi-squared test was conducted, depending on whether the variable being tested had two or more than two levels. With a significance level of 0.05, we used an adjusted p-value of 0.01 for the BRT survey, 0.00625 for the PT Users survey, and 0.0125 for the JICA survey to determine significance. The variables which had p-values below their respective threshold value for that survey are marked with an asterisk in the tables above. In the BRT survey, extra cost for comfort/reliability and trip purpose were significantly different between respondents with and without a car. In the PT users survey, frequency of use, travel cost, willingness to pay extra for safety, travel purpose and satisfaction in regards to security were found to be significantly different between those with and without a car. Finally, for the JICA survey, preferred travel mode, travel time, travel purpose and number of trips taken were significantly different between respondents with and without a car. Most of the tested variables showed a statistically significant difference between those with and those without a car. Thus, we can conclude that there is a difference in travel patterns and travel experiences on public transportation between those with and without a car.

Summary

Both the JICA and PT Users survey showed that those without a car take trips less than 30 minutes more often, while those with a car take trips longer than 30 minutes more often. Those without a car pay less on average for trips, while those with a car pay more for trips. However; females pay more for trips regardless of car ownership. In the PT Users survey, those with a car are much more likely to pay extra for safer public transport compared to those without a car. This difference is much more prominent for females. Travelers with a car are much more likely to be traveling for work whereas individuals without a car are more likely to be travelling for home, school, or shopping purposes. Respondents of the PT Users survey who do not have a car are less satisfied with public transit, and females are less satisfied than males, despite car ownership status. Overall, those without a car take shorter but more frequent trips for the purposes of home, school, and shopping; while also paying less for these trips.

Question 5: How does Mobility of Care influence Travel Purpose?

One idea we wanted to incorporate into our research to take it further is the "mobility of care". According to background literature, this represents the fact that when survey respondents report their trip purposes, the purpose of caring for another is often not listed as an option. Were it to be represented, research indicates that this would embody many of the responses that otherwise fall under the umbrella of "shopping" or "other". To see what our data would look like with the mobility of care principles, we revisited our previous analysis exploring the intersection of gender and trip purpose with this in mind.

Respondents of Age 18-55:

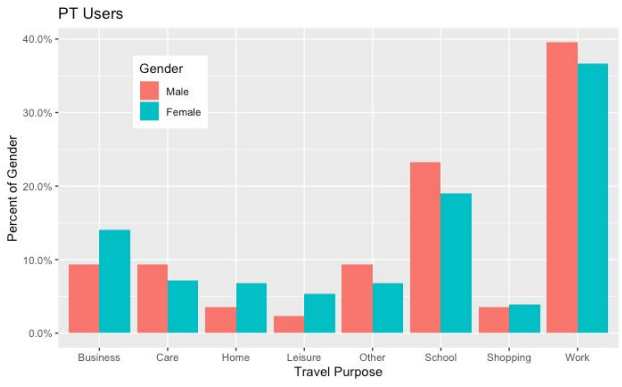

Fig. 5.1.1 *Travel Purpose by Gender with Mobility of Care from Respondents of Ages 18-55 (PT Users)*

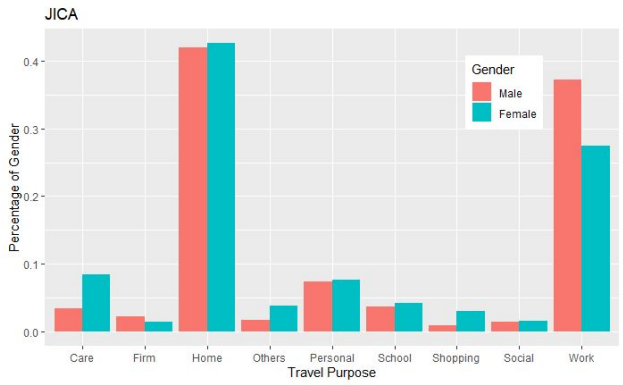

Fig. 5.1.2 *Travel Purpose by Gender with Mobility of Care from Respondents of Ages 18-55 (JICA)*

The above figure analizes trip purpose, with a mobility of care perspective. Only respondents from age 18-55 are considered. 'Care' trips are two-thirds of the original 'shopping' trips as well as one-third of the original 'other' trips and one-third of the original 'social' trips. With this new allocation, we see differences in trip purpose between genders. In both surveys, there is a larger proportion of female respondents taking trips for care than males but a much larger proportion of males taking trips for work. A Chi-square test was run to see if there was a statistically significant difference between travel purpose between men and women for both surveys. For the PT Users survey, the test statistic was 12.335 which resulted in a p-value of 0.09005, which is not significant. For the JICA survey, the test statistic was 989.15 with a p-values < 0.01, which is statistically significant. This is most likely due to the JICA survey having a much larger sample size, giving it more statistical power, or the ability to detect differences when differences are truly present. Since the JICA survey had a more representative sample, we have evidence that travel purpose is significantly different between men and women of ages 18-55.

Respondents of Age 18-35:

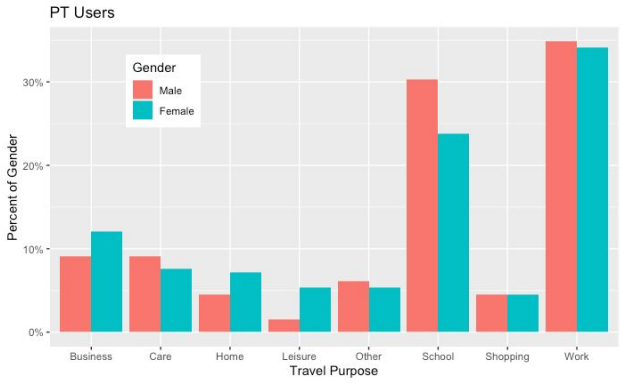

Fig. 5.2.1 *Travel Purpose by Gender with Mobility of Care from Respondents of Ages 18-35 (PT Users)*

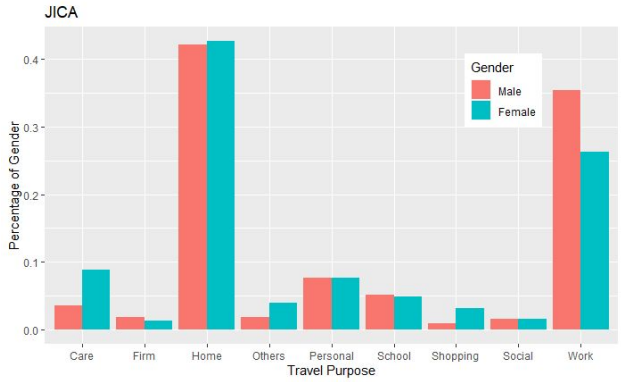

Fig. 5.2.2 *Travel Purpose by Gender with Mobility of Care from Respondents of Ages 18-35 (JICA)*

The graphs above show the distribution of travel purpose of respondents to the JICA and PT Users survey of ages 18-35. Again, we can see that females take care trips more often than men while females take work trips less often than men. The other trip purposes show inconsistent results between the two surveys. A Chi-square test was run to see if there was a statistically significant difference between travel purpose between men and women. For the PT Users survey, the test statistic was 11.606 which resulted in a p-value of 0.1143, which is not significant. For the JICA survey, the test statistic was 706.86 with a p-values < 0.01, which is statistically significant. Again, this is due to the large sample size of the JICA survey and increased power.

Respondents of Age 36-55:

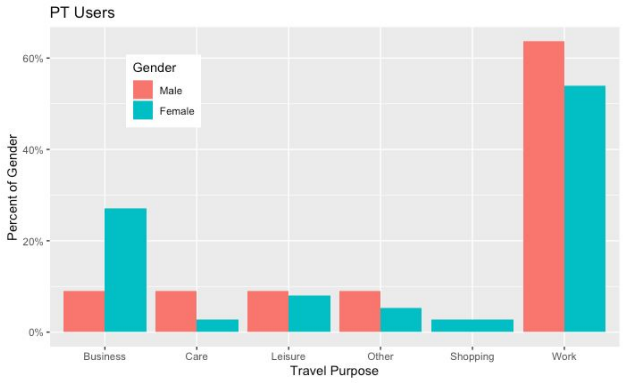

Fig. 5.3.1 *Travel Purpose by Gender with Mobility of Care from Respondents of Ages 35-55 (PT Users)*

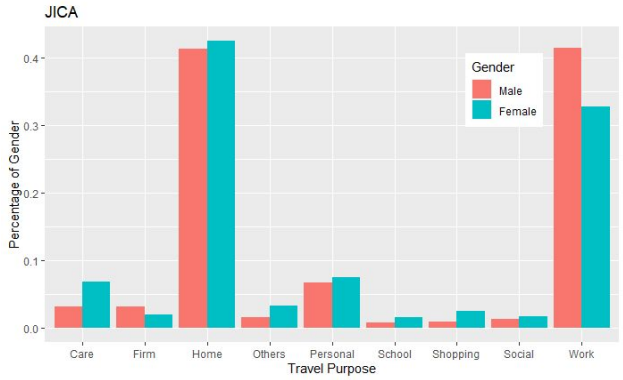

Fig. 5.3.2 *Travel Purpose by Gender with Mobility of Care from Respondents of Ages 35-55 (JICA)*

Now we examine respondents that are between 36 and 55. The JICA data shows very similar results to the other age categories with females taking more care trips and males taking more work trips. However, for the PT Users survey, the distribution is much different. The PT Users survey shows males taking more care trips and more work trips. Females are now taking a much larger proportion of trips for business as well. This is most likely due to the fact that with the age limits, there are only 11 males and 37 females, so with six different categories there are very few respondents in each category. A Chi-square test was run to see if there was a statistically significant difference between travel purpose between men and women on both data sets. The test statistic was 2.6506 which resulted in a p-value of 0.7537, which is not significant. However; due to the small sample size this result is not very reliable. For the JICA data, the test statistic was 195.08 which resulted in a p-value of less than 0.01, which is statistically significant. Even with these age restrictions, the JICA data still had a very large sample size. Thus, we can conclude that there is a statistically significant difference between males and females in regards to travel purpose.

Summary

Taking the mobility of care principles into account, JICA data showed differences in trip purpose among genders in the overall age bracket (18-55) and the two individual age brackets within it (18-35 and 36-55). The relationship between gender and trip purpose also did not differ greatly when comparing 18-55 year olds to all other age groups. The JICA survey showed that less than 10% of participants from each gender went on trips related to care, with more female respondents going on these trips than male respondents. However, data from the PT User survey did show more differences among individual age brackets, as seen below.

For respondents aged 18 to 55 in the PT Users survey, similar to in the JICA data, less than 10% of participants from each gender went on trips related to care, with more female respondents going on these trips than male respondents. For respondents aged 36 to 55 in the PT Users survey, unlike in the JICA data, over 10% of female respondents reported going on these trips compared to less than 5% of male respondents. The respondents aged 36 to 55 in the PT Users survey had a lower proportion of participants going on trips related to care than any other age group. Unlike the other age groups, more male respondents in this bracket are going on these trips than female respondents.

Logistic Regression

Logistic regression is a technique used to model a binary response variable, a variable with only two outcomes. We used logistic regression to predict if

the main mode of the trip was walking or not as well as to predict if the trip was inter-zone or intra-zone. For both of these models, we used personal monthly income, age and gender of the respondent as explanatory variables. We used the JICA data set for both models. The predictions from the logistic regression model are most easily interpreted in the form of odds, which are a ratio of the likelihood event A occurs compared to the likelihood event B occurs. For example, the odds of a coin landing on heads and not tails would be 1 to 1 odds, expressed as 1 / 1, or 1.

P-values can be computed for each of the explanatory variables. If these p-values are statistically significant, we can conclude that that explanatory variable is statistically significant in predicting the response variable, after adjusting for the remaining explanatory variables in the model. As seen in the tables below, all variables were statistically significant.

Predicting if Main Mode was Walking

Variable	**Estimate**	**Change in Odds**	**P-value**
Personal Monthly Income	-0.21596	-19.4232	<.001*
Age	-0.0101	-1.0049	<.001*
Gender	0.2804	32.3659	<.001*

The p-values for income, age and gender are all below any reasonable significance level. Thus, we can conclude that individually each of these predictors is significant after adjusting for the other two.

Accounting for age and gender, each increase in personal monthly income bracket level is associated with a decrease in the odds that a respondent was walking by 19.4232%. Also, accounting for income level and gender, each one year increase in age is associated with a decrease in the odds that a respondent was walking by 1.0049%. Finally, accounting for age and income level, a respondent being female is associated with an increase in the odds that a respondent was walking by 32.3659%.

Predicting if Trip was Intra-zone or Inter-zone

Variable	**Estimate**	**Change in Odds**	**P-value**
Personal Monthly Income	-0.128964	-12.09944	<.001*
Age	-0.008675	-0.8637481	<.001*
Gender	0.302674	35.34732	<.001*

The p-values for income, age and gender are all below any reasonable significance level. Thus, we can conclude that individually each of these predictors is significant after adjusting for the other two.

Accounting for age and gender, each increase in personal monthly income bracket level is associated with a decrease in the odds that a respondent traveled within their census zone by 12.0944%. Also, accounting for income level and gender, each one year increase in age is associated with a decrease in the odds that a respondent traveled within their census zone by .8637481%. Lastly, accounting for age and income level, a respondent being female is associated with an increase in the odds that a respondent traveled within their census zone by 35.34732%.

Multiple Linear Regression

We created a multiple linear regression model, which is a statistical technique that uses several explanatory variables, to predict one response variable. This is very similar to the logistic regression which was previously used. However, in linear regression the explanatory variable is quantitative instead of binary. We attempted to predict the difference in slum build up of the start zone to the destination zone of the trip. The explanatory variables that were used were age, gender, and income. One model was made using all trips as valid inputs, and another took only inter-zone trips as valid inputs. This was to account for the fact that trips in the same zone had slum buildup percentage differences of 0.

Model 1: All trips

Variable	**Estimate**	**P-value**
Personal Monthly Income	0.0082	0.903
Age	-0.001	0.953
Gender	0.023	0.944

The above table shows that none of the explanatory variables have significant p-values. Thus, income, age and gender are not significant predictors of difference in slum build up from start to end zone.

Model 2: Inter-Zone Trips

Variable	**Estimate**	**P-value**
Personal Monthly Income	0.006	0.951
Age	-0.002	0.935
Gender	0.054	0.924

For the second model, we only looked at trips that ended in a different zone than the starting zone. Again we see that none of the explanatory variables have significant p-values. Thus, income, age and gender are not significant predictors of difference in slum build up from start to end zone for inter-zone trips.

Unsuccessful Approaches

While many parts of the investigation did yield meaningful and interesting results, not everything we attempted was successful and not all data sources came to the same conclusions. The JICA and PT Users surveys showed statistically significant differences by gender, but overall the data from the BRT survey did not. It also did not contain data on economic status so it could not be used in main questions 2 and 3, which considered respondents by income level. The following table shows the results of t-tests on questions from the BRT survey by gender. We conducted 7 t-tests, so an adjusted p-value of 0.0071 was used to account for multiple testing on the data. At the 0.05 overall significance level, none of the tested variables gave strong evidence of statistically significant differences by gender.

Variable	**P-value**
Can't Travel When I Want(1-5)	0.3849
More Than One Vehicle(1-5)	0.1878
Takes Too Long(1-5)	0.0065
Too Expensive(1-5)	0.2747
Travel Unsafe(1-5)	0.3728

Uncomfortable(1-5)	0.3519
Wait Long Time To Board(1-5)	0.7681

As mentioned earlier, we would give less consideration to the BRT data than the other data sources. This was because the BRT survey had a smaller sample size, along with gender analysis not being one of its goals. With this being said, the BRT data was less useful in finding differences in public transportation usage and experience between genders.

Potential Future Work

We discovered a good deal about the different uses of Kenyan public transportation in this investigation, but future explorations to this topic could go deeper. Our analysis began to incorporate predictive models to take a different approach to understanding what factors make public transportation usage and experience different for males and females. If this investigation were to continue on, our group would have wanted to improve our models to attempt to better illustrate how male and female public transportation needs differ. In particular, it would have been great if we could have gotten a decision tree model where we could see how relatively important gender is in determining various response variables. This would give a good visual representation of how different some response variables were different between genders.

The project also contained free response data which our group did not analyze. Free text response data was a significant part of the PT Users survey. The information in these responses could help us learn about the more complicated relationships in the data. It may be difficult to analyze, but it may also be a worthwhile approach. The surveys also included data on each census zone in the city and how much "slumliness" it had accumulated. It would be interesting to examine the inter-zone trips and explore which factors best predict how this figure would change.

Conclusion

Our investigation of the differences in public transportation usage between genders in Kenya has shown that different groups use the Kenyan public transportation system in different ways. First, we established that travel patterns differed between males and females in Kenya. Then we found differences in public transportation usage between low-income and high-income Kenyans, then differences between low-income and high-income Kenyan females. Finally, we saw strong evidence of differences in public transportation usage based on car ownership.

After searching for differences in public transportation between explicit groups, we tried to estimate proportions of trips that were "care trips" made by both men and women, to see if these trips were different from "non-care trips". Then we tried to use 2 different logistic regression models to predict whether or not a traveler was walking to their destination and then if the traveler was moving within the zone they started in or to another zone. These models found that the demographic variables of gender, age, and income level were all useful predictors of each response. Lastly we made an attempt to run a multiple linear regression model to predict the difference in percentage slum-buildup of trips, but no predictors came up as having a significant association with the response.

We used many different perspectives in this project to look at how the Kenyan public transportation system is used, and learned about a number of meaningful associations. The original goals of the investigation centered around differences between genders, and our findings suggest that these were

almost certainly present. If changes were made to systems such as the Kenyan public transportation system that accounted for the needs and interests of various demographic groups, the system would better serve the collective community. This being said, changes with consideration to the interests of both genders would likely create a more equally valuable public transportation system experience for both male and female travelers.

References

Deborah Salon & Sumila Gulyani (2010) Mobility, Poverty, and Gender: Travel 'Choices' of Slum Residents in Nairobi, Kenya, Transport Reviews, 30:5, 641-657, DOI: 10.1080/01441640903298998

Inés Sánchez de Madariaga (2013) FROM WOMEN IN TRANSPORT TO GENDER IN TRANSPORT: CHALLENGING CONCEPTUAL FRAMEWORKS FOR IMPROVED POLICYMAKING, Journal of International Affairs, Vol. 67, No. 1.