Philadelphia Household Travel Survey

The household ID we extracted is 100702. By referring to the headers and data in Excel, along with the corresponding data dictionary, we can obtain basic information about the household.

Salary

From table "1_Household_Public", it is possible to find that the selected household's "INCOME" is "7". Referring to the data dictionary's "Value" field, we can see that "4" represents a household annual income in the "$35,000 to $49,999" range.

Race

From table "2_Person_Public," it can be found that "RACE" is 988. Referring to the data dictionary's "Value" field, this corresponds to "Pilot Record (Non-Hispanic)."

Composition

From the same "2_Person_Public" table, we can find that the "PERSON_NUM" indicates that the family consists of 2 people. Therefore, the family is likely a married couple based on their income and other data.

County

From the table "1_Household_Public," we can see that "H_STATE" is "34," and "H_COUNTY" is "34005." According to the data dictionary, "H_STATE" stands for "HH State based on Geocoded location," and "34" represents "New Jersey." "H_COUNTY" stands for "HH County based on Geocoded location," and "34005" represents "Burlington." So, their place of residence is in Burlington, New Jersey.

Car

From the same "1_Household_Public" table, we can see that the selected family has "TOT_VEH" as "2." According to the data dictionary, "TOT_VEH" stands for "Total Household Vehicles," indicating that the extracted family owns two vehicles.

Philadelphia Household Travel Survey

2) Describe the daily activities and travel of the household's members using the trip data.

Through the "4_Trip_Public" table and the data dictionary, we've learned that "ACTIV1" represents "Activity 1 at Origin," and "ARRIVE" stands for "Arrival time (MILITARY TIME)." By examining the data and the data dictionary's explanations, we can see that for this family, "ACTIV1" has values of "1, 5, 8, 11," which correspond to the following meanings: "1 Home activities not related to work, school, or online," "5 Work for pay," "8 Everyday shopping (grocery, drug store, gas, etc.)," and "11 Eat out (restaurant, drive-thru, etc.)." Based on the arrival times, we can deduce that their activities at 9:38 AM and 5:45 PM are related to work, while activities at 12:17 PM, 4:25 PM, and 7:25 PM involve non-work-related activities. The activity at 3:23 PM is related to shopping, and the activity at 7:10 PM is eating out.

Furthermore, the "MODE" value is "5," and "MODE" represents the "Mode of Transportation," with "05" indicating "Auto / Van / Truck." Another entry, "MODE_AGG," represents the "Aggregate of transportation modes," and its value is "3," indicating "Private Vehicle." Therefore, this family primarily uses private vehicles for transportation.

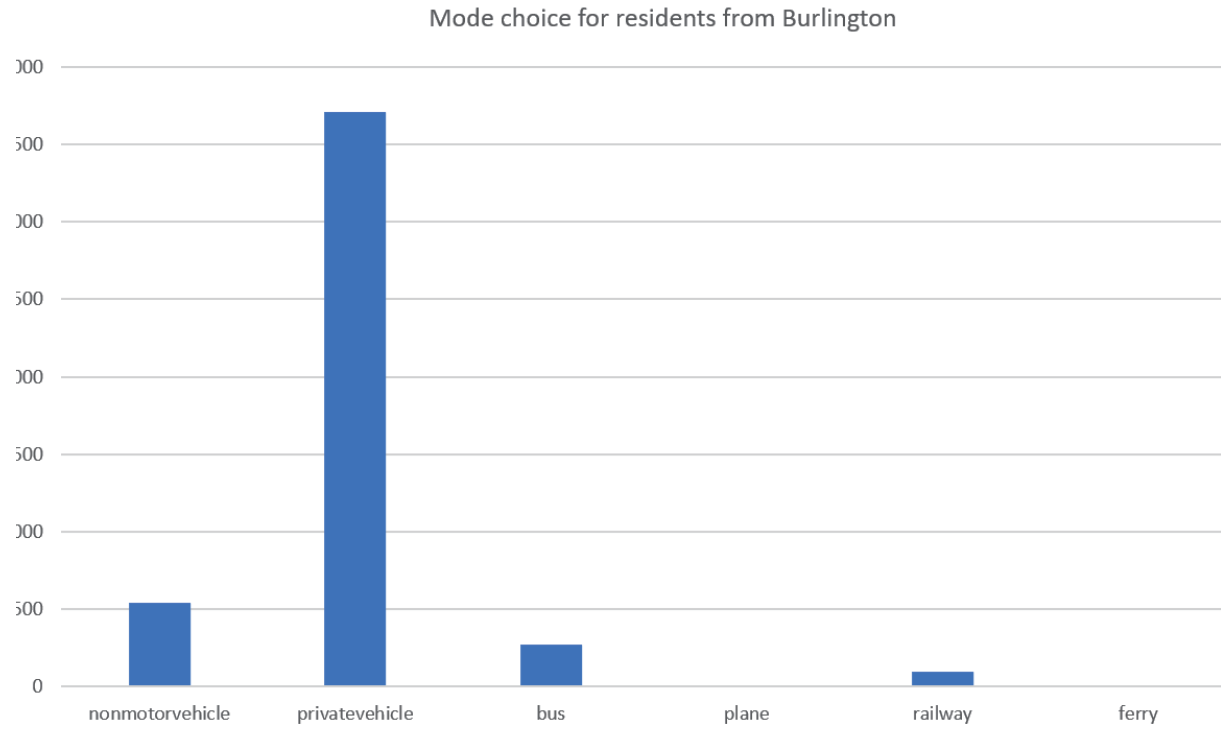
In summary, this family consists of two members, and based on other data, we can infer that they are a married couple. One of them likely commutes to work while the other stays at home, or both of them may commute to work. The morning activity at 9:38 AM appears to be related to commuting, and the midday activity at 12:17 PM suggests non-work-related activities. The afternoon activity at 3:23 PM is a shopping trip, followed by more non-work-related activities at 4:25 PM. In the evening, at 7:10 PM, the two individuals go out to eat, and at 7:25 PM, they engage in other activities. This describes their daily routines and travel patterns, with both members using private vehicles (indicating they have two cars). Due to the variety of daytime activities, it is likely that one person has a regular daytime job, while the other may work from home or handle household chores, shopping, and additional family-related activities. This information provides insight into their daily activities, travel methods, and purposes using travel data.

Philadelphia Household Travel Survey

3) Provide an estimate of the total number of bicycle trips represented by the survey data. Note that you should use the Person Weight to make this estimate. This number represents the total number of people each person in the survey is supposed to represent.

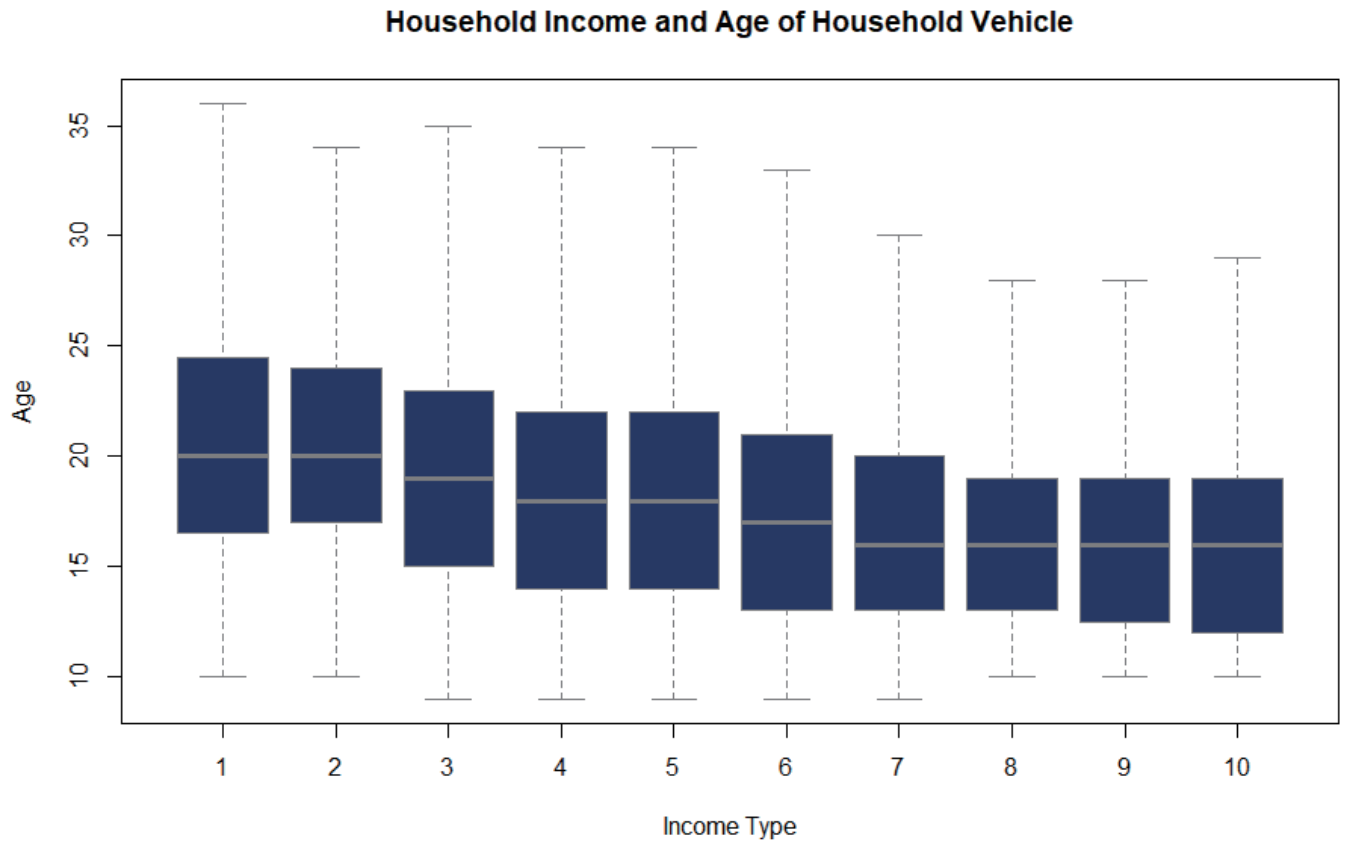
The estimated total number of bicycle trips represented by the survey data is 143641.9.

4) Make a table that shows the mode choice for residents from the county of the household from question 1. Note that you will have to make a choice about how to group the modes together.



The household in question 1 is located in Burlington County, as shown in the table. The most common mode choice for residents here is private vehicles, and it significantly outweighs other mode choices.

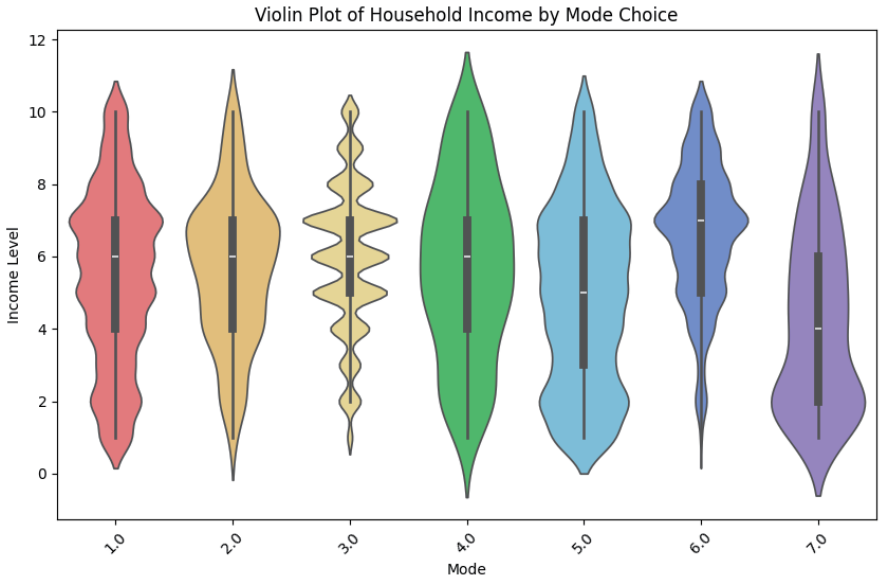
5) The relationship between household income and the age of household vehicles.



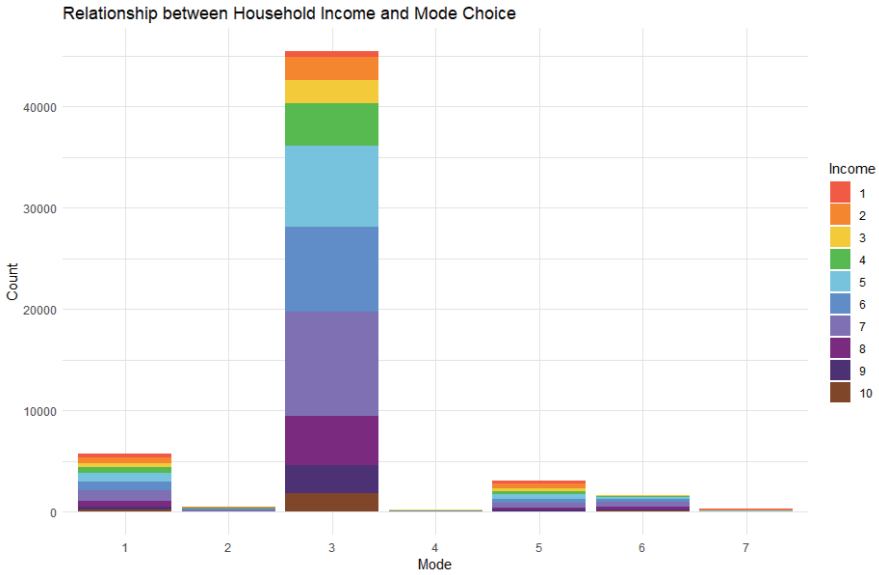
In the chart, the horizontal axis represents "Income Type," and the vertical axis represents the "Age of Household Vehicle." The Income Type ranges from 1 to 10 incrementally. For example, households with Income Type 1 (annual income $0 to $9,999) have household vehicles aged between 16 and 24 years. Households with Income Type 4 (annual income $35,000 to $49,999) have household vehicles aged between 14 and 23 years. Households with Income Type 7 (annual income $100,000 to $149,999) have household vehicles aged between 13 and 20 years. Lastly, households with Income Type 10 (annual income $250,000 or more) have household vehicles aged between 12 and 19 years. The overall trend indicates that as household income increases, the age of household vehicles gradually decreases. This suggests that higher-income households have more financial resources and can afford to replace their older vehicles, while lower-income households tend to keep their household vehicles for a longer period.

Philadelphia Household Travel Survey

6) The relationship between household income and mode choice for all households.



Graphic 1



Graphic 2

6) The relationship between household income and mode choice for all households.

In the four tables mentioned above, we can determine from the data dictionary that Income Level ranges from 1 to 10, progressively representing the following household income intervals: "1 $0 to $9,999, 2 $10,000 to $24,999, 3 $25,000 to $34,999, 4 $35,000 to $49,999, 5 $50,000 to $74,999, 6 $75,000 to $99,999, 7 $100,000 to $149,999, 8 $150,000 to $199,999, 9 $200,000 to $249,999, 10 $250,000 or more." Mode Type represents the "Mode of Transportation," which includes the following options: "1 Walk; 2 Bike; 3 Private Vehicle; 4 Private Transit; 5 Public Transit; 6 School Bus; 7 Other."

From the previous four tables, we can see that households in any income interval predominantly use "Mode 3," which is traveling by Private Vehicle. The number of people using this mode of transportation is significantly higher than other modes. Walking is the second most common mode of transportation, followed by Public Transit, School Bus, Biking, and so on. Therefore, the use of private vehicles is the most common mode of transportation among many households.

Additionally, for households in the income interval "2" ($10,000 to $24,999), biking is the most commonly used mode of transportation. Among those in the income interval "7" ($100,000 to $149,999), households using private vehicles for transportation constitute the largest group. (For households in income intervals 5-7, they make up the most significant number of private vehicle users, but it's worth noting that this could be because households in the "$50,000 to $149,999" income range are already the most numerous.)

Philadelphia Household Travel Survey

7) a. According to the regression model, each additional job within a half mile of a station correlates with how many more daily transit riders per day?

Through mathematical computation, it can be determined that, the coefficient for "Jobs within a half mile" in the regression output is 0.179.

So according to the regression model, for each additional job within half a mile of the station, it is estimated that there will be an increase of 0.179 daily bus passengers.

7) b. If there are an average of 1,793 boardings per station and 3,130 jobs around each station, what percentage increase in ridership does the model predict due to a doubling of the number of jobs around a station?

Mathematical Computation:
Elasticity measures the percentage change in one variable (in this case, boardings) due to a 1% change in another variable (jobs).

$$\text{Elasticity} = \beta \times \left(\frac{\text{Average Number of Jobs}}{\text{Average Boardings}} \right)$$

The model predicts that a doubling of the number of jobs around a station would result in an approximately 31.25% increase in ridership.

7) c. Describe the relationship between AM peak service frequency and light rail boardings.

The coefficient for "AM peak service frequency" is 105.822. This means that for every unit increase in the AM peak service frequency, the daily light rail boardings are expected to increase by 105.822, all else being equal. This is a positive relationship, indicating that higher AM peak service frequencies are associated with higher expected daily light rail boardings.

Philadelphia Household Travel Survey

Additional information about regression indicators:

Observations: 406

This indicates that the regression model is estimated based on 406 data points or observations. In other words, this is the number of records in your dataset.

R^2: 0.408 (Coefficient of Determination)

The coefficient of determination, represented as R^2, reflects the percentage of variability in the data that the model explains. In this model, the value is 0.408, meaning the model accounts for 40.8% of the variability in the data. This is a moderate value, indicating that the model has some explanatory power but leaves a significant portion of the variation unexplained.

Adjusted R^2: 0.402

Adjusted R^2 is an R^2 value that has been adjusted for the number of predictor variables in the model to prevent overfitting. When additional predictor variables are added to the model, even if they do not contribute to the response variable, R^2 can increase. Adjusted R^2 corrects for this situation. In this model, the adjusted R^2 value is 0.402, indicating that the model accounts for 40.2% of the variability in the data, taking into account the number of predictor variables in the model.

In summary, these three statistical measures provide information about the quality of the model and the size of the dataset. R^2 and adjusted R^2 values inform us about the model's explanatory power, while the number of observations tells us how many data points the model is based on.