Household Income and Transportation Usage

Jasmine Mangat

May 9 2021

Abstract

With rising concerns over the impact of climate change on our planet, specifically urban communities, there has been a recent push in designing cities to be more eco-friendly through investing in public transportation and creating neighborhoods more accessible to bicyclists and pedestrians. However, these changes may be feeding into another urban issue that has been making headlines: gentrification. Coined as 'eco-gentrification,' researchers are pointing to the inherent issues with touting the benefits of sustainable urban planning without considering equity (Rice). Through this study, I focus on the current state of transportation options available to those living and working in the Boston–Cambridge–Newton, MA–NH Metropolitan Statistical Area depending on their income level and where they have to travel to work. This study could potentially play a role in studies relating to transportation accessibility and its potential role in promoting eco-gentrification in urban areas.

Introduction

Transportation is one of the key factors which influences urban planning because of its power to connect different people and industries throughout a given area. A study in Brazil conducted interviews with the poorer population in a given city and found how unaffordable and unexpansive transportation systems can inhibit access to jobs, businesses, and other resources that are too far by foot or bicycles. Although American and Brazilian land models differ, this phenomenon is important to consider when considering how gentrification can often lead to people being kicked out of their neighborhoods into those "less-desirable" and "inaccessible."

Transportation has also been a topic of interest in the recent surge in the sustainable and smart city movement which helps promote more climate-conscience and data-driven urban planning. However, recent studies have found that efforts such as more sustainable and cheaper transportation systems such as bikes may lead to increased housing prices (Rice). Thus, eco-gentrification points out how increased sustainable transportation accessibility efforts may be targeting may lead to less accessibility overall to a city's overall resources for low-income populations.

Through this study, I look at how different household incomes travel from their homes to work while controlling for different living and working locations in a metropolitan area. Through this study, we hope to gain a better understanding of how transportation choices among lower and upper-income groups might differ depending on their travel needs. While I do not consider transportation accessibility directly, I provide suggestions on how this study can be extended to do so. Additionally, I will be discussing the nuances that come from studies involving transportation choices and accessibility due to various hidden variables and flawed assumptions that are often made.

Data

I use a 2019 sample from the American Community Survey consisting of 43,951 observations. The sample consists of people who live and work in the Boston–Cambridge–Newton, MA–NH Metropolitan Statistical

Area which encompasses eastern Massachusetts and southern New Hampshire through Suffolk, Norfolk, Plymouth, Middlesex, Essex, Rockingham, and Strafford counties. I chose this area because although it is considered to be one of the wealthiest parts of the two states, studies have shown that it is also the area with the most income inequality (Sommeiller). This area is also dominated by the MBTA in terms of commuting to, from, and within the city of Boston and surrounding towns. However, this area also includes counties such as Rockingham and Strafford which do not have as much access to MBTA services such as the Commuter Rail. However, since Metropolitan Statistical Areas are often defined by commuting patterns, Rockingham and Strafford residents and workers may still travel to and from the city. Thus, this sample should be able to capture a diverse set of transportation methods.

My sample focuses on the following variables: HHINCOME, METRO, PWTYPE, TRANWORK, and CARPOOL. HHINCOME gives the total household income of the individual observed. I do not include other family members of the individual in the final list of observations. My final model uses a new variable called HHINCOME_10000 which divides each number in HHINCOME by 10000 dollars. I filtered out any values below 0 to avoid negative incomes and also any values above 500 due to outliers which can be seen in Table 1 below.

METRO is a categorical variable that indicates whether or not a person lives in a metropolitan area and if so, which part. I only include those living in metropolitan areas which include the following categories: central/principal city, not in central/principal city, and central/principal city status indeterminable (mixed). PWTYPE is another categorical variable similar to METRO for where the observed individual works. Again, I only include individuals working in metropolitan areas within the same categories as METRO. I combine METRO and PWTYPE using a new variable called HOME_TO_WORK_NAME to establish four different types of geographical travel patterns: "City to city," "City to suburbs," "Suburbs to city," and "Suburbs to suburbs." The first location indicates where the person lives while the second indicates where they work. The mixed category in METRO and PWTYPE was included within those who lived in the city and those who worked in the city respectively. A city's transportation system, including Boston's, often extends outside of the city itself into surrounding areas. So, these areas, have access to many of the transportation options available to those living within the city.

Finally, TRANWORK indicates which mode of transportation the observed individual uses to get from home to work while CARPOOL shows whether or not the person carpools to work. I combine a more filtered down version of TRANWORK and CARPOOL into TRANWORK_SM. All modes of transportation involving privately owned vehicles such as cars or vans and those who do not carpool are combined into the "Private vehicle" category in TRANWORK_SM. I create a separate category for "Carpool." All public transportation options are combined into the "Public transport" category. I also include categories for "Bicycle" and "Walk only." All NA individuals with NA values in any of these categories were removed.

Below, I included a graph to show the distribution of observations across different incomes on the x-axis in different transportation types on the y-axis while differentiating between HOUSE_TO_WORK_NAME.

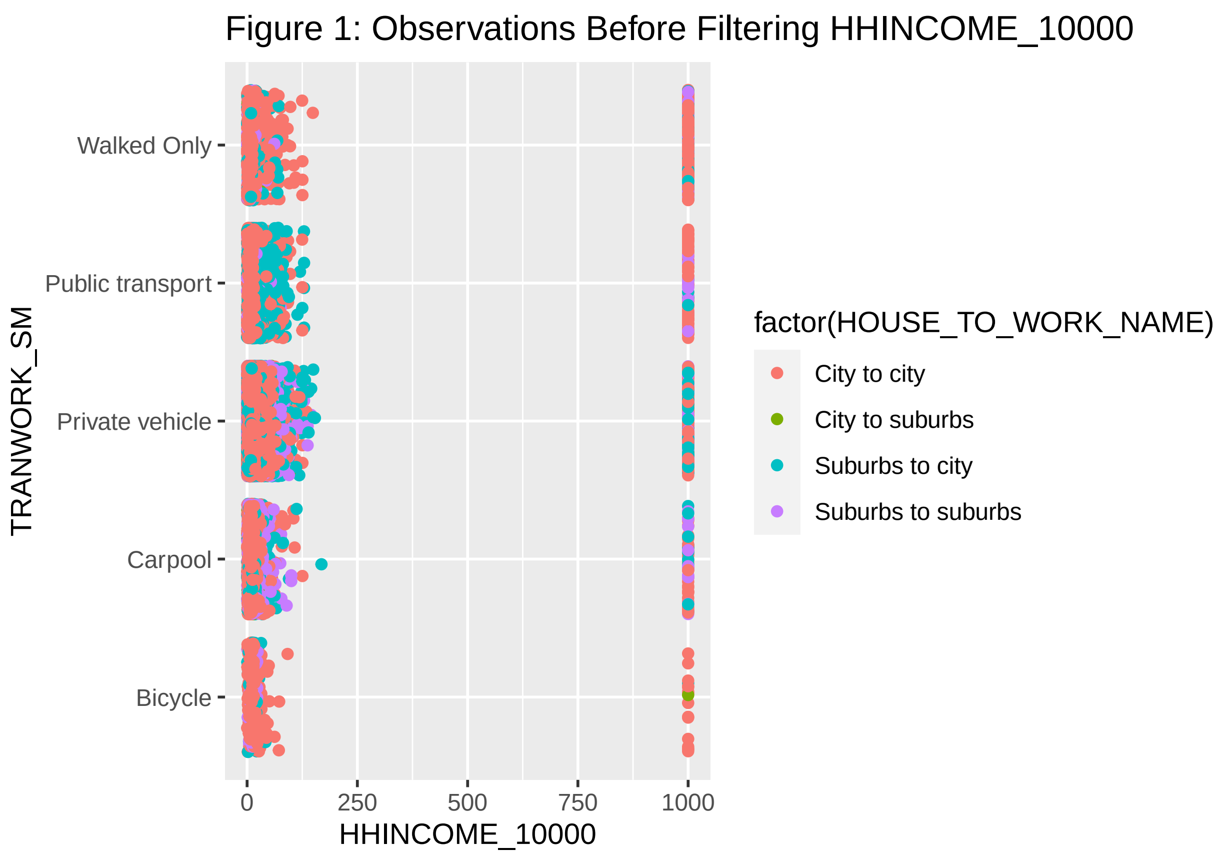


Figure 1: Observations Before Filtering HHINCOME_10000
Walked Only
Public transport
Private vehicle
Carpool
Bicycle
TRANWORK_SM
0
250
500
750
1000
HHINCOME_10000
factor(HOUSE_TO_WORK_NAME)
City to city
City to suburbs
Suburbs to city
Suburbs to suburbs

Figure 2: Observations After Filtering HHINCOME_10000

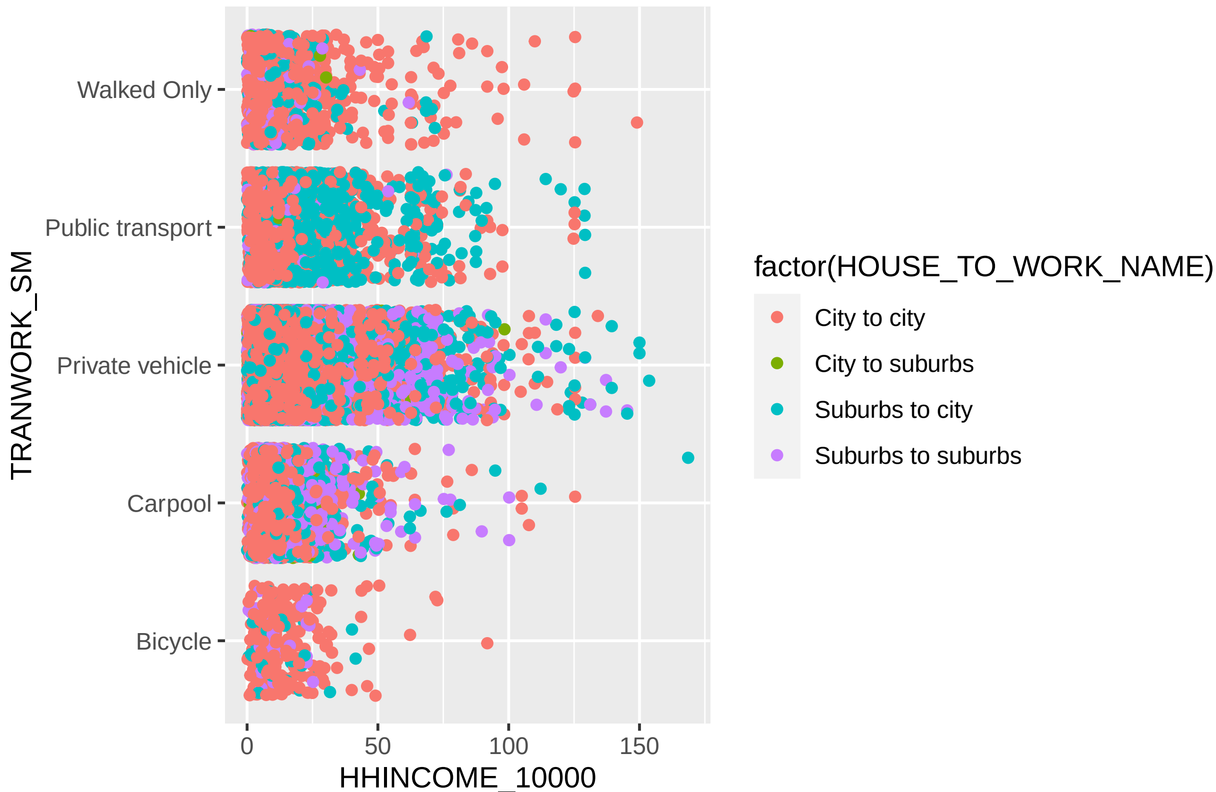


Figure 1 shows a large gap between people with an HHINCOME_10000 value of aroudn 170 and those with a value of 1000. These points made up 965 of the observations. Due to the small amount of points compared to the total number of observations that lie around 1000 for HHINCOME_10000, I exluded these points from my final list of observations. Figure 2 then gives a much more understandable depiction of the observations.

Table 1: Table 1: Summary Statistics for HHINCOME_10000

HHINCOME_10000
Min. : 0.0004
1st Qu.: 6.7080
Median : 11.1100
Mean : 13.9860
3rd Qu.: 17.3300
Max. :168.6000

Table 2: Table 2: Count for HOUSE_TO_WORK_NAME

HOUSE_TO_WORK_NAME	Total
City to city	20364
City to suburbs	986
Suburbs to city	12548
Suburbs to suburbs	10051

Table 3: Table 3: Count for TRANWORK_SM

TRANWORK_SM	Total
Bicycle	299
Carpool	3546
Private vehicle	34423
Public transport	4250
Walked Only	1431

Model and Hypothesis

I will use multinomial logistic regression where the dependent variable is TRANWORK_SM. The independent variables are HHINCOME_10000 and HOME_TO_WORK_NAME. I also included an interaction term between the two independent variables because I suspect that the impact of household income on transportation choice will differ depending on people's geographic travel patterns.

$$log(\pi_j/\pi_1) = \beta_0 + \beta_1 HHINCOME\_10000 + \beta_2 HOME\_TO\_WORK\_NAME +$$
$$\beta_3 HHINCOME\_10000 * HOME\_TO\_WORK\_NAME + \epsilon_i$$

$$\pi_1, \pi_2, \pi_3, \pi_4, \pi_5$$

= probability of Individual, probability of Carpool, probability of Public Transit, probability of Walking, probability of Biking

"Private vehicle" will be my baseline dependant variable for comparison since this is the most common transportation type in this metropolitan area. Since HOME_TO_WORK_NAME is a categorical variable, I will be treating "Suburb to suburb" as my reference category.

I hypothesize that as household income increases, I will see that the probability of using bikes or walking to work will increase when traveling within the city because wealthier people, who tend to work in wealthy areas of a city, can afford to live nearby. I suspect that biking and walking will have very low probabilities when looking at travel within the suburbs or between the suburbs and the city among all income groups. Due to more accessibility to bikes and walking, I suspect that the probability of using public transportation to get to work will decrease as income increases within the city.

It is worth noting that after dividing up HHINCOME_10000 into three income brackets (described in the Results section of this paper), I found that lower, middle and upper-income people tend to have 2.97, 2.49, and 2.81 private vehicles respectively. However, Among all travel patterns, I still suspect that the probability of carpooling will decrease and the probability of using private vehicles will increase are income increases. This is because higher-income people can afford services like car repair and gas more frequently. So, although many lower-income people do seem to have access to cars, they might be less willing to use them daily.

Additionally, I hypothesize that there is a higher probability for higher income brackets to use public transportation to move between the city and the suburbs due to the higher cost of using services such as the commuter rail as compared to the subway within the city.

Results

Below, I will report the log odds the result from the model to analyze the signs and significance.

Table 4: Table 4: Computed log odds from multinomial logistic regression

y.level	term	estimate	std.error	statistic	p.value
Bicycle	(Intercept)	-5.2907024	0.3337612	-1.585176e+01	0.0000000
Bicycle	HHINCOME_10000	-0.0345716	0.0240976	-1.434651e+00	0.1513867
Bicycle	HOUSE_TO_WORK_NAMECity to city	0.9421549	0.3462062	2.721369e+00	0.0065012
Bicycle	HOUSE_TO_WORK_NAMECity to suburbs	1.5021121	0.0003934	3.818462e+03	0.0000000
Bicycle	HOUSE_TO_WORK_NAMESuburbs to city	-0.1788708	0.4511608	-3.964679e-01	0.6917599
Bicycle	HHINCOME_10000:HOUSE_TO_WORK_NAMECity to city	0.0501856	0.0244945	2.048855e+00	0.0404763
Bicycle	HHINCOME_10000:HOUSE_TO_WORK_NAMECity to suburbs	-12.9935963	0.0000693	-1.873999e+05	0.0000000
Bicycle	HHINCOME_10000:HOUSE_TO_WORK_NAMESuburbs to city	0.0195652	0.0301698	6.485030e-01	0.5166597
Carpool	(Intercept)	-2.3237009	0.0596592	-3.894961e+01	0.0000000
Carpool	HHINCOME_10000	-0.0050207	0.0033828	-1.484170e+00	0.1377638
Carpool	HOUSE_TO_WORK_NAMECity to city	0.3006487	0.0719133	4.180712e+00	0.0000291
Carpool	HOUSE_TO_WORK_NAMECity to suburbs	0.6032771	0.1949497	3.094527e+00	0.0019713
Carpool	HOUSE_TO_WORK_NAMESuburbs to city	0.1424983	0.0840215	1.695974e+00	0.0898908
Carpool	HHINCOME_10000:HOUSE_TO_WORK_NAMECity to city	-0.0072506	0.0043821	-1.654608e+00	0.0980039
Carpool	HHINCOME_10000:HOUSE_TO_WORK_NAMECity to suburbs	-0.0233861	0.0122950	-1.902083e+00	0.0571604
Carpool	HHINCOME_10000:HOUSE_TO_WORK_NAMESuburbs to city	-0.0071231	0.0048107	-1.480693e+00	0.1386884
Public transport	(Intercept)	-4.1683779	0.1488201	-2.800952e+01	0.0000000
Public transport	HHINCOME_10000	-0.0073894	0.0086858	-8.507373e-01	0.3949153
Public transport	HOUSE_TO_WORK_NAMECity to city	2.1653304	0.1520380	1.424203e+01	0.0000000
Public transport	HOUSE_TO_WORK_NAMECity to suburbs	1.3399181	0.3982824	3.364241e+00	0.0007675
Public transport	HOUSE_TO_WORK_NAMESuburbs to city	2.0682192	0.1541792	1.341439e+01	0.0000000
Public transport	HHINCOME_10000:HOUSE_TO_WORK_NAMECity to city	0.0229336	0.0088270	2.598132e+00	0.0093732
Public transport	HHINCOME_10000:HOUSE_TO_WORK_NAMECity to suburbs	-0.0460199	0.0288597	-1.594610e+00	0.1107995
Public transport	HHINCOME_10000:HOUSE_TO_WORK_NAMESuburbs to city	0.0255449	0.0088390	2.890003e+00	0.0038524

y.level	term	estimate	std.error	statistic	p.value
Walked Only	(Intercept)	-3.1646228	0.1729216	-1.830091e+01	0.0000000
Walked Only	HHINCOME_10000	-0.1070847	0.0170068	-6.296570e+00	0.0000000
Walked Only	HOUSE_TO_WORK_NAMECity to city	0.3101214	0.1786045	1.736358e+00	0.0825006
Walked Only	HOUSE_TO_WORK_NAMECity to suburbs	-0.7507377	0.5956816	-1.260300e+00	0.2075610
Walked Only	HOUSE_TO_WORK_NAMESuburbs to city	-0.1497501	0.2112927	-7.087331e-01	0.4784901
Walked Only	HHINCOME_10000:HOUSE_TO_WORK_NAMECity to city	0.1235989	0.0171427	7.209991e+00	0.0000000
Walked Only	HHINCOME_10000:HOUSE_TO_WORK_NAMECity to suburbs	0.0683736	0.0428308	1.596364e+00	0.1104076
Walked Only	HHINCOME_10000:HOUSE_TO_WORK_NAMESuburbs to city	0.0763869	0.0188486	4.052650e+00	0.0000506

It is interesting to note that all HHINCOME_10000 variables besides the one for "Walking Only" are insignificant. However, I still kept this variable in my model since I expected that as income grows, people are less likely to use other forms of transportation compared with private vehicles that are individually driven due to more access to these vehicles. Other insignificant variables include the category "Suburbs to city" using bicycle. However, the negative sign of the coefficient is expected since the relative odds of using a bike as opposed to a vehicle to travel from the suburbs to the city in reference to traveling just within the suburbs should be lower. However, if we look at the interaction term between Suburbs to city and HHINCOME_10000 for the bicycle, we see another insignificant coefficient with a positive sign indicating that the relative odds of using a bike in this situation increases as unexpected income increases.

We also see man insignificant variables stem from the Carpool option. The signs of the coefficients are expected, however. Carpooling may overall be preferred over using private vehicles when traveling from the suburbs to the city in reference to traveling through just the suburbs. For the interaction terms, it makes sense why carpooling will overall have lower relative odds of being picked over private vehicles as income increases due to increased access.

With the choice of public transit, we see that the interaction term between City to suburbs and household income is insignificant. However, the sign of the coefficient is not unexpected. One could reason that the less robust transportation systems in the suburbs may not be enough to get someone home from the city and thus, the relative odds of using a private vehicle is higher in reference to traveling just within the suburbs.

There are also many variables for the walking choice that are insignificant. Again, the signs seem to match my hypotheses and not seem unexpected. Besides City city traveling, it makes sense why the relative odds of walking in all other travel patterns would be lower than using a private vehicle in reference to traveling within the suburbs. The positive sign for the interaction term between income and City to suburbs is unexpected, however.

Since more of these insignificant variables have signs that make sense in the context of my hypothesis and the data, I will keep them in my model. Having insignificant variables is surprising to me overall due to the large sample size and relatively low standard error.

Since odds are often difficult to interpret, I instead calculated predicted probabilities on a new dataset I created. This dataset consists of data points for each whole value between 0 and 170 to represent 10000 dollar increases in household income for each increase in the value. Each of these discrete values is paired with every combination of TRANWORK_SM and HOUSE_TO_WORK_NAME.

Below, there is a chart that shows the averaged predicted probabilities for each value of the continuous variable HHINCOME_10000 within each category of HOUSE_TO_WORK_NAME.

Table 5:

```
## predicted_prob_hhincome$HOUSE_TO_WORK_NAME: City to city
##  Private vehicle          Bicycle          Carpool Public transport
##       0.52804365       0.02691889       0.03416008       0.27894326
##      Walked Only
##       0.13193412
## ------------------------------------------------------------
## predicted_prob_hhincome$HOUSE_TO_WORK_NAME: City to suburbs
##  Private vehicle          Bicycle          Carpool Public transport
##     0.9582538668     0.0001033258     0.0332054880     0.0057399140
##      Walked Only
##     0.0026974054
## ------------------------------------------------------------
## predicted_prob_hhincome$HOUSE_TO_WORK_NAME: Suburbs to city
##  Private vehicle          Bicycle          Carpool Public transport
##      0.585718815      0.001051294      0.032179290      0.375850504
##      Walked Only
##      0.005200098
## ------------------------------------------------------------
## predicted_prob_hhincome$HOUSE_TO_WORK_NAME: Suburbs to suburbs
##  Private vehicle          Bicycle          Carpool Public transport
##     0.9281519773     0.0007779211     0.0608068972     0.0081188725
##      Walked Only
##     0.0021443319
```

Looking at the average household income may not be that useful, however. We can breakdown our predicted probabilities into lower (less than 30,000 dollars), middle (between 30,000 and 90,000 dollars), and upper (above 90,000 dollars) income brackets (Sommeilier).

Table 6: Lower Income Bracket

```
## predicted_prob_hhincome_lower$HOUSE_TO_WORK_NAME: City to city
##  Private vehicle          Bicycle          Carpool Public transport
##       0.72817183       0.01198185       0.08075356       0.12493237
##      Walked Only
##       0.05416039
## ------------------------------------------------------------
## predicted_prob_hhincome_lower$HOUSE_TO_WORK_NAME: City to suburbs
##  Private vehicle          Bicycle          Carpool Public transport
##     0.8614594141     0.0005699586     0.1028738288     0.0250490643
##      Walked Only
##     0.0100477342
## ------------------------------------------------------------
## predicted_prob_hhincome_lower$HOUSE_TO_WORK_NAME: Suburbs to city
##  Private vehicle          Bicycle          Carpool Public transport
##      0.778399706      0.002644524      0.073738757      0.126645068
##      Walked Only
##      0.018571944
## ------------------------------------------------------------
## predicted_prob_hhincome_lower$HOUSE_TO_WORK_NAME: Suburbs to suburbs
```

```
##  Private vehicle          Bicycle          Carpool Public transport
##      0.892380676      0.002793689      0.081063198      0.012377536
##      Walked Only
##      0.011384901
```

Table 7: Middle Income Bracket

```
## predicted_prob_hhincome_middle$HOUSE_TO_WORK_NAME: City to city
##  Private vehicle          Bicycle          Carpool Public transport
##       0.62304087       0.02084818       0.04110615       0.21666552
##      Walked Only
##       0.09833928
## ------------------------------------------------------------
## predicted_prob_hhincome_middle$HOUSE_TO_WORK_NAME: City to suburbs
##  Private vehicle          Bicycle          Carpool Public transport
##     9.593749e-01    6.141243e-174     3.491864e-02     3.398066e-03
##      Walked Only
##     2.308366e-03
## ------------------------------------------------------------
## predicted_prob_hhincome_middle$HOUSE_TO_WORK_NAME: Suburbs to city
##  Private vehicle          Bicycle          Carpool Public transport
##      0.694723134      0.001253846      0.039296484      0.259949558
##      Walked Only
##      0.004776977
## ------------------------------------------------------------
## predicted_prob_hhincome_middle$HOUSE_TO_WORK_NAME: Suburbs to suburbs
##  Private vehicle          Bicycle          Carpool Public transport
##     0.9227556443     0.0006953536     0.0670631446     0.0092355095
##      Walked Only
##     0.0002503480
```

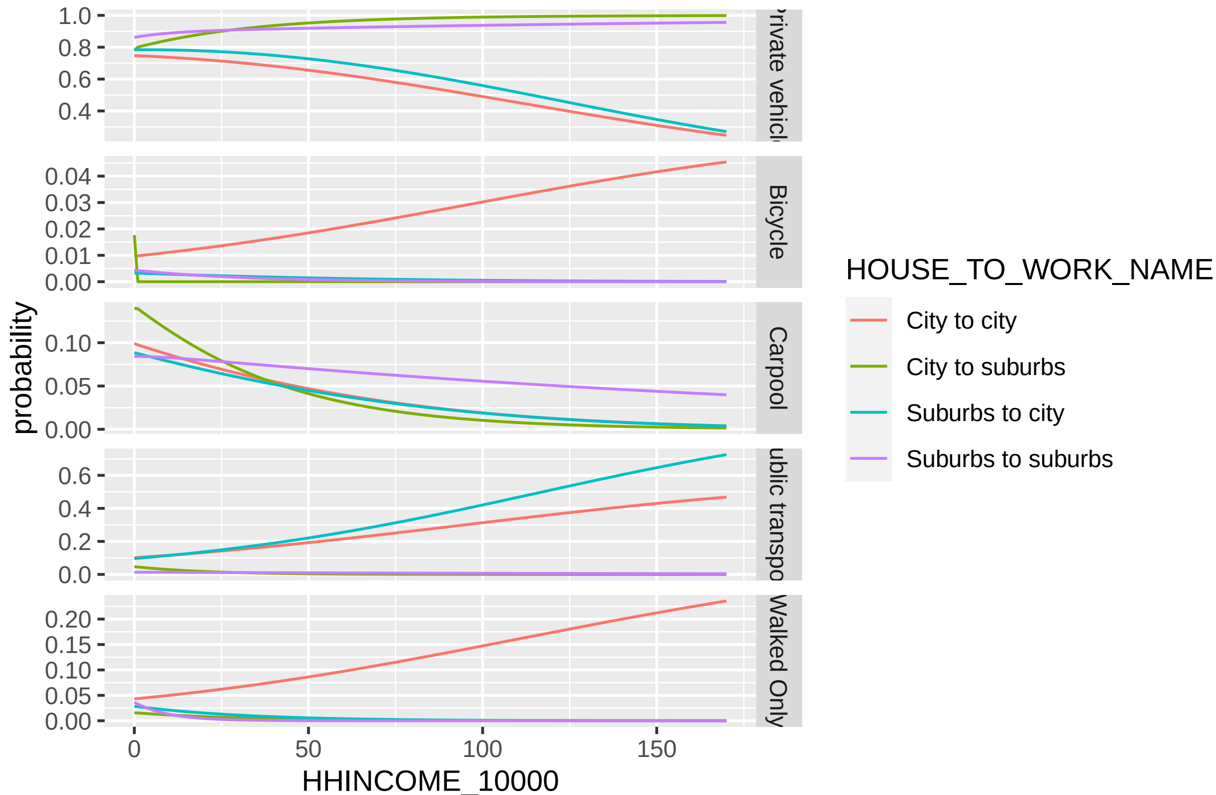
Table 8: Upper Income Bracket

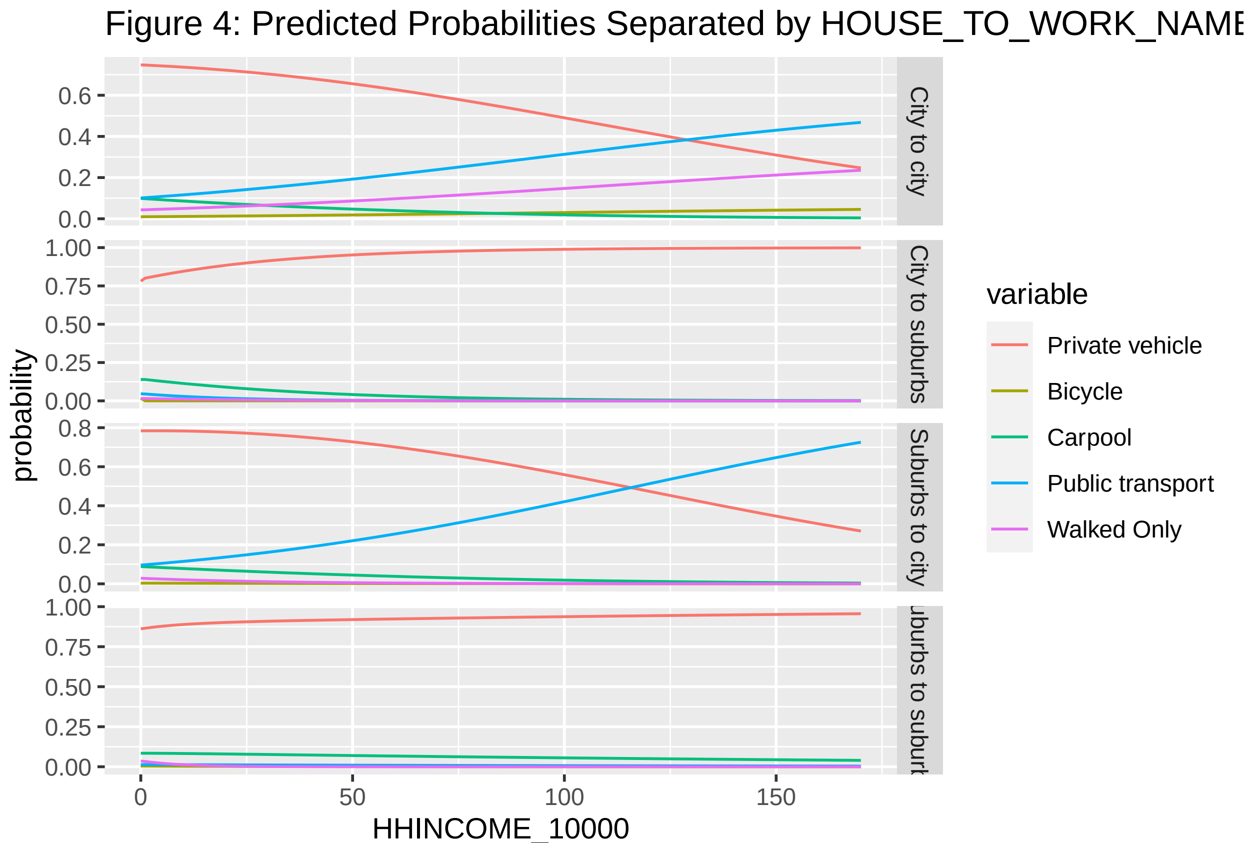
```
## predicted_prob_hhincome_upper$HOUSE_TO_WORK_NAME: City to city
##  Private vehicle          Bicycle          Carpool Public transport
##       0.38206123       0.03706474       0.01133340       0.38332232
##      Walked Only
##       0.18621831
## ------------------------------------------------------------
## predicted_prob_hhincome_upper$HOUSE_TO_WORK_NAME: City to suburbs
##  Private vehicle          Bicycle          Carpool Public transport
##     0.9942372383     0.0000000000     0.0054621839     0.0001120515
##      Walked Only
##     0.0001885262
## ------------------------------------------------------------
## predicted_prob_hhincome_upper$HOUSE_TO_WORK_NAME: Suburbs to city
##  Private vehicle          Bicycle          Carpool Public transport
##     0.4322716597     0.0002964944     0.0111404762     0.5558649004
##      Walked Only
##     0.0004264693
## ------------------------------------------------------------
## predicted_prob_hhincome_upper$HOUSE_TO_WORK_NAME: Suburbs to suburbs
##  Private vehicle          Bicycle          Carpool Public transport
```

```
##     9.457437e-01     7.205610e-05     4.850537e-02     5.678571e-03
##      Walked Only
##     3.137072e-07
```

To analyze the overall trends among different income brackets, I will analze results from a visual representation of these predicted probabilities.

Figure 3: Predicted Probabilities Separated by TRANWORK_SM





Discussion and Conclusion

As discussed earlier, many of the coefficents in this model are insignificant but have the expected directions. Thus, I will be referring to Figures 3 and 4 to analyze the model across different income brackets rather than looking at specific predicted probabilities provided in Tables 5 through 8.

From Figure 4, we see the probability of using of private vehicles when traveling from the city to the suburbs and within the suburbs as income increases as expected.

However, it was unexpected that lower-income brackets seem to be using private vehicles more when traveling from the suburbs to the city or within the city itself.

What is even more interesting is the upper-income bracket's increased probability in using public transportation when traveling from the suburbs to the city compared to the lower and middle-income brackets in Figure 3. In Figure 4, we see that at a certain income level within the higher income bracket, people have a higher probability of using public transit over private vehicles while private vehicles remain the most likely to be used when traveling from the city to the suburbs. On the other hand, the increased probability of using public transportation within the city is expected among the upper-income bracket as explained in the Model and Hypothesis section. With further study, the low probability of using public transportation in the city or from the suburbs to the city among lower-incomes could potentially indicate a lack of accessibility to public transportation for lower incomes.

Walking seems to increase among higher income brackets within the city, potentially indicating increased access to industries and businesses due to home location. However, walking seems to be less preferred compared to individual vehicles and public transportation among higher income brackets.

Finally, as hypothesized, lower-income brackets have the highest probability of using carpooling and the probability decreases as household income increases. We see a less drastic dip in probability as income brackets increase for carpooling when looking at travel within the suburbs only which makes sense due to the lower accessibility to public transit, walking, and biking.

The expected trends for walking and biking further follow what past studies have suspected about the increased accessibility that the wealthy have to businesses within the city when living in it while poorer communities may have to take farther commutes for which biking or walking may be too strenuous (Rice). It could be interesting to further investigate where city biking services are more prominent to see whether they are mainly targeted toward wealthier neighborhoods to promote the notion of sustainability. One could conduct a similar study to look at walkway renovations to study in which neighborhoods walking has become more accessible overtime.

I would also be interested in investigating public transit accessibility in terms of location and price. The increased probability of using public transportation over private vehicles among the wealthy while private vehicles remain the main source of transit among lower classes could be explained by a lack of public transportation services among poorer suburbs and/or lack of affordability.

It should be noted, however, that the probabilities among the lower income brackets may be deflated since they make up a smaller proportion of the population.

It also seems as if increased income may not have as large of an impact on the use of private vehicles. Although having a larger income indeed means that one can afford to use their vehicles daily, public transportation into and in the city and more sustainable forms of transportation such as walking and biking may be more attractive and perhaps more convenient.

This last point brings up another important limitation in this study. I do not establish a metric by which I can judge how people make decisions when it comes to using one transportation type over the other. It could be assumed that each individual is making the “best” choice available to them due to factors such as finances and convenience. However, this is not always the case. It would not be surprising that public transportation simply is not an option in certain poor areas. Thus, even if for some individuals, public transit is their “best” option, they simply would not be able to use it.

To try to combat this issue, a future study could potentially create an “accessibility index” to evaluate the accessibility to “affordable” and “convenient” transportation types in different communities. This index could be compared with similar graphs and charts shown in this study to infer why might people be making the choices they are and whether or not their choice is truly the “best” one available to them.

Bibliography

Lima, Jassica, et al. “Income vs. Travel Time: Why Do the Poorest and Richest Travel Fastest in Northeastern Brazil?” ScienceDirect, 10 July 2016.

Dargay, Joyce, and Jos Van Ommeren. “The Effect of Income on Commuting Time - an Analysis Based on Panel Data.” ECONSTOR, 23 Aug. 2005.

Sommeiller, Estelle, and Mark Price. “The New Gilded Age Income Inequality in the U.S. by State, Metropolitan Area, and County.” Economic Policy Institute, 19 July 2018.

Rice, Jennifer L., et al. “CONTRADICTIONS OF THE CLIMATE-FRIENDLY CITY: New Perspectives on Eco-Gentrification and Housing Justice.” INTERNATIONAL JOURNAL OF URBAN AND REGIONAL RESEARCH.