Data Appendix to “The role of homeownership in transit use in the Allston-Brighton neighborhood of Boston”

Victoria Tse

Contents


```
## Warning: package 'summarytools' was built under R version 3.6.3
```

1 Appendix description

Your Data Appendix should begin with a brief statement explaining its purpose like the following one.

This Data Appendix documents the data used in “The role of homeownership in transit use in the Allston-Brighton neighborhood of Boston”. It was prepared in a Rmarkdown document that contains both the documentation and the R code used to prepare the data used in the final estimation. It also includes descriptive statistics for both the original data and the final dataset, with a discussion of any issues of note.

The datasets used directly by the final analysis are saved in `processed-data/` at the end of this file.

Note: this document structure will require you to re-run steps of your analysis multiple times. If your code takes a long time, please come talk with me about strategies to reduce run time or save earlier results.

2 Instructions for Use

This document includes instructions for how to create your Data Appendix. Outside of this section, instruction paragraphs are listed in *italics* (like the first paragraph above). Instructions should be removed before submission.

To start creating your own data appendix, follow these steps:

1. Replace the title and author in the section at the top of the file (called the YAML).

2. Commit your changes with a message like "customizing data appendix".
3. Delete this instruction section of the document.
4. Remove any other instructions in italics and examples from the completed sections of the document.

Remember that you will submit your assignment by committing and then pushing your versions to your repository. I encourage you to commit your changes often as you work, but there are three specific points at which you need to both submit and push changes, corresponding to course deadlines:

1. You must submit a version with the original data section completed by the Data Appendix 1 deadline. This will include the .Rmd file, the .pdf file, and the html data summary files stored in the output folder.
2. You must submit a version with all parts completed by the Data Appendix 2 deadline.

3. You must submit a final version of this document that is consistent with your final paper by the final project deadline.

While creating your data appendix, refer regularly to the assignment descriptions posted on Moodle.

A few tips:

- When creating a list like this one, be sure to put an empty line above the list. If you don't do this, your entries won't be formatted a list.
- Make sure you have empty lines above and below section and subsection headings.
- When creating numbered lists, you can number all items in your list with 1. Rmarkdown will number them sequentially when it creates your final document.

3 Raw data

Each dataset you use will have its own documentation section. The next subsection in this document (Dataset description) is a template. You can copy this section and paste it into your document each time you need to add a section for a new dataset. Note that each line in the Dataset description section ***must*** *end with two spaces.* This section documents the datasets used in this analysis.

3.1 Dataset description

Citation:

Massachusetts Department of Transportation. (2012, June). Massachusetts Travel Survey.

DOI:

N/A

Date Downloaded:

3/7/20

Filename(s):

- raw_data/filename.csv

- HH.txt
- PER.txt
- PLACE.txt

Unit of observation:

Each unit is an individual person

Dates covered:

6/14/10 - 10/24/11

3.1.1 To obtain a copy

To obtain a copy, be a part of the Smith College institution and ask Professor Sayre! Also, agree to the necessary privacy measures to ensure the safety of the data.

3.1.2 Importable version (if necessary)

N/A

Filename(s):

- raw_data/filename.csv

- HH.txt
- PER.txt
- PLACE.txt
- VEH.txt

3.1.3 Variable descriptions

- **RIBUS:** Use transit on a regular basis during the week

  1 - Yes

  2- No

- **HHVEH:** Number of household vehicles in working condition

  Range 1-7 8 - 8 or more

- **HISP:** Hipanic or Latino

  1- Yes

  2- No

- **RACE:** Ethnicity or race

  1 White alone

  2 Black or African American alone

  3 Indigenous/ Alaska Native alone

  4 Asian alone

  5 Native Hawaiian/ Pacific Islander alone

  6 Some other race alone

  7 Two or more races

- **OWN:** Homeownership

  1 Own/Mortgaged

  2 Renter

- **INCOME:** Household Income in 2009

  1 Less than $15,000

  2 $15,00-$24,999

  3 $25,000-$34,999

  4 $35,000- $49,999

  5 $50,000-$74,999

  6 $75,000-$99,999

  7 $100,00- $149,999

  8 $150,000 or more

- **AGE:** Age in years
- **LIC:** Valid Driver's License

  1- Yes

  2- No

- **TRANS:** Have a transit pass?

  1- Yes

  2- No

- **EMPLY:** Employed

  1- Yes

  2- No

- **EDUCA:** Level of Education

  1- Not a high school graduate, 12th or less (Includes very young children)

  2-High school diploma or GED

  3- Some college credit but no degree

  4- Associate or technical school degree

  5- Bachelor's or undergraduate degree

  6- Graduate Degree (Includes professional degrees, MD, DDs, JD)

- **ENROL:** Student

  1- Yes

  2- No

- **BIKFA:** Bike facilities available at work/school? [ENROL = 1, WORK=1]

  1- Yes

  2- No

3.1.4 Data import code and summary

Once you've described the variables, enter an R chunk by selecting Code -> Insert Chunk, or Ctrl+Alt+I, give it a name to describe the dataset you are importing. After importing, export a dataframe summary using the command.

```r
hh_data <- read_csv(file.path("raw-data","mts-survey-data","HH.txt"), guess_max = 15000)
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   MPO = col_character(),
##   O_RESTY = col_character(),
##   O_OWN = col_character(),
##   HCITY = col_character(),
##   HSTATE = col_character(),
##   HZIP = col_character(),
##   HTOWN = col_character(),
##   HPUMA10 = col_character()
## )

## See spec(...) for full column specifications.
```

```r
per_data <- read_csv(file.path("raw-data","mts-survey-data","PER.txt"), guess_max = 15000)
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   O_PASST = col_character(),
##   O_WKSTAT = col_character(),
##   O_WMODE = col_character(),
##   WTRSB = col_character(),
##   O_FLEXP = col_character(),
##   O_EDUCA = col_character(),
##   O_SCHOL = col_character(),
##   O_SMODE = col_character(),
##   STRSB = col_character(),
##   O_TYPDY = col_character(),
##   O_TYPPL = col_character(),
##   O_NOGO = col_character(),
##   WCITY = col_character(),
##   WSTATE = col_character(),
##   WZIP = col_character(),
##   WTOWN = col_character(),
##   WPUMA10 = col_character(),
##   SCITY = col_character(),
##   SSTATE = col_character(),
##   SZIP = col_character()
##   # ... with 2 more columns
## )

## See spec(...) for full column specifications.
```

```r
place_data <- read_csv(file.path("raw-data","mts-survey-data","PLACE.txt"), guess_max = 15000)
```

```
## Parsed with column specification:
## cols(
```

```
##   .default = col_double(),
##   O_TPURP = col_character(),
##   O_TPUR2 = col_logical(),
##   O_MODE = col_character(),
##   O_PRKTY = col_character(),
##   PRKLC = col_character(),
##   HOV = col_character(),
##   O_FARE = col_character(),
##   ROUTE = col_character(),
##   CITY = col_character(),
##   STATE = col_character(),
##   ZIP = col_character(),
##   TOWN = col_character(),
##   PUMA10 = col_character()
## )

## See spec(...) for full column specifications.

## Warning: 46 parsing failures.
##   row     col           expected       actual                                 file
## 36226 O_TPUR2 1/0/T/F/TRUE/FALSE VOLUNTEERING 'raw-data/mts-survey-data/PLACE.txt'
## 42538 O_PRKCS a double           DK/RF        'raw-data/mts-survey-data/PLACE.txt'
## 43711 O_PRKCS a double           DK/RF        'raw-data/mts-survey-data/PLACE.txt'
## 57718 O_PRKCS a double           DK/RF        'raw-data/mts-survey-data/PLACE.txt'
## 61029 O_PRKCS a double           DK/RF        'raw-data/mts-survey-data/PLACE.txt'
## ..... ....... .................. ............ ...................................
## See problems(...) for more details.
```

```
export_summary_table(dfSummary(dataset_name))
```

While it will make your resulting file long, you should not modify the chunk options to suppress printing of code and output. I would likely not include this in the documentation for an actual paper I was submitting, but including them here will let me read your code and the output message from R and may help identify data import concerns early in the process. Since these files will exist only electronically, their length is less of a concern. If you like to print out files to proofread and want me to help you shorten the printed versions, let me know. We can temporarily modify the chunk options for printing and restore them before you submit the assignment.

```
# Data Processing and Combination
*This section should include a discussion of the processing and merging steps needed to create your basi
```

```
Dummy_1 <- merge(hh_data, per_data,
                                 by = "SAMPN")
```

```
Preliminary_data <- merge(Dummy_1, place_data,
                  by ="SAMPN")
```

```
  Final_data_half1 <- Preliminary_data[ which(Preliminary_data$ZIP=='02134'), ]
```

```
  Final_data_half2 <- Preliminary_data[ which(Preliminary_data$ZIP=='02135'), ]
```

```
Final_data <- rbind(Final_data_half1, Final_data_half2)
```

```
Modified_final_data<- Final_data %>%

  mutate(RIBUS_factor = as.factor(case_when(RIBUS == 1 ~ "Ride", RIBUS== 2 ~ "Doesn't_Ride")))
```

```
Modified_final_data<- Final_data %>%
   mutate(HISP_factor = as.factor(case_when(HISP == 1 ~ "Hispanic", HISP== 2 ~ "Non_Hispanic")))
```

```
Modified_final_data<- Final_data %>%
    mutate(OWN_factor = as.factor(case_when(OWN == 1 ~ "Homeowner", OWN== 2 ~ "Non_Homeowner")))
```

```
Modified_final_data<- Final_data %>%
  mutate(LIC_factor = as.factor(case_when(LIC == 1 ~ "License", LIC== 2 ~ "No_License")))
```

```
Modified_final_data<- Final_data %>%
  mutate(TRANS_factor = as.factor(case_when(TRANS== 1 ~ "Pass", TRANS== 2 ~ "No_Pass")))
```

```
Modified_final_data<- Final_data %>%
   mutate(TRANS_factor = as.factor(case_when(TRANS== 1 ~ "Pass", TRANS== 2 ~ "No_Pass")))
```

```
Modified_final_data<- Final_data %>%
  mutate(ENROL_factor = as.factor(case_when(ENROL == 1 ~ "Student", WORKS== 2 ~ "Non_Student")))
```

```
Modified_final_data<- Final_data %>%
  mutate(BIKFA_factor = as.factor(case_when(BIKFA == 1 ~ "Bike_Accessible", BIKFA == 2 ~ "No_Bike_Access
```

```
Modified_final_data<- Final_data %>%
  mutate(RACE_factor = as.factor(case_when(RACE == 1 ~ "White", RACE== 2 ~ "Black", RACE== 3 ~ "Indigen
```

```
Modified_final_data<- Final_data %>%
  mutate(INCOME_factor = as.factor(case_when(INCOME == 1 ~ "15", INCOME== 2 ~ "15-24", INCOME== 3 ~ "25
```

```
Modified_final_data<- Final_data %>%
   mutate(EDUCA_factor = as.factor(case_when(EDUCA== 1 ~ "K-12", EDUCA== 2 ~ "HS-GED", EDUCA== 3 ~ "Som
```

```
Summary_statistics <- Modified_final_data %>%
  select(RIBUS, OWN, HHVEH, LIC, TRANS, BIKFA, RACE, HISP, INCOME, EDUCA, ENROL, EMPLY, AGE)

stargazer(Summary_statistics, type = "latex",
          title = "Summary Statistics",
          header = F,
          digits = 2)
```

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
RIBUS	4,206	1.34	0.47	1	1	2	2
OWN	4,206	1.37	0.64	1	1	2	7
HHVEH	4,206	1.36	0.95	0	1	2	5
LIC	3,138	1.20	0.52	1.00	1.00	1.00	9.00
TRANS	3,138	1.46	0.77	1.00	1.00	2.00	9.00
BIKFA	3,287	2.48	2.43	1.00	1.00	2.00	9.00
RACE	4,206	1.84	1.95	1	1	1	9
HISP	4,206	1.96	0.45	1	2	2	9
INCOME	4,206	13.42	25.73	1	5	7	99
EDUCA	4,206	3.77	2.16	1	1	6	9
ENROL	4,206	1.74	0.69	1	1	2	9
EMPLY	3,138	1.27	0.44	1.00	1.00	2.00	2.00
AGE	4,206	36.98	23.93	0	15	54	99

```
Regression_1 <- lm(RIBUS ~ factor(OWN), data = Modified_final_data)

stargazer(Regression_1, type = "latex",
          report = 'vct',
          intercept.bottom = F,
          header = F)
```

Table 2:

	Dependent variable:
	RIBUS
Constant	1.409
	t = 160.764
factor(OWN)2	−0.211
	t = −13.973
factor(OWN)7	0.591
	t = 6.254
Observations	4,206
R^2	0.055
Adjusted R^2	0.055
Residual Std. Error	0.461 (df = 4203)
F Statistic	122.235*** (df = 2; 4203)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
Regression_2 <- lm(RIBUS ~ factor(OWN) + HHVEH + factor(LIC) + factor(TRANS) + factor(BIKFA), data = Mod

stargazer(Regression_2, type = "latex",
          report = 'vct',
```

```
          intercept.bottom = F,
          header = F)
```

Table 3:

	Dependent variable:
	RIBUS
Constant	1.209
	t = 51.491
factor(OWN)2	−0.143
	t = −7.012
factor(OWN)7	0.424
	t = 3.935
HHVEH	0.074
	t = 7.197
factor(LIC)2	−0.007
	t = −0.269
factor(LIC)9	−0.152
	t = −0.867
factor(TRANS)2	0.250
	t = 12.902
factor(TRANS)8	0.468
	t = 4.984
factor(TRANS)9	−0.358
	t = −0.841
factor(BIKFA)2	−0.031
	t = −1.635
factor(BIKFA)8	−0.131
	t = −4.937
factor(BIKFA)9	0.223
	t = 3.134
Observations	2,462
R^2	0.202
Adjusted R^2	0.199
Residual Std. Error	0.425 (df = 2450)
F Statistic	56.537*** (df = 11; 2450)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
Regression_3 <- lm(RIBUS ~ factor(OWN) + factor(RACE) + factor(HISP), data = Modified_final_data)

stargazer(Regression_3, type = "latex",
          report = 'vct',
          intercept.bottom = F,
          header = F)
```

Table 4:

	Dependent variable:
	RIBUS
Constant	1.682
	t = 43.321
factor(OWN)2	−0.241
	t = −15.289
factor(OWN)7	0.460
	t = 4.748
factor(RACE)2	−0.225
	t = −5.743
factor(RACE)4	0.117
	t = 3.946
factor(RACE)6	−0.251
	t = −6.106
factor(RACE)7	0.014
	t = 0.329
factor(RACE)9	−0.296
	t = −6.499
factor(HISP)2	−0.259
	t = −6.955
factor(HISP)9	−0.386
	t = −2.694
Observations	4,206
R^2	0.090
Adjusted R^2	0.088
Residual Std. Error	0.453 (df = 4196)
F Statistic	45.989*** (df = 9; 4196)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
Regression_4 <- lm(RIBUS ~  factor(OWN) + factor(EDUCA), data = Modified_final_data)
```

```r
stargazer(Regression_4, type = "latex",
          report = 'vct',
          intercept.bottom = F,
          header = F)
```

Table 5:

	Dependent variable:
	RIBUS
Constant	1.403
	t = 103.460
factor(OWN)2	−0.217
	t = −14.248
factor(OWN)7	0.644
	t = 6.757
factor(EDUCA)2	−0.070
	t = −2.484
factor(EDUCA)3	−0.028
	t = −0.870
factor(EDUCA)4	0.094
	t = 2.275
factor(EDUCA)5	0.089
	t = 4.436
factor(EDUCA)6	−0.029
	t = −1.574
factor(EDUCA)8	−0.349
	t = −1.519
factor(EDUCA)9	0.121
	t = 1.678
Observations	4,206
R^2	0.068
Adjusted R^2	0.066
Residual Std. Error	0.458 (df = 4196)
F Statistic	34.010*** (df = 9; 4196)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
Regression_5 <- lm(RIBUS ~  factor(OWN) + factor(INCOME), data = Modified_final_data)

stargazer(Regression_5, type = "latex",
          report = 'vct',
```

```r
           intercept.bottom = F,
           header = F)
```

Table 6:

	Dependent variable:
	RIBUS
Constant	1.402
	t = 36.028
factor(OWN)2	−0.196
	t = −11.204
factor(OWN)7	0.595
	t = 6.082
factor(INCOME)2	0.105
	t = 2.243
factor(INCOME)3	0.001
	t = 0.010
factor(INCOME)4	0.004
	t = 0.079
factor(INCOME)5	−0.156
	t = −3.813
factor(INCOME)6	0.014
	t = 0.362
factor(INCOME)7	−0.034
	t = −0.834
factor(INCOME)8	0.117
	t = 2.720
factor(INCOME)99	0.093
	t = 2.097
Observations	4,206
R^2	0.084
Adjusted R^2	0.082
Residual Std. Error	0.454 (df = 4195)
F Statistic	38.580*** (df = 10; 4195)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
Regression_6 <- lm(RIBUS ~  factor(OWN) + factor(EMPLY) + factor(ENROL), data = Modified_final_data)

stargazer(Regression_6, type = "latex",
```

```
          report = 'vct',
          intercept.bottom = F,
          header = F)
```

Table 7:

	Dependent variable:
	RIBUS
Constant	1.360
	t = 53.623
factor(OWN)2	−0.238
	t = −14.072
factor(OWN)7	0.567
	t = 4.984
factor(EMPLY)2	−0.002
	t = −0.091
factor(ENROL)2	0.073
	t = 2.987
factor(ENROL)8	−0.014
	t = −0.104
factor(ENROL)9	−0.360
	t = −2.700
Observations	3,138
R^2	0.078
Adjusted R^2	0.077
Residual Std. Error	0.453 (df = 3131)
F Statistic	44.337*** (df = 6; 3131)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
Regression_8 <- lm(RIBUS ~  factor(OWN) + AGE, data = Modified_final_data)

stargazer(Regression_8, type = "latex",
          report = 'vct',
          intercept.bottom = F,
          header = F)
```

4 Analysis Variables

The dependent variable of the seven regressions given is RIBUS. Massachusetts Travel Survey describes RIBUS as "use transit on a regular basis during the week." While "regular basis" is undefined, we can conclude that individuals who identified "yes" for RIBUS use public transit at least weekly. Unlike most

Table 8:

	Dependent variable:
	RIBUS
Constant	1.354
	t = 94.733
factor(OWN)2	−0.206
	t = −13.682
factor(OWN)7	0.603
	t = 6.391
AGE	0.001
	t = 4.818
Observations	4,206
R^2	0.060
Adjusted R^2	0.059
Residual Std. Error	0.460 (df = 4202)
F Statistic	89.657*** (df = 3; 4202)
Note:	*p<0.1; **p<0.05; ***p<0.01

indicator variables, "1" indicates "yes" and "2" indicates "no". This applies not only for RIBUS, but or all the indicator variables used. Therefore, when interpreting coefficient, a positive coefficient is one that moves closer to "no", indicating that the individual is less likely to use public transit regularly during the week. A negative coefficient moves closer to "yes", indicating that the individual is more likely to regularly use public transit.

OWN is the homeownership variable whose coefficient we pay close attention to throughout the eight regressions. As stated before, OWN is an indicator variable in which "1" signifies ownership and "2" denotes rentership. "7" indicate other, which we disregard so as to focus on the dynamics between homeowning and renting. We hypothesize that renters are more likely to use public transit than homeowners for a variety of reasons. The main reason is that renters tend to be more temporary members of the community, especially in Allston-Brighton, as many of the renters in this area are Boston University students. This lack of permanency could indicate not having access to cars or other services that could be used in lieu of public transit.

The variables used in Regression 2 speak more to the survey individual's access to transportation. HHVEH is a measure of household vehicles, ranging from 0 to 7. Anything above 8 is specified in a different variable, however, assuming diminishing returns, we have chosen to disregard households with eight or more vehicles. Next is LIC, which is an indicator variable of a driver's license, "1" signifying "yes" and "2" signifying "no". TRANS indicates the individual having a transit pass, again, "1" signifying "yes" and "2" signifying "no". Finally, BIKFA examines whether or not bike facilities are available at an individuals school or work, "1" signifying "yes" and "2" signifying "no". These variables were included to account for an individual's access to diferent modes of transit. We predict that those with a value of "1" for TRANS will prove more like to be using public transit regularly, as the pass provides greater access and might also be an indicator of access. Those with a value of "1" for LIC and BIKFA as well as a value of 1 or higher for HHVEH are predicted to be less likely to use public transit, as they indicated access to other modes of transportation.

Regression 3 variables look at race and ethnicity. For variable RACE, a value of "1" signifies "White alone". "2" means "Black or African American alone", "3" "Indigenous/ Alaska Native alone","4" "Asian alone", "5" "Native Hawaiian/ Pacific Islander alone", "6" "Some other race alone", and "7" "two or more races".

There is also a HISP indicator variable, with a “1” denoting “yes” and a 2 denoting “no”. Race, ethnicity, and socioeconomic standing are very intertwined and in many cities, public transit is widely used by people of color, so it is likely that non white groups may demonstrate a greater likelihood of using public transit. However, given that this neighborhood is known for housing college students, this particularly nuance could potentially throw off predicted assumptions.

Regressions 4 and 5 cover education and income respectively. EDUCA is a factor categorical variable with six groups, with each group indicating a different level of education. Similarly, INCOME is factor categorical variable consisting of right groups, ranging from under $15,000 per year to $150,00+ per year. Education and income are closely interwined. One might assume that those who are higher educate and have higher income are less likely to regularly use public transit. This makes logical sense, as those with more disposable income have access to more expensive forms of transportation. However, given how transit accessible the main business district in Boston is and how difficult it is to find parking in many of the white collar business districts, there could be a reverse effect of those with higher income and education being more likely to take public transit regularly.

Regression 7 looks at two indicator variables EMPLY and ENROL. EMPLY is an indicator variable for whether or not an individual is employed, “1” meaning “yes” and “2” meaning “no”. ENROL is an indicator variable for whether or not an individual is a student, “1” as “yes” and “2” as “no”. We predict that students and workers are more likely to regularly use public transit as means of commuting.

Finally, Regression at the effect of age through variable AGE. Age is a continuous variable by year. We predict that younger individuals are more likely to be using public transit than older individuals. (This might be most effect to either include and age squared variable or categorically set groups. This may be added to the final draft.)

5 Discussion of Data

Regression 1 looks solely at the relationship between regular weekly transit use (RIBUS) and homeownership (OWN). OWN is set to be a factor categorical variable, which when indicated in the regression, calls for one of the categories to be dropped. In the case of OWN, we observe that when OWN = “2” or “no”, the corresponding coefficient is -0.211, which is noted with statistical significance. This means that if an individual is a renter, the chance of them using public transit regularly increases by 21.1%, which aligns with our hypothesis. However, since this regression only includes the OWN variable, it is very likely that this coefficient is not the true coefficient, as many other explanatory factors are not included, which is why we ran seven other regressions to observe how the OWN coefficient changed.

In Regression 2, we regress RIBUS onto OWN, as well as LiC (driver’s license), TRANS (transit pass), and BIKFA (bike facilities accessible). This regression was meant to assess the effect of an individual’s access to various modes of transport on the OWN coefficient. We predicted that, when controlling for this access, the magnitude of the coefficient for homeownership should decrease. Our prediction was correct, as this regression saw the greatest decrease in magnitude of all of the six regressions that controlled for other factors. The coefficient’s absolute value decreased from -0.211 to -0.114, meaning that with this update regression, renters are now 11.4% more likely to report using public transit regularly than homeowners. This finding was statistically significant. As predicted as, the presence of an additional household vehicle (HHVEH) and well as the possession of a driver’s license (LIC) and access to bike facilities (BIKFA), decreased the likelihood of reporting “yes” to regularly using public transit. Similarly, if individuals reported having a transit pass (TRANS), they were more likely to report “yes” to regularly using transit.

Regression 3 focused the effects of race and ethnicity on homeownership, regressing RIBUS on OWN, RACE, and HISP. RACE is a factor categorical variable that encompasses seven race categories and HISP is a Hispanic indicator variable. The presence of the race variables increased the magnitude of the coefficient very slightly from -0.211 to -0.241. Though statistically significant, the results show that controlling for race in this regression has a small but positive effect on the rentership coefficient. Black identifying participants show a coefficient of -0.225, meaning that they are 22.5% more likely than white participants to report regularly using public transit.

Regression 4 controlled for education (EDUCA). EDUCA was yet another factor variable with nine categories representing levels of education ranging from no high school diploma to graduate diploma. The inclusion of the EDUCA variable only changes the homeownership coefficient magnitude by a very small amount, from -0.211 to -0.217. Many of the groups showed statistically insignificant results.

Regression 5 controls for income (INCOME). With the inclusion of the INCOME variable, the OWN coefficient sees a relatively small decrease in magnitude, from -0.211 to -0.196, which we noted with statistical significance. Most income groups above the lowest income group (under \$15,000) showed a lower likelihood of reporting “yes” to regular transit use. The exceptions were income groups 5 (\$50,000-\$74,999) and 7 (\$100,000-\$149,000), although income group 7, along with 6 and 3, were not statistically significant.

Regression 6 regressed RIBUS onto OWN as well as EMPLY (worker indicator variable) and ENROL (student indicator variable). The presence of these two variables changed the maginatude of the OWN coefficient slightly from -0.211 to -0.238. While the EMPLY coefficient found a very weak and staistically insignificant relationship, the student coefficient saw a statistically significant, but very small coefficient. The regression found that those who are not identified as students are 7.3% more likely to report “no” for regular transit use.

Regression seven controlled for AGE, age being a continuous variable by year. The OWN coefficient decreased very slightly in magnitude from -0.211 to -0.206 with statistical significance. The AGE coefficient was a statistically significant value of 0.001, meaning that very every year of age, an individual is 0.1% more likely to report “no” for regular public transit use. (PERHAPS adding an age squared or making age categorical can help with better results.)

For my reader: Do you think it would be a good use of my time to change the 1 as “yes” and 2 as “no” scale to the traditional 0 as “no” and 1 as “yes”? Is interpreting this too confusing? Also, I am considering doing summary statistics for homeowners and renters so that my conclusions can be less speculative. Additionally, as I have mentioned, I am also thinking of altering the age category to get better results. Let me know what you think of these ideas!