Assignment 5

Miguel Perez Luna

10/5/2020

```r
library(tidyverse)
```

```
## -- Attaching packages -----------------------------------------------------------------------

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts -----------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(jtools)
library(interactions)
library(knitr)
```

Reloading my original linear model from Assignment 4

```r
transpo_data <- read.csv("transpo_data.csv") %>%
  mutate(meansTW = case_when(
    JWTR_label=="Bicycle"~"Bicycle",
    JWTR_label=="Bus or trolley bus"~"Bus or\ntrolley bus",
    JWTR_label=="Car, truck, or van"~"1Car, truck,\nor van",
    JWTR_label=="Motorcycle"~"Motorcycle",
    JWTR_label=="Other method"~"Other",
    JWTR_label=="Railroad"~"Railroad",
    JWTR_label=="Streetcar or trolley car (carro publico in Puerto Rico)"~"Streetcar or\ntrolley car",
    JWTR_label=="Subway or elevated"~"Subway or\nelevated",
    JWTR_label=="Taxicab"~"Taxicab",
    JWTR_label=="Walked"~"Walk")) %>%
  mutate(edu = case_when(
    SCHL_label=="Some college, but less than 1 year"~"Less 1 year of college",
    SCHL_label=="Regular high school diploma"~"1High school diploma",
    SCHL_label=="Bachelor's degree"~"Bachelor's degree",
    SCHL_label=="1 or more years of college credit, no degree"~"More 1 yr of college,\nno degree",
    SCHL_label=="Master's degree"~"Master's degree",
    SCHL_label=="GED or alternative credential"~"GED",
```

```r
    SCHL_label=="Doctorate degree"~"Doctorate degree",
    SCHL_label=="Associate's degree"~"Associate's degree",
    SCHL_label=="Grade 8"~"Grade 8",
    SCHL_label=="Grade 7"~"Grade 7",
    SCHL_label=="Grade 10"~"Grade 10",
    SCHL_label=="12th grade - no diploma"~"Grade 12 -\nno diploma",
    SCHL_label=="Grade 6"~"Grade 6",
    SCHL_label=="Grade 11"~"Grade 11",
    SCHL_label=="Professional degree beyond a bachelor's degree"~"Professional degree",
    SCHL_label=="Grade 9"~"Grade 9",
    SCHL_label=="No schooling completed"~"No schooling",
    SCHL_label=="Grade 5"~"Grade 5",
    SCHL_label=="Grade 4"~"Grade 4",
    SCHL_label=="Nursery school, preschool"~"Preschool",
    SCHL_label=="Grade 3"~"Grade 3",
    SCHL_label=="Grade 1"~"Grade 1",
    SCHL_label=="Kindergarten"~"Kindergarten",
    SCHL_label=="Grade 2"~"Grade 2"))

model <- lm(PINCP ~ SEX_label + meansTW + edu + JWMNP + vehicle + GRNTP + AGEP,
           data = transpo_data)
options(scipen = 999)
summary(model)
```

```
##
## Call:
## lm(formula = PINCP ~ SEX_label + meansTW + edu + JWMNP + vehicle +
##     GRNTP + AGEP, data = transpo_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -106817  -17346   -3709   11007  469745
##
## Coefficients:
##                                       Estimate  Std. Error t value
## (Intercept)                        -13892.1871   2217.9958  -6.263
## SEX_labelMale                       12660.7552    883.5243  14.330
## meansTWBicycle                     -12587.9144   3254.9841  -3.867
## meansTWBus or\ntrolley bus          -8378.8701   2599.6721  -3.223
## meansTWMotorcycle                    -358.6898   9079.4633  -0.040
## meansTWOther                         2407.8624   4677.2547   0.515
## meansTWRailroad                      4135.6233   6445.9454   0.642
## meansTWStreetcar or\ntrolley car     3528.8252  11489.2225   0.307
## meansTWSubway or\nelevated          -5633.2840   6253.3617  -0.901
## meansTWTaxicab                      -2357.8065   8343.2471  -0.283
## meansTWWalk                         -3823.7025   2043.6837  -1.871
## eduAssociate's degree                5033.8675   1816.2136   2.772
## eduBachelor's degree                18153.2293   1376.4953  13.188
## eduDoctorate degree                 25053.5591   3828.7967   6.543
## eduGED                               2293.3529   2321.6023   0.988
## eduGrade 1                          13149.4021  25631.7847   0.513
## eduGrade 10                         -7968.1836   3946.6425  -2.019
## eduGrade 11                         -6731.8987   3143.6728  -2.141
```

```
## eduGrade 12 -\nno diploma            2566.6930   3584.7370   0.716
## eduGrade 2                          -25435.4778  25640.2934  -0.992
## eduGrade 3                           -2292.4248  20947.5795  -0.109
## eduGrade 4                           -6182.5758  20964.2283  -0.295
## eduGrade 5                           -1099.9128  14845.7144  -0.074
## eduGrade 6                           -6435.2441   5141.6076  -1.252
## eduGrade 7                          -10891.5747   9423.6588  -1.156
## eduGrade 8                           -4502.4662   7471.4426  -0.603
## eduGrade 9                           -6764.9026   4955.2926  -1.365
## eduKindergarten                     -34218.5166  25640.8474  -1.335
## eduLess 1 year of college             4572.3733   1977.4870   2.312
## eduMaster's degree                   25879.3543   1879.1949  13.772
## eduMore 1 yr of college,\nno degree   2966.3574   1499.8275   1.978
## eduNo schooling                      -5635.8698   5237.6577  -1.076
## eduPreschool                        -13198.9113  36245.7783  -0.364
## eduProfessional degree               62647.4429   3668.0107  17.079
## JWMNP                                   42.6007     20.5903   2.069
## vehicle                              -2917.3641    449.7630  -6.486
## GRNTP                                   15.2303      0.8123  18.751
## AGEP                                   652.8961     35.1781  18.560
##                                                   Pr(>|t|)
## (Intercept)                            0.0000000003994 ***
## SEX_labelMale                       < 0.0000000000000002 ***
## meansTWBicycle                                0.000111 ***
## meansTWBus or\ntrolley bus                    0.001274 **
## meansTWMotorcycle                             0.968488
## meansTWOther                                  0.606708
## meansTWRailroad                               0.521164
## meansTWStreetcar or\ntrolley car              0.758744
## meansTWSubway or\nelevated                    0.367704
## meansTWTaxicab                                0.777492
## meansTWWalk                                   0.061390 .
## eduAssociate's degree                         0.005593 **
## eduBachelor's degree                < 0.0000000000000002 ***
## eduDoctorate degree                    0.0000000000644 ***
## eduGED                                        0.323270
## eduGrade 1                                    0.607960
## eduGrade 10                                   0.043528 *
## eduGrade 11                                   0.032276 *
## eduGrade 12 -\nno diploma                     0.474012
## eduGrade 2                                    0.321227
## eduGrade 3                                    0.912860
## eduGrade 4                                    0.768071
## eduGrade 5                                    0.940941
## eduGrade 6                                    0.210758
## eduGrade 7                                    0.247816
## eduGrade 8                                    0.546779
## eduGrade 9                                    0.172239
## eduKindergarten                               0.182074
## eduLess 1 year of college                     0.020795 *
## eduMaster's degree                  < 0.0000000000000002 ***
## eduMore 1 yr of college,\nno degree           0.047991 *
## eduNo schooling                               0.281952
## eduPreschool                                  0.715757
```

```
## eduProfessional degree              < 0.0000000000000002 ***
## JWMNP                                          0.038586 *
## vehicle                                  0.0000000000940 ***
## GRNTP                               < 0.0000000000000002 ***
## AGEP                                < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36210 on 6876 degrees of freedom
## Multiple R-squared:  0.2159, Adjusted R-squared:  0.2117
## F-statistic: 51.18 on 37 and 6876 DF,  p-value: < 0.00000000000000022
```

Model 1: Reconsindering the relationships between my variables

In my linear model from Assignment 4, I examined to what extent each of my variables correlated with Income. I treated Income as my dependent variable, and the rest as independent variables. However, I think Income itself, at least in the set of my variables, can function as an independent variable.

More precisely, I can see the 'most' independent variables being Sex, Age, and Educational Attainment (of course, other factors can influence educational attainment, but in this case I'm treating this as a fairly 'fixed' variable). I'm assuming that these three have the biggest effect on an individual's Income. Then, I can see how Income can influence an individual's monthly Rent, Number of Vehicles Accessible, and Means of Transportation to work. Lastly, the three of these variables can influence Travel Time to Work. Certainly, any and all of the variables can correlate with Travel Time to Work, but this is the general chain of influence I've thought of.

To this end, I am going to examine the effect that these variables have on Travel Time to Work. I will keep the mutated variables above (comparing the change in education to receiving a high school diploma, and using a car, truck, or van as a means of travel to work.

A regular linear regression is below:

```r
model2 <- lm(JWMNP ~ SEX_label + meansTW + edu + PINCP + vehicle + GRNTP + AGEP,
            data = transpo_data)
options(scipen = 999)
summary(model2)
```

```
##
## Call:
## lm(formula = JWMNP ~ SEX_label + meansTW + edu + PINCP + vehicle +
##     GRNTP + AGEP, data = transpo_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.532 -12.227  -4.787   6.096 143.049
##
## Coefficients:
##                                       Estimate    Std. Error t value
## (Intercept)                        17.089702533   1.285946691  13.290
## SEX_labelMale                       1.812345329   0.524524348   3.455
## meansTWBicycle                     -6.354531842   1.906356007  -3.333
## meansTWBus or\ntrolley bus         22.416406429   1.499102660  14.953
## meansTWMotorcycle                  -7.202296257   5.315403092  -1.355
## meansTWOther                       14.821254811   2.732790936   5.423
```

```
## meansTWRailroad                    15.089272725   3.769886055   4.003
## meansTWStreetcar or\ntrolley car     26.351585374   6.719584763   3.922
## meansTWSubway or\nelevated           19.175564568   3.654309207   5.247
## meansTWTaxicab                     -7.811200906   4.884170555  -1.599
## meansTWWalk                        -8.156246458   1.192852330  -6.838
## eduAssociate's degree              -0.127063824   1.064003080  -0.119
## eduBachelor's degree               -2.287942325   0.815613849  -2.805
## eduDoctorate degree                -2.257654179   2.248601068  -1.004
## eduGED                              2.279653501   1.359138637   1.677
## eduGrade 1                         -4.149462019  15.007857404  -0.276
## eduGrade 10                        -2.240658365   2.311323635  -0.969
## eduGrade 11                        -0.825443603   1.841237099  -0.448
## eduGrade 12 -\nno diploma           1.185810570   2.098927150   0.565
## eduGrade 2                         -6.139047140  15.013527249  -0.409
## eduGrade 3                         -1.379969667  12.265006317  -0.113
## eduGrade 4                        -14.100603057  12.273654716  -1.149
## eduGrade 5                         -5.910371651   8.692018178  -0.680
## eduGrade 6                          2.028721427   3.010703641   0.674
## eduGrade 7                         -5.320410809   5.517804714  -0.964
## eduGrade 8                         -2.836025164   4.374582510  -0.648
## eduGrade 9                          2.275068633   2.901634378   0.784
## eduKindergarten                    18.510954433  15.013244494   1.233
## eduLess 1 year of college           0.521940179   1.158270355   0.451
## eduMaster's degree                 -1.984745366   1.115100523  -1.780
## eduMore 1 yr of college,\nno degree  -1.508533730   0.878224627  -1.718
## eduNo schooling                     2.888296099   3.066758740   0.942
## eduPreschool                       -2.645341913  21.222428433  -0.125
## eduProfessional degree             -4.689161092   2.192008534  -2.139
## PINCP                               0.000014604   0.000007059   2.069
## vehicle                             0.644859034   0.264030442   2.442
## GRNTP                               0.002084342   0.000486945   4.280
## AGEP                                0.065316745   0.021092038   3.097
##                                              Pr(>|t|)
## (Intercept)                       < 0.0000000000000002 ***
## SEX_labelMale                                0.000553 ***
## meansTWBicycle                               0.000863 ***
## meansTWBus or\ntrolley bus         < 0.0000000000000002 ***
## meansTWMotorcycle                            0.175467
## meansTWOther                         0.00000006043718 ***
## meansTWRailroad                      0.00006331393881 ***
## meansTWStreetcar or\ntrolley car     0.00008881514838 ***
## meansTWSubway or\nelevated           0.00000015888690 ***
## meansTWTaxicab                               0.109802
## meansTWWalk                          0.00000000000875 ***
## eduAssociate's degree                        0.904946
## eduBachelor's degree                         0.005043 **
## eduDoctorate degree                          0.315401
## eduGED                                       0.093534 .
## eduGrade 1                                   0.782183
## eduGrade 10                                  0.332367
## eduGrade 11                                  0.653944
## eduGrade 12 -\nno diploma                    0.572119
## eduGrade 2                                   0.682625
## eduGrade 3                                   0.910420
```

```
## eduGrade 4                                              0.250657
## eduGrade 5                                              0.496542
## eduGrade 6                                              0.500438
## eduGrade 7                                              0.334967
## eduGrade 8                                              0.516815
## eduGrade 9                                              0.433029
## eduKindergarten                                         0.217627
## eduLess 1 year of college                               0.652277
## eduMaster's degree                                      0.075140 .
## eduMore 1 yr of college,\nno degree                     0.085895 .
## eduNo schooling                                         0.346324
## eduPreschool                                            0.900806
## eduProfessional degree                                  0.032454 *
## PINCP                                                   0.038586 *
## vehicle                                                 0.014616 *
## GRNTP                                           0.00001890497925 ***
## AGEP                                                    0.001964 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.2 on 6876 degrees of freedom
## Multiple R-squared:  0.06604,    Adjusted R-squared:  0.06101
## F-statistic: 13.14 on 37 and 6876 DF,  p-value: < 0.00000000000000022
```

The multiple R-squared value for this model is 0.06101, which means that this model predicts about 6% of the variation in Travel Time to Work in this dataset.

Mutating Education and Income

I've decided to group some of the categorical values of both Educational Attainment, and I've decided to group Income into different intervals. I've decided to group educational into 8 values rather than 24. I'll still compare all categories to earning just a high school diploma.

1. No High School Diploma
2. High School Diploma
3. Some College
4. Associate's Degree
5. Bachelor's Degree
6. Master's Degree
7. Other Professional Degree
8. Doctorate Degree

I've also decided to convert my Income variable (continous) to a categorical variable by breaking up the income distribution (arbitrarily) into different intervals to capture differences between very low wealth, a broad middle class, and extreme wealth. Thank you to Ryan Johnson for helping me with the code for dividing the income into these intervals. I recognize that how I named these intervals may not truly represent where the median income lies. I will compare all values to the "Low to Mid Income", which is where I suggest most of the data lies.

1) Less than $25,000 (Low Income)
2) $25,000 to $49,999.99 (Low to Mid Income)
3) $50,000 to $74,999.99 (Mid Income)
4) $75,000 to $99,999.99 (Mid to High Income)
5) $100,000 and higher (High Income)

```
transpo_data2 <- read_csv("transpo_data.csv") %>%
  mutate(edu_attainment = case_when(
    SCHL_label=="Some college, but less than 1 year"~"Some College",
    SCHL_label=="Regular high school diploma"~"1High school diploma",
    SCHL_label=="Bachelor's degree"~"Bachelor's degree",
    SCHL_label=="1 or more years of college credit, no degree"~"Some College",
    SCHL_label=="Master's degree"~"Master's degree",
    SCHL_label=="GED or alternative credential"~"1High school diploma",
    SCHL_label=="Doctorate degree"~"Doctorate degree",
    SCHL_label=="Associate's degree"~"Associate's degree",
    SCHL_label=="Grade 8"~"No High School Diploma",
    SCHL_label=="Grade 7"~"No High School Diploma",
    SCHL_label=="Grade 10"~"No High School Diploma",
    SCHL_label=="12th grade - no diploma"~"No High School Diploma",
    SCHL_label=="Grade 6"~"No High School Diploma",
    SCHL_label=="Grade 11"~"No High School Diploma",
    SCHL_label=="Professional degree beyond a bachelor's degree"~"Professional degree",
    SCHL_label=="Grade 9"~"No High School Diploma",
    SCHL_label=="No schooling completed"~"No High School Diploma",
    SCHL_label=="Grade 5"~"No High School Diploma",
    SCHL_label=="Grade 4"~"No High School Diploma",
    SCHL_label=="Nursery school, preschool"~"No High School Diploma",
    SCHL_label=="Grade 3"~"No High School Diploma",
    SCHL_label=="Grade 1"~"No High School Diploma",
    SCHL_label=="Kindergarten"~"No High School Diploma",
    SCHL_label=="Grade 2"~"No High School Diploma")) %>%
  mutate(income = case_when(
    PINCP < 25000 ~ "Low Income",
    (PINCP >= 25000) & (PINCP < 50000) ~ "1Low to Mid Income",
    (PINCP >= 50000) & (PINCP < 75000) ~ "Mid Income",
    (PINCP >= 75000) & (PINCP < 100000) ~ "Mid to High Income",
    PINCP > 100000 ~ "High Income")) %>%
  mutate(meansTW = case_when(
    JWTR_label=="Bicycle"~"Bicycle",
    JWTR_label=="Bus or trolley bus"~"Bus or\ntrolley bus",
    JWTR_label=="Car, truck, or van"~"1Car, truck,\nor van",
    JWTR_label=="Motorcycle"~"Motorcycle",
    JWTR_label=="Other method"~"Other",
    JWTR_label=="Railroad"~"Railroad",
    JWTR_label=="Streetcar or trolley car (carro publico in Puerto Rico)"~"Streetcar or\ntrolley car",
    JWTR_label=="Subway or elevated"~"Subway or\nelevated",
    JWTR_label=="Taxicab"~"Taxicab",
    JWTR_label=="Walked"~"Walk"))
```

```
## Parsed with column specification:
```

```
## cols(
##   SEX_label = col_character(),
##   JWTR_label = col_character(),
##   SCHL_label = col_character(),
##   vehicle = col_double(),
##   PINCP = col_double(),
##   JWMNP = col_double(),
##   GRNTP = col_double(),
##   AGEP = col_double()
## )
```

Model 2: A Linear Regression with mutated variables

I can run the linear regression with my newly mutated Education and Income variables.

```r
model3 <- lm(JWMNP ~ SEX_label + meansTW + edu_attainment + income + vehicle + GRNTP + AGEP,
           data = transpo_data2)
options(scipen = 999)
summary(model3)
```

```
##
## Call:
## lm(formula = JWMNP ~ SEX_label + meansTW + edu_attainment + income +
##     vehicle + GRNTP + AGEP, data = transpo_data2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.989 -12.264  -4.809   5.935 145.435
##
## Coefficients:
##                                       Estimate Std. Error t value
## (Intercept)                         18.9083034  1.3553199  13.951
## SEX_labelMale                        1.5806744  0.5313131   2.975
## meansTWBicycle                      -6.2920002  1.9232878  -3.271
## meansTWBus or\ntrolley bus          22.8378707  1.5014058  15.211
## meansTWMotorcycle                   -7.5640817  5.3189898  -1.422
## meansTWOther                        15.2411518  2.7346329   5.573
## meansTWRailroad                     14.8498681  3.7721258   3.937
## meansTWStreetcar or\ntrolley car    26.9445702  6.7233517   4.008
## meansTWSubway or\nelevated          19.1063292  3.6562654   5.226
## meansTWTaxicab                      -7.1028701  4.8881263  -1.453
## meansTWWalk                         -7.9338532  1.1999592  -6.612
## edu_attainmentAssociate's degree    -0.8109459  1.0332505  -0.785
## edu_attainmentBachelor's degree     -3.1656630  0.7780543  -4.069
## edu_attainmentDoctorate degree      -3.5821077  2.2631085  -1.583
## edu_attainmentMaster's degree       -2.8043728  1.0995001  -2.551
## edu_attainmentNo High School Diploma -0.3480509  1.0505946  -0.331
## edu_attainmentProfessional degree   -4.6035677  2.2077314  -2.085
## edu_attainmentSome College          -1.4613775  0.7649433  -1.910
## incomeHigh Income                   -0.6726311  1.3340220  -0.504
## incomeLow Income                    -1.9459021  0.6334875  -3.072
## incomeMid Income                     1.3511240  0.7382858   1.830
## incomeMid to High Income             1.9480822  1.1285427   1.726
```

```
## vehicle                                 0.6546097  0.2645167   2.475
## GRNTP                                   0.0020185  0.0004904   4.116
## AGEP                                    0.0618933  0.0210997   2.933
##                                                    Pr(>|t|)
## (Intercept)                           < 0.0000000000000002 ***
## SEX_labelMale                                       0.00294 **
## meansTWBicycle                                      0.00108 **
## meansTWBus or\ntrolley bus            < 0.0000000000000002 ***
## meansTWMotorcycle                                   0.15505
## meansTWOther                               0.0000000259376 ***
## meansTWRailroad                            0.0000834177027 ***
## meansTWStreetcar or\ntrolley car           0.0000619885659 ***
## meansTWSubway or\nelevated                 0.0000001786935 ***
## meansTWTaxicab                                      0.14625
## meansTWWalk                                0.0000000000408 ***
## edu_attainmentAssociate's degree                    0.43257
## edu_attainmentBachelor's degree            0.0000478104372 ***
## edu_attainmentDoctorate degree                      0.11351
## edu_attainmentMaster's degree                       0.01078 *
## edu_attainmentNo High School Diploma                0.74044
## edu_attainmentProfessional degree                   0.03709 *
## edu_attainmentSome College                          0.05612 .
## incomeHigh Income                                   0.61413
## incomeLow Income                                    0.00214 **
## incomeMid Income                                    0.06728 .
## incomeMid to High Income                            0.08436 .
## vehicle                                             0.01336 *
## GRNTP                                      0.0000390659465 ***
## AGEP                                                0.00336 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.22 on 6830 degrees of freedom
##   (59 observations deleted due to missingness)
## Multiple R-squared:  0.06674,    Adjusted R-squared:  0.06346
## F-statistic: 20.35 on 24 and 6830 DF,  p-value: < 0.00000000000000022
```

This model has an r-squared value of 0.06346, which means that this model predicts about 6% of the variation in Travel Time to Work in this dataset. This r-squared value is barely higher than the previous model, with ungrouped educational outcomes and a continuous income set. 0.06346 compared to 0.06101.

Interesting to note is that in the education variable, the only value to gain statistical significance was the coefficient for having earned a master's degree, suggesting that those with a masters degree will travel 2.8 minutes less to work than those with just a high school diploma. I thought that aggregating all of the K-12 grades might have brought about a significant value because it captures more people, but it didn't.

SEX_labelMale became more statistically significant with this model, which makes me think that there's an intereaction between sex and education or income.

Grouping income into intervals did increase the statistical significance from when it was continuous. The only statistical significant vale was the Low Income interval, suggesting that those who make less than $25,000 will actually travel about 1.9 minutes less to work than those who make between $25,000 and $50,000 per year. In the state of Colorado, jobs are fairly spread out throughout cities and not necessarily concentrated in hubs. Jobs that tend to be concentrated in hubs like downtown Denver or Boulder tend to be higher income jobs, so perhaps it's not too difficult for low-income earners to find housing closer to their jobs.

Speaking about housing, Gross Rent is statistically significant, but the coefficient is not very telling, suggesting that there's no real difference in travel time to work for every additional dollar paid for rent.

Model 3: Interaction Between Income and Sex, and Between Education and Sex

Seeing that statistical significance increased for Sex in my previous model, I wonder if that was due to the change in income from continuous to categorical data. I think that the effect of sex on income can tell us more about this model.

```
model4 <- lm(JWMNP ~ SEX_label + meansTW + edu_attainment + income + vehicle + GRNTP + AGEP + income:SEX
            data = transpo_data2)
options(scipen = 999)
summary(model4)
```

```
##
## Call:
## lm(formula = JWMNP ~ SEX_label + meansTW + edu_attainment + income +
##     vehicle + GRNTP + AGEP + income:SEX_label + edu_attainment:SEX_label,
##     data = transpo_data2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.767 -12.246  -4.708   6.111 145.194
##
## Coefficients:
##                                                      Estimate Std. Error
## (Intercept)                                        17.5586255  1.5350380
## SEX_labelMale                                       3.3061258  1.2928974
## meansTWBicycle                                     -6.0858897  1.9254434
## meansTWBus or\ntrolley bus                         22.8825220  1.5015675
## meansTWMotorcycle                                  -7.4577637  5.3168045
## meansTWOther                                       15.4539260  2.7363212
## meansTWRailroad                                    15.3298053  3.7743455
## meansTWStreetcar or\ntrolley car                   26.8878400  6.7210029
## meansTWSubway or\nelevated                         18.9807947  3.6560631
## meansTWTaxicab                                     -7.3645277  4.8860736
## meansTWWalk                                        -7.8613246  1.2001047
## edu_attainmentAssociate's degree                   -0.3122799  1.5016609
## edu_attainmentBachelor's degree                    -0.9983539  1.1652308
## edu_attainmentDoctorate degree                      0.2256841  3.4709294
## edu_attainmentMaster's degree                       0.2683471  1.5928253
## edu_attainmentNo High School Diploma               -2.6364238  1.7151183
## edu_attainmentProfessional degree                  -4.2047347  3.4430923
## edu_attainmentSome College                         -0.4470599  1.1634211
## incomeHigh Income                                   1.4543482  2.4511385
## incomeLow Income                                   -1.6013566  0.8868667
## incomeMid Income                                    0.4741122  1.1649160
## incomeMid to High Income                            1.8888916  1.8877411
## vehicle                                             0.6453199  0.2645896
## GRNTP                                               0.0020914  0.0004909
## AGEP                                                0.0646033  0.0211116
## SEX_labelMale:incomeHigh Income                    -2.3571458  2.8855071
## SEX_labelMale:incomeLow Income                     -0.2177640  1.2524374
```

```
## SEX_labelMale:incomeMid Income                       1.3239531  1.5008065
## SEX_labelMale:incomeMid to High Income               0.1579242  2.3423614
## SEX_labelMale:edu_attainmentAssociate's degree      -0.5737742  2.0744692
## SEX_labelMale:edu_attainmentBachelor's degree       -3.8531214  1.5489561
## SEX_labelMale:edu_attainmentDoctorate degree        -6.4442337  4.5640119
## SEX_labelMale:edu_attainmentMaster's degree         -5.7393018  2.1921818
## SEX_labelMale:edu_attainmentNo High School Diploma  3.5232929  2.1712811
## SEX_labelMale:edu_attainmentProfessional degree    -0.6711702  4.4740206
## SEX_labelMale:edu_attainmentSome College           -1.6900155  1.5446683
##                                                    t value             Pr(>|t|)
## (Intercept)                                         11.439 < 0.0000000000000002
## SEX_labelMale                                        2.557              0.01057
## meansTWBicycle                                      -3.161              0.00158
## meansTWBus or\ntrolley bus                          15.239 < 0.0000000000000002
## meansTWMotorcycle                                   -1.403              0.16076
## meansTWOther                                         5.648      0.0000000169157
## meansTWRailroad                                      4.062      0.0000492879982
## meansTWStreetcar or\ntrolley car                     4.001      0.0000638587278
## meansTWSubway or\nelevated                           5.192      0.0000002145308
## meansTWTaxicab                                      -1.507              0.13179
## meansTWWalk                                         -6.551      0.0000000000615
## edu_attainmentAssociate's degree                    -0.208              0.83527
## edu_attainmentBachelor's degree                     -0.857              0.39159
## edu_attainmentDoctorate degree                       0.065              0.94816
## edu_attainmentMaster's degree                        0.168              0.86622
## edu_attainmentNo High School Diploma                -1.537              0.12430
## edu_attainmentProfessional degree                   -1.221              0.22205
## edu_attainmentSome College                          -0.384              0.70080
## incomeHigh Income                                    0.593              0.55298
## incomeLow Income                                    -1.806              0.07102
## incomeMid Income                                     0.407              0.68403
## incomeMid to High Income                             1.001              0.31705
## vehicle                                              2.439              0.01476
## GRNTP                                                4.260      0.0000207047889
## AGEP                                                 3.060              0.00222
## SEX_labelMale:incomeHigh Income                     -0.817              0.41402
## SEX_labelMale:incomeLow Income                      -0.174              0.86197
## SEX_labelMale:incomeMid Income                       0.882              0.37772
## SEX_labelMale:incomeMid to High Income               0.067              0.94625
## SEX_labelMale:edu_attainmentAssociate's degree      -0.277              0.78210
## SEX_labelMale:edu_attainmentBachelor's degree       -2.488              0.01289
## SEX_labelMale:edu_attainmentDoctorate degree        -1.412              0.15801
## SEX_labelMale:edu_attainmentMaster's degree         -2.618              0.00886
## SEX_labelMale:edu_attainmentNo High School Diploma   1.623              0.10470
## SEX_labelMale:edu_attainmentProfessional degree     -0.150              0.88076
## SEX_labelMale:edu_attainmentSome College            -1.094              0.27395
##
## (Intercept)                                        ***
## SEX_labelMale                                      *
## meansTWBicycle                                     **
## meansTWBus or\ntrolley bus                         ***
## meansTWMotorcycle
## meansTWOther                                       ***
## meansTWRailroad                                    ***
```

```
## meansTWStreetcar or\ntrolley car              ***
## meansTWSubway or\nelevated                     ***
## meansTWTaxicab
## meansTWWalk                                    ***
## edu_attainmentAssociate’s degree
## edu_attainmentBachelor’s degree
## edu_attainmentDoctorate degree
## edu_attainmentMaster’s degree
## edu_attainmentNo High School Diploma
## edu_attainmentProfessional degree
## edu_attainmentSome College
## incomeHigh Income
## incomeLow Income                               .
## incomeMid Income
## incomeMid to High Income
## vehicle                                        *
## GRNTP                                          ***
## AGEP                                           **
## SEX_labelMale:incomeHigh Income
## SEX_labelMale:incomeLow Income
## SEX_labelMale:incomeMid Income
## SEX_labelMale:incomeMid to High Income
## SEX_labelMale:edu_attainmentAssociate’s degree
## SEX_labelMale:edu_attainmentBachelor’s degree  *
## SEX_labelMale:edu_attainmentDoctorate degree
## SEX_labelMale:edu_attainmentMaster’s degree    **
## SEX_labelMale:edu_attainmentNo High School Diploma
## SEX_labelMale:edu_attainmentProfessional degree
## SEX_labelMale:edu_attainmentSome College
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 21.2 on 6819 degrees of freedom
##   (59 observations deleted due to missingness)
## Multiple R-squared:  0.06979,    Adjusted R-squared:  0.06502
## F-statistic: 14.62 on 35 and 6819 DF,  p-value: < 0.00000000000000022
```

This model has an r-squared value of 0.06502, which means that this model predicts about 6% of the variation in Travel Time to Work in this dataset. This r-squared value is barely higher than the previous model without interaction terms. This is my best model yet. 0.06502 compared to 0.06346.

Interestingly, neither of the interactions between Sex and Income were statistically significant. However, interactions between Sex and Bachelor’s Degree and Sex and Master’s Degree were statistically significant.

The first interaction suggests that men with a bachelor’s degree will, on average and controlling for all other variables, travel for 3.8 minutes less to work than women with a bachelor’s degree.

The second interaction suggests that men with a master’s degree will, on average and controlling for all other variables, travel for 5.7 minutes less to work than women with a master’s degree.

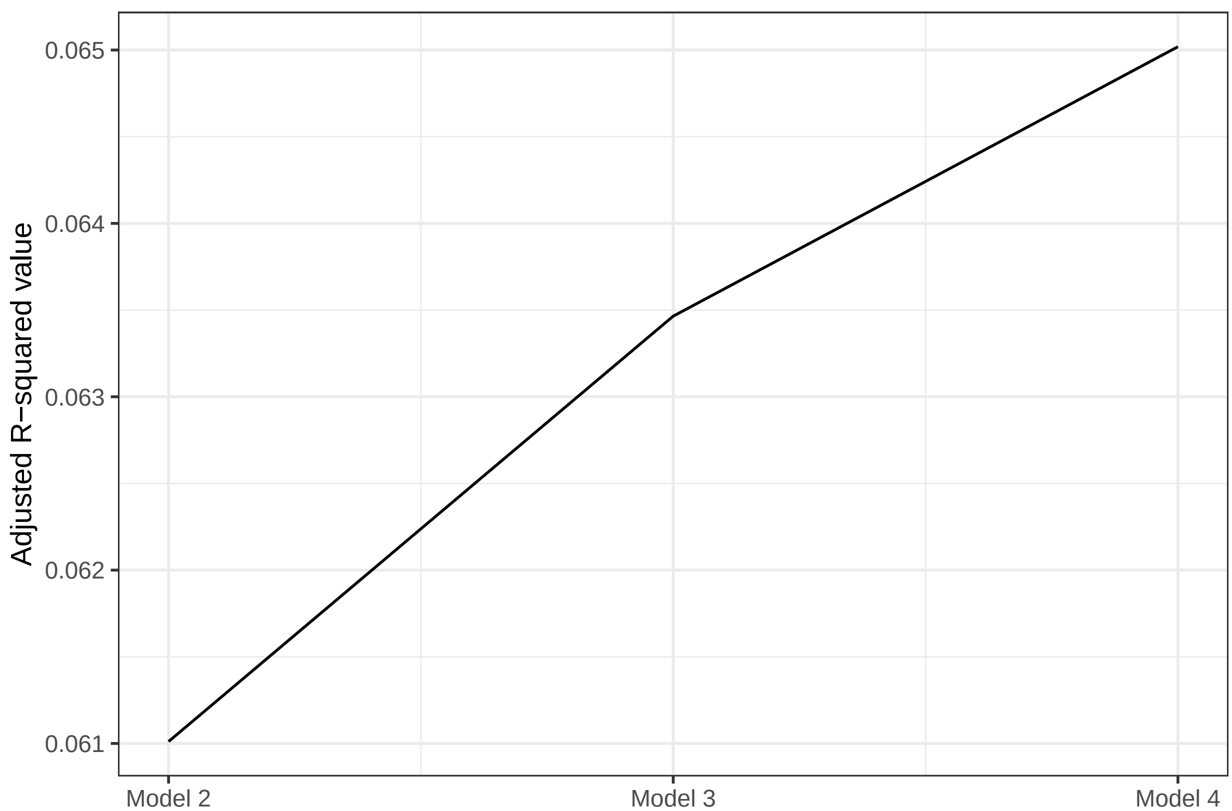
In the previous two models before examining interactions between Sex and Income/Education, men corresponded with greater travel time, so this model tells a more nuanced story of how Sex and Education interact with each other in correlating with travel time.

Visualizing Model Fits

The graph below compares the r-squared values for my four models.

```
ModelFit <- tibble(model = c(2, 3, 4),
                   R_square = c(summary(model2)$adj.r.squared,
                                summary(model3)$adj.r.squared,
                                summary(model4)$adj.r.squared))

ggplot(ModelFit, aes(x = model, y = R_square)) +
  geom_line() +
  scale_x_continuous(name = "",
                   breaks = breaks <- seq(1, 4, by = 1),
                   labels = paste("Model", breaks)) +
  scale_y_continuous(name = "Adjusted R-squared value") +
  theme_bw()
```



We can see from this that Model 4, with the interaction terms, had the highest r-squared value, but the difference across the three is quite small.

Visualzing the Education and Sex Interaction on Commute Time

Lastly, to visualze the effect of the interaction that I found in model four, I can make this graph. It shows the relationship between travel time to work and sex, by educational attainment, and holding all other variables constant.

```r
cat_plot(model4, pred = edu_attainment, modx = SEX_label, interval = TRUE) +
 scale_x_discrete(name = "Educational Attainment ",
                   labels = c("High\nSchool\nDiploma", "Associate's\nDegree",
                              "Bechelor's\nDegree", "Doctorate\nDegree","Master's\nDegree", "No High\ns
  scale_y_continuous(name = "Travel time to work (minutes)",
                    breaks = seq(0, 45, by = 2)) +
  scale_color_discrete(name = "") +
  scale_linetype_discrete(name = "") +
  scale_fill_discrete(name = "")
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.

## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.

## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```