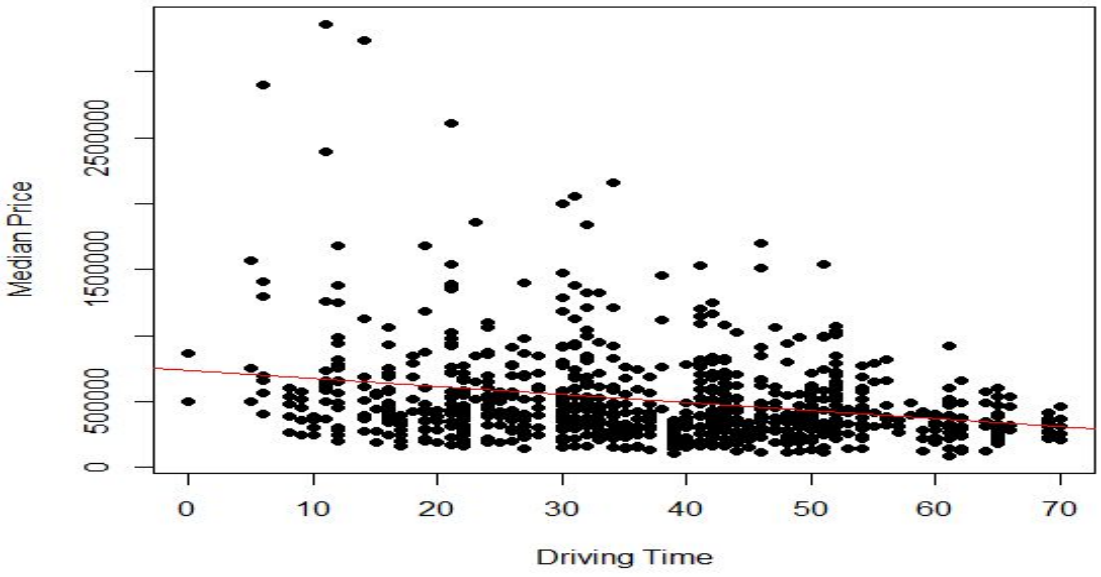
1) setwd("c:/Users/btdan/Desktop/ECN 145/hw")

```
> #2
> str(ps1)
'data.frame':   1140 obs. of  10 variables:
 $ Zcta5          : int  1431 1431 1431 1431 1431 1432 1432 1432 1432 1432 ...
 $ City           : Factor w/ 150 levels " Amesbury, MA",..: 33 33 33 33 33 3 3 3 3 3 ...
 $ Bedroom        : int  1 2 3 4 5 1 2 3 4 5 ...
 $ MedianPrice    : int  NA 195600 243100 278600 NA NA 193200 287500 336800 NA ...
 $ DrivingDistance: num  54.8 54.8 54.8 54.8 54.8 ...
 $ DrivingTime    : int  99 99 99 99 99 72 72 72 72 72 ...
 $ TransitTime    : int  150 150 150 150 150 86 86 86 86 86 ...
 $ EnglishScore   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ MathScore      : int  NA NA NA NA NA NA NA NA NA NA ...
 $ ScienceScore   : int  NA NA NA NA NA NA NA NA NA NA ...
> names(ps1)
 [1] "Zcta5"           "City"            "Bedroom"         "MedianPrice"     "DrivingDistance" "DrivingTime"     "TransitTime"     "EnglishScore"
 [9] "MathScore"       "ScienceScore"
> summary(ps1)
     Zcta5                  City        Bedroom       MedianPrice      DrivingDistance  DrivingTime     TransitTime      EnglishScore     MathScore
 Min.   :1431   Boston, MA     :140   Min.   :1   Min.   :  88700   Min.   : 0.00   Min.   : 0.00   Min.   :  1.00   Min.   : 8.0   Min.   :18.00
 1st Qu.:1884   Newton, MA     : 50   1st Qu.:2   1st Qu.: 290700   1st Qu.:10.96   1st Qu.:27.00   1st Qu.: 56.25   1st Qu.:29.0   1st Qu.:40.00
 Median :2076   Cambridge, MA  : 25   Median :3   Median : 404200   Median :22.20   Median :41.00   Median : 83.00   Median :44.0   Median :55.00
 Mean   :2080    Marshfield, MA: 20   Mean   :3   Mean   : 495307   Mean   :22.59   Mean   :39.99   Mean   : 92.42   Mean   :45.4   Mean   :57.22
 3rd Qu.:2301   Lowell, MA     : 20   3rd Qu.:4   3rd Qu.: 587100   3rd Qu.:31.95   3rd Qu.:52.00   3rd Qu.:119.50   3rd Qu.:64.0   3rd Qu.:74.00
 Max.   :2770   Lynn, MA       : 20   Max.   :5   Max.   :3358500   Max.   :62.44   Max.   :99.00   Max.   :221.00   Max.   :81.0   Max.   :90.00
                (Other)        :865               NA's   :259                                                        NA's   :140    NA's   :140
  ScienceScore
 Min.   : 2.00
 1st Qu.:17.00
 Median :32.00
 Mean   :33.67
 3rd Qu.:48.00
 Max.   :76.00
 NA's   :140
```

2) Observations: 1140 of 10 variables
Average driving time: 39.99 minutes
Average transit time: 92.42 minutes
Longest commute to Downtown Boston: 62.44

3)



From the regression line, we can see a negative correlation between median prices and driving times. Home prices will decline as commute time increases, and vice versa.

4)

```
> #4
> fit <- lm(MedianPrice ~ DrivingTime, data = ps1)
> summary(fit)

Call:
lm(formula = MedianPrice ~ DrivingTime, data = ps1)

Residuals:
    Min      1Q  Median      3Q     Max
-465000 -203777  -80290  108402 2693362

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 731941.0    32078.4  22.817  < 2e-16 ***
DrivingTime  -6073.0      799.8  -7.593 8.84e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 347700 on 787 degrees of freedom
Multiple R-squared:  0.06826,   Adjusted R-squared:  0.06708
F-statistic: 57.66 on 1 and 787 DF,  p-value: 8.843e-14
```

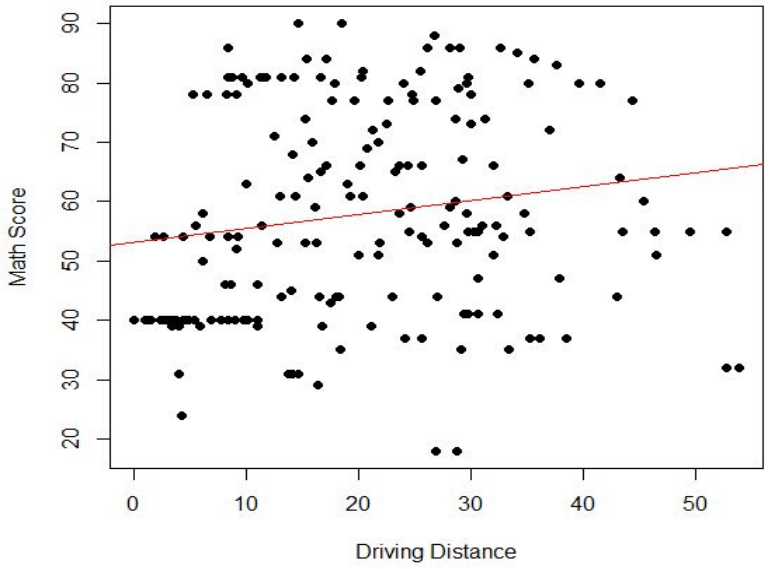
The regression relates to the scatter plot by finding lines that fit best into the data. Our coefficients then give us the slope and intercept of that line.
For every one minute increase in Driving Time (slope), there will be a -6,073 reduction in housing prices.
Intercept = 731,941

5)

```
> #5
> fitbed <- lm(MedianPrice ~ DrivingTime + factor(Bedroom), data=ps1)
> summary(fitbed)

Call:
lm(formula = MedianPrice ~ DrivingTime + factor(Bedroom), data = ps1)

Residuals:
    Min      1Q  Median      3Q     Max
-584405 -163258  -37330   91742 2570711

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      493073.5    36077.2  13.667  < 2e-16 ***
DrivingTime       -7236.9      697.4 -10.378  < 2e-16 ***
factor(Bedroom)2 158185.5    35833.2   4.414 1.15e-05 ***
factor(Bedroom)3 284683.4    35949.4   7.919 8.20e-15 ***
factor(Bedroom)4 374321.1    36400.8  10.283  < 2e-16 ***
factor(Bedroom)5 572258.4    38447.8  14.884  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 300600 on 783 degrees of freedom
Multiple R-squared:  0.3072,    Adjusted R-squared:  0.3028
F-statistic: 69.45 on 5 and 783 DF,  p-value: < 2.2e-16
```

By comparing the cost of adding a bedroom relative to the cost of a single bedroom, we see a positive trend. For example, this would mean that it would cost $158,185 more to own a two-bedroom house than a single one. Its R-squared is higher than before, from .07 to .31, which means that more of the data fits into the regression model. Our predictions become more accurate with the addition of bedroom controls.

6)



The scatter plots show a positive correlation between math scores and commute distances. For every increase in math scores, they will have to commute an extra mile.

7)

```
> #7
> fitfull <- lm(MedianPrice ~ DrivingTime + factor(Bedroom) + MathScore, data = ps1)
> summary(fitfull)

Call:
lm(formula = MedianPrice ~ DrivingTime + factor(Bedroom) + MathScore,
    data = ps1)

Residuals:
    Min      1Q  Median      3Q     Max
-503594 -133684  -37220   70837 2653533

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       173745.3    42023.4   4.134 3.94e-05 ***
DrivingTime        -9173.2      657.9 -13.944  < 2e-16 ***
factor(Bedroom)2 134228.6    32879.2   4.082 4.91e-05 ***
factor(Bedroom)3 254149.7    33021.3   7.697 4.22e-14 ***
factor(Bedroom)4 341663.8    33446.9  10.215  < 2e-16 ***
factor(Bedroom)5 515213.8    35520.3  14.505  < 2e-16 ***
MathScore          7261.6      590.4  12.300  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 275300 on 782 degrees of freedom
Multiple R-squared:  0.4195,    Adjusted R-squared:  0.4151
F-statistic:  94.2 on 6 and 782 DF,  p-value: < 2.2e-16
```

The coefficient on math score is the correct sign and is positive. In other words, a one-unit change in math scores will raise home prices by 7261.6.

8) The positive relationship between math scores and housing prices from question 6 indicates that there be another variable that further reduces home prices for every one unit added in commute time.

9) Driving time coefficient = -9173.2, so home prices are reduced by $9,173.2 for every minute increase in commute time. The magnitude of the coefficient seems plausible when calculating a wider time horizon and considering if benefits exceed additional costs to travel farther (time).

10) An explanation for this case might be that students are responsible for the positive relationship between math scores and driving distance. They are expected to test well and have cheaper preferences.

You may regress math scores to the variables driving time and bedroom to test their relationship with large and small homes.

Code

```
setwd("c:/Users/btdan/Desktop/ECN 145/hw")

#1
library(readr)
ps1 <- read.csv("ps1_data.csv")
View(ps1)

Zcta5 <- c(ps1$V1)
City <- c(ps1$V2)
Bedroom <- c(ps1$V3)
MedianPrice <- c(ps1$V4)
DrivingDistance <- c(ps1$V5)
DrivingTime <- c(ps1$V6)
MathScore <- c(ps1$V7)

#2
dim(ps1)
nrow(ps1)

names(ps1)
summary(ps1)

str(ps1)
is.na(ps1)

ps1[!complete.cases(ps1),]
getOption("max.print")

ps1 <- na.omit(ps1)
summary(ps1)


#3
plot(x = ps1$DrivingTime, y = ps1$MedianPrice, xlab = "Driving Time", ylab = "Median Price",
pch = 19)
abline(lm(ps1$MedianPrice ~ ps1$DrivingTime), col = "red")

#4
```

```
fit <- lm(MedianPrice ~ DrivingTime, data = ps1)
summary(fit)

#5
fitbed <- lm(MedianPrice ~ DrivingTime + factor(Bedroom), data=ps1)
summary(fitbed)


#6
plot(x = ps1$DrivingDistance, y = ps1$MathScore, xlab = "Driving Distance", ylab = "Math
Score", pch = 19)
abline(lm(ps1$MathScore ~ ps1$DrivingDistance), col = "red")
fitmath <- lm(MathScore ~ DrivingDistance, data = ps1)
summary(ps1)

#7
fitfull <- lm(MedianPrice ~ DrivingTime + factor(Bedroom) + MathScore, data = ps1)
summary(fitfull)

fitsize <- lm(MathScore ~ DrivingTime + factor(Bedroom), data = ps1)
summary(fitsize)
```