ECON 573 Problem Set 2

Part 1

Ex 1, 3, 4, from Chapter 3 of ISL.

1) Describe the null hypotheses to which the p-values given in Table 3.4 correspond. Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of sales, TV, radio, and newspaper, rather than in terms of the coefficients of the linear model.

The null hypothesis is that TV, radio, and newspapper advertising have no effect on sales. After conduting the test, we reject the null hypothesis that sales TV and radio have no effect on sales. However, we fail to reject the null that newspaper advertising has no effect on sales

3) Suppose we have a data set with five predictors, X1 =GPA, X2 = IQ, X3 = Gender (1 for Female and 0 forMale), X4 = Interaction between GPA and IQ, and X5 = Interaction between GPA and Gender. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and get ˆB0 = 50, ˆB1 = 20, ˆB2 = 0.07, ˆB3 = 35, ˆB4 = 0.01, ˆB5 = -10.

3a) Which answer is correct, and why? Male = 50 + 20(gpa) + 0.07(iq) + 0.01(iqandgpa) Female = 85 + 10(gpa) + 0.07(iq) + 0.01(iqandgpa) Point iii is the most valid one

3b) Predict the salary of a female with IQ of 110 and a GPA of 4.0. Female = 85 + 10(4) + 0.07(110) + 0.01(110*4) = 137.1

3c) True or false: Since the coefficient for the GPA/IQ interaction term is very small, there is very little evidence of an interaction effect. Justify your answer. Flase, just because the coefficient is realtively small doesnt mean that there is littler interaction effect. A true test would be to test if the interaction term is 0

4) I collect a set of data (n = 100 observations) containing a single predictor and a quantitative response. I then fit a linear regression model to the data, as well as a separate cubic regression, i.e. Y = B0 + B1X + B2X2 + B3X3 + .

4a) Suppose that the true relationship between X and Y is linear, i.e. Y = B0 + B1X + . Consider the training residual sum of squares (RSS) for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer. It should be expected that the cubic regression RSS is less than that of the linear regression due to more flexibility with the cubic

4b) Answer (a) using test rather than training RSS.

4c) Suppose that the true relationship between X and Y is not linear, but we don't know how far it is from linear. Consider the training RSS for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.

4d) Answer (c) using test rather than training RSS.

Part 2

8) This question involves the use of simple linear regression on the Auto data set. 8a) Use the lm() function to perform a simple linear regression with mpg as the response and horsepower as the predictor. Use the summary() function to print the results. Comment on the output.

```
library(ISLR)
data("Auto")
names(Auto)
```

```
## [1] "mpg"          "cylinders"    "displacement" "horsepower"   "weight"
## [6] "acceleration" "year"         "origin"       "name"
```

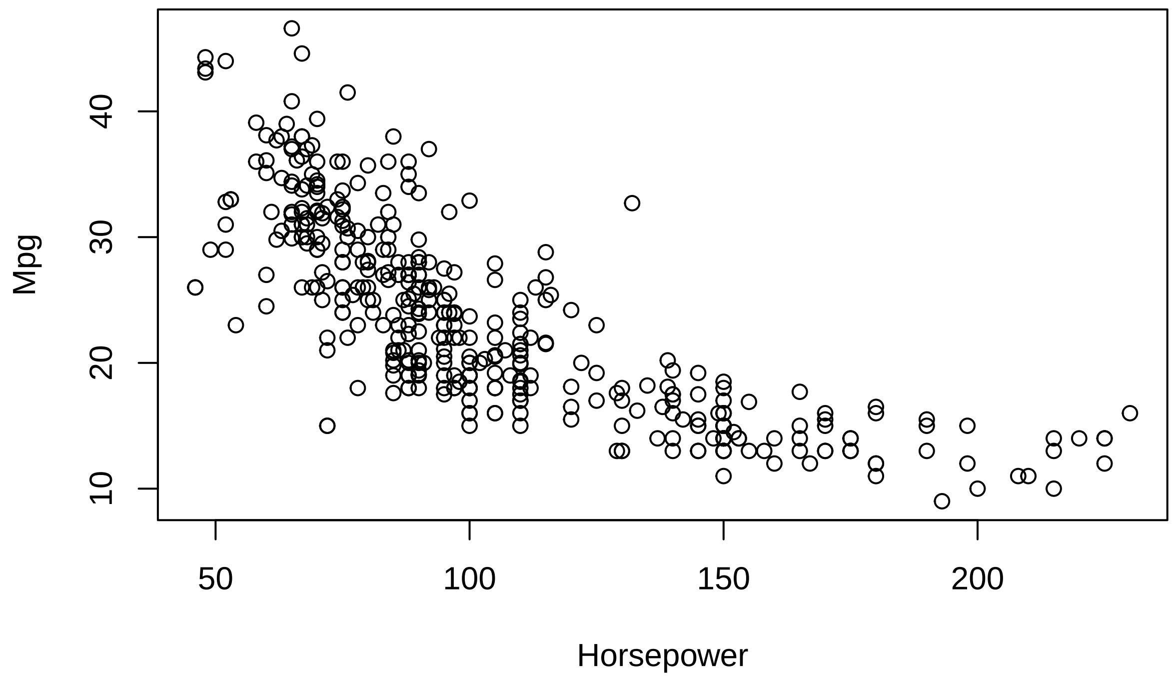
```
linreg = lm(mpg ~ horsepower, data = Auto)
summary(linreg)
```

```
##
## Call:
## lm(formula = mpg ~ horsepower, data = Auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.5710  -3.2592  -0.3435   2.7630  16.9240
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 39.935861   0.717499   55.66   <2e-16 ***
## horsepower  -0.157845   0.006446  -24.49   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.906 on 390 degrees of freedom
## Multiple R-squared:  0.6059, Adjusted R-squared:  0.6049
## F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

At all standard significance levels the p values are significant, indicating association between mpg and horsepower. The R^2 values suggests that 60.5% of the variation in mpg is due to horsepower. There seems to be a moderately strong negative association between mpg and horsepower.

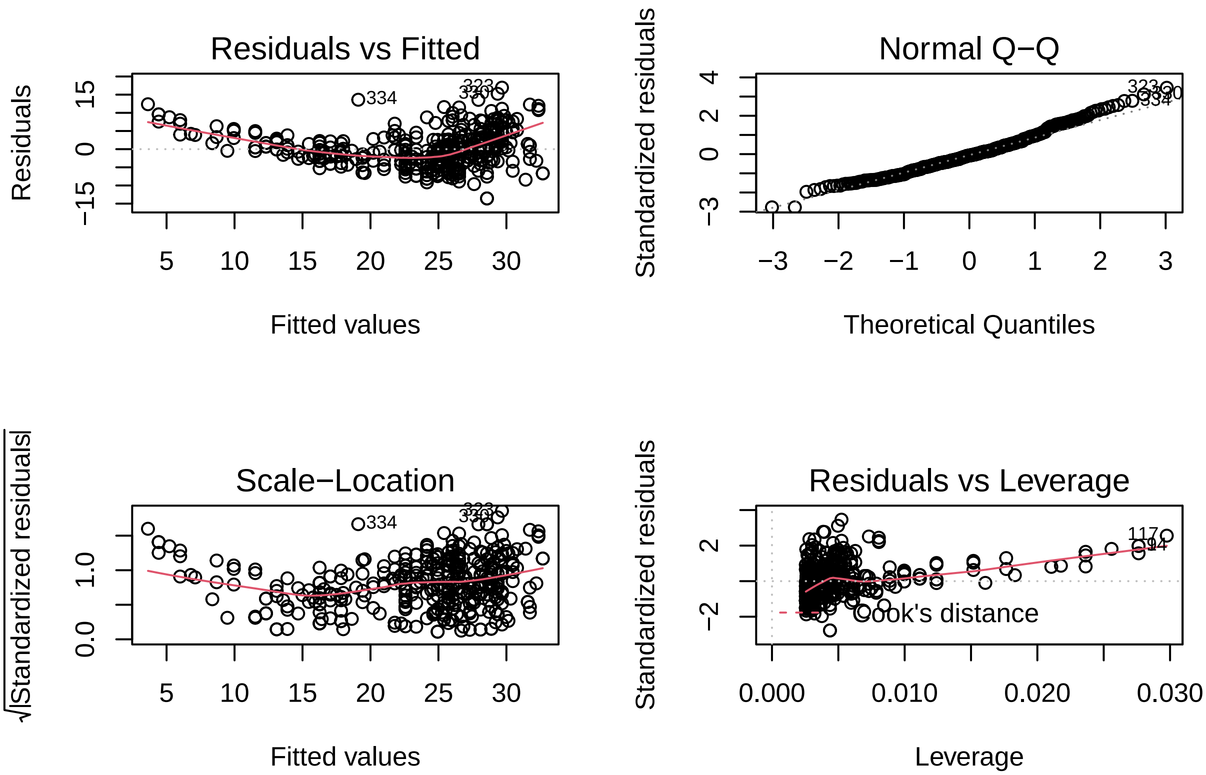
8b) Plot the response and the predictor. Use the abline() function to display the least squares regression line.

```
plot(Auto$mpg ~ Auto$horsepower, xlab="Horsepower", ylab="Mpg")
```



8c) Use the plot() function to produce diagnostic plots of the least squares regression fit. Comment on any problems you see with the fit.

```
par(mfrow=c(2,2))
plot(linreg)
```



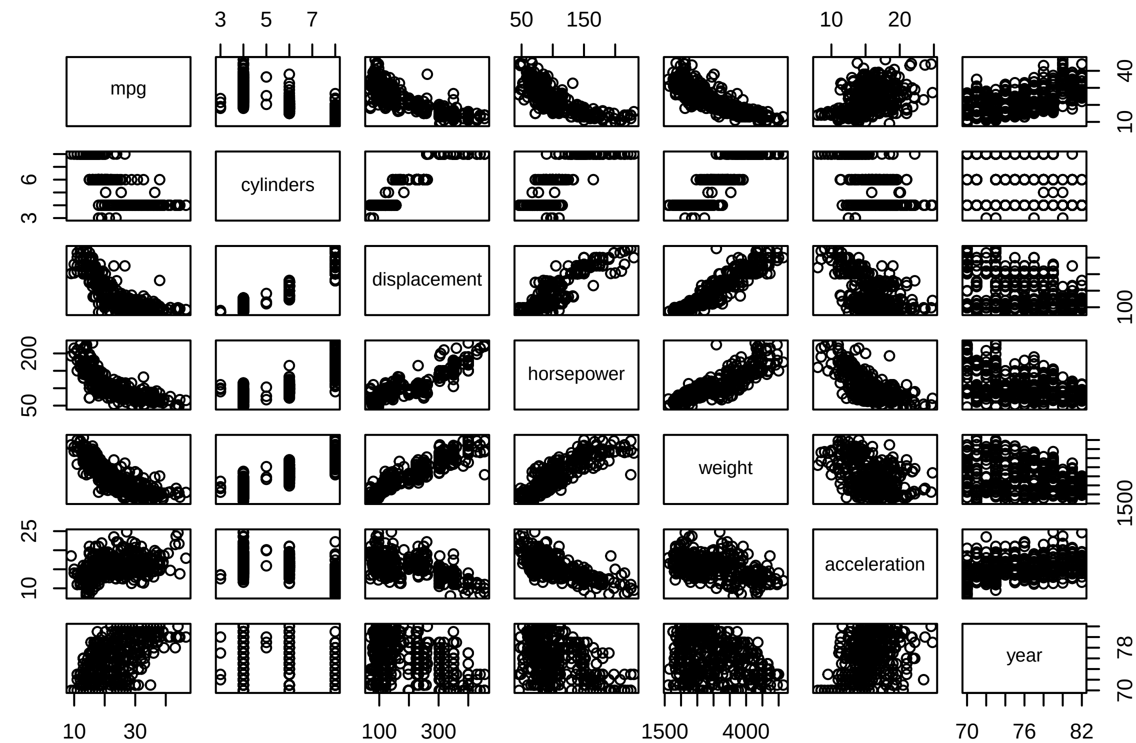
From the plots it can be seen that the relationship between is not linear, normally distributed, constant variance, an no major leverage points

9) This question involves the use of multiple linear regression on the Auto data set. 9a) Produce a scatterplot matrix which includes all of the variables in the data set.

```
Auto$horsepower <- as.numeric(type.convert(Auto$horsepower))
```

```
## Warning in type.convert.default(Auto$horsepower): 'as.is' should be specified by
## the caller; using TRUE
```

```
pairs(Auto[,1:7])
```



9b) Compute the matrix of correlations between the variables using the function cor(). You will need to exclude the name variable, cor() which is qualitative.

```
cor(Auto[,1:7])
```

```
##                     mpg  cylinders displacement horsepower     weight
## mpg           1.0000000 -0.7776175   -0.8051269 -0.7784268 -0.8322442
## cylinders    -0.7776175  1.0000000    0.9508233  0.8429834  0.8975273
## displacement -0.8051269  0.9508233    1.0000000  0.8972570  0.9329944
## horsepower   -0.7784268  0.8429834    0.8972570  1.0000000  0.8645377
## weight       -0.8322442  0.8975273    0.9329944  0.8645377  1.0000000
## acceleration  0.4233285 -0.5046834   -0.5438005 -0.6891955 -0.4168392
## year          0.5805410 -0.3456474   -0.3698552 -0.4163615 -0.3091199
##              acceleration       year
## mpg             0.4233285  0.5805410
## cylinders      -0.5046834 -0.3456474
## displacement   -0.5438005 -0.3698552
## horsepower     -0.6891955 -0.4163615
## weight         -0.4168392 -0.3091199
## acceleration    1.0000000  0.2903161
## year            0.2903161  1.0000000
```

9c) Use the lm() function to perform a multiple linear regression with mpg as the response and all other variables except name as the predictors. Use the summary() function to print the results. Comment on the output.

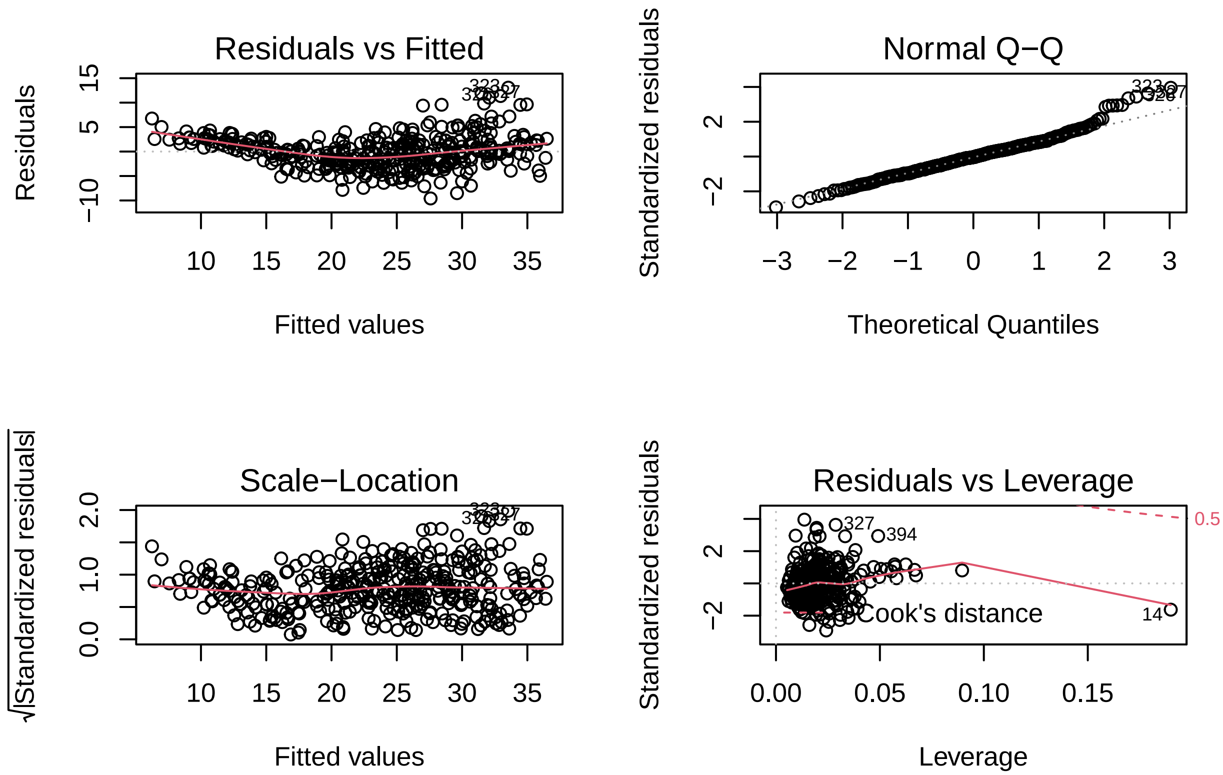
```r
attach(Auto)
multiple = lm(mpg ~. -name, data = Auto)
summary(multiple)
```

```
##
## Call:
## lm(formula = mpg ~ . - name, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5903 -2.1565 -0.1169  1.8690 13.0604
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -17.218435   4.644294  -3.707  0.00024 ***
## cylinders     -0.493376   0.323282  -1.526  0.12780
## displacement   0.019896   0.007515   2.647  0.00844 **
## horsepower    -0.016951   0.013787  -1.230  0.21963
## weight        -0.006474   0.000652  -9.929  < 2e-16 ***
## acceleration   0.080576   0.098845   0.815  0.41548
## year           0.750773   0.050973  14.729  < 2e-16 ***
## origin         1.426141   0.278136   5.127 4.67e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.328 on 384 degrees of freedom
## Multiple R-squared:  0.8215, Adjusted R-squared:  0.8182
## F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

All but cylinders, acceleration, and horsepower are statistically significant. The R^2 value of 82.15% implies that 82% of the variation in mpg is explained by the said reggressors.

9d) Use the plot() function to produce diagnostic plots of the linear regression fit. Comment on any problems you see with the fit. Do the residual plots suggest any unusually large outlines? Does the leverage plot identify any observations with unusually high leverage?

```r
par(mfrow=c(2,2))
plot(multiple)
```



There seems to be a non-linear relationship between the aggressors, while the residuals are normally distributed there does seem to be some that are skewed to the right. There also seem to be one outlier leberage poin in the fourth graph

9e) Use the * and : symbols to fit linear regression models with interaction effects. Do any interactions appear to be statistically significant?

```
multiple = lm(mpg ~. -name + horsepower*weight + year*acceleration + cylinders*weight + cylinders*displa
summary(multiple)
```

```
##
## Call:
## lm(formula = mpg ~ . - name + horsepower * weight + year * acceleration +
##     cylinders * weight + cylinders * displacement, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.7506 -1.3918 -0.0726  1.2024 11.3285
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.231e+02  2.027e+01   6.070 3.10e-09 ***
## cylinders              -1.369e+00  1.070e+00  -1.279  0.20156
## displacement            9.235e-03  2.276e-02   0.406  0.68519
## horsepower             -1.918e-01  3.756e-02  -5.106 5.22e-07 ***
## weight                 -1.303e-02  2.238e-03  -5.822 1.24e-08 ***
## acceleration           -7.449e+00  1.209e+00  -6.164 1.81e-09 ***
```

```
## year                   -7.516e-01  2.523e-01  -2.980  0.00307 **
## origin                  6.578e-01  2.482e-01   2.651  0.00836 **
## horsepower:weight       4.203e-05  9.555e-06   4.399 1.41e-05 ***
## acceleration:year       9.648e-02  1.580e-02   6.106 2.53e-09 ***
## cylinders:weight        5.983e-04  4.318e-04   1.386  0.16671
## cylinders:displacement -1.813e-03  3.185e-03  -0.569  0.56947
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.795 on 380 degrees of freedom
## Multiple R-squared:  0.8754, Adjusted R-squared:  0.8718
## F-statistic: 242.7 on 11 and 380 DF,  p-value: < 2.2e-16
```

From this regressions outcomes that are statistically significant include the interaction term between acceleration and year as well as horsepower and weight

9f) Try a few different transformations of the variables, such as log(X), sqrt(X), X2. Comment on your findings.

```r
multiple = lm(mpg ~ . - name + log(horsepower) + sqrt(weight) + I(displacement^2),data=Auto)
summary(multiple)
```

```
##
## Call:
## lm(formula = mpg ~ . - name + log(horsepower) + sqrt(weight) +
##     I(displacement^2), data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.2591 -1.5052 -0.1238  1.4466 11.9749
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        7.444e+01  1.574e+01   4.730 3.17e-06 ***
## cylinders          1.449e-01  3.341e-01   0.434   0.6647
## displacement      -3.472e-02  2.003e-02  -1.734   0.0837 .
## horsepower         7.827e-02  3.284e-02   2.383   0.0177 *
## weight             6.377e-03  4.242e-03   1.503   0.1335
## acceleration      -1.955e-01  9.992e-02  -1.957   0.0511 .
## year               7.670e-01  4.504e-02  17.031  < 2e-16 ***
## origin             5.479e-01  2.658e-01   2.061   0.0400 *
## log(horsepower)   -1.531e+01  3.762e+00  -4.069 5.75e-05 ***
## sqrt(weight)      -1.133e+00  4.940e-01  -2.293   0.0224 *
## I(displacement^2)  6.416e-05  3.446e-05   1.862   0.0634 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.899 on 381 degrees of freedom
## Multiple R-squared:  0.8656, Adjusted R-squared:  0.8621
## F-statistic: 245.4 on 10 and 381 DF,  p-value: < 2.2e-16
```

In this regression only the log of horsepower proved to be significant

13) In this exercise you will create some simulated data and will fit simple linear regression models to it. Make sure to use set.seed(1) prior to starting part (a) to ensure consistent results. 13a) Using the rnorm() function, create a vector, x, containing 100 observations drawn from a N(0, 1) distribution. This represents a feature, X.

```
X = rnorm(100, mean = 0, sd = 1)
```

13b) Using the rnorm() function, create a vector, eps, containing 100 observations drawn from a N(0, 0.25) distribution i.e. a normal distribution with mean zero and variance 0.25.

```
eps = rnorm(100, mean = 0, sd = sqrt(0.25))
```

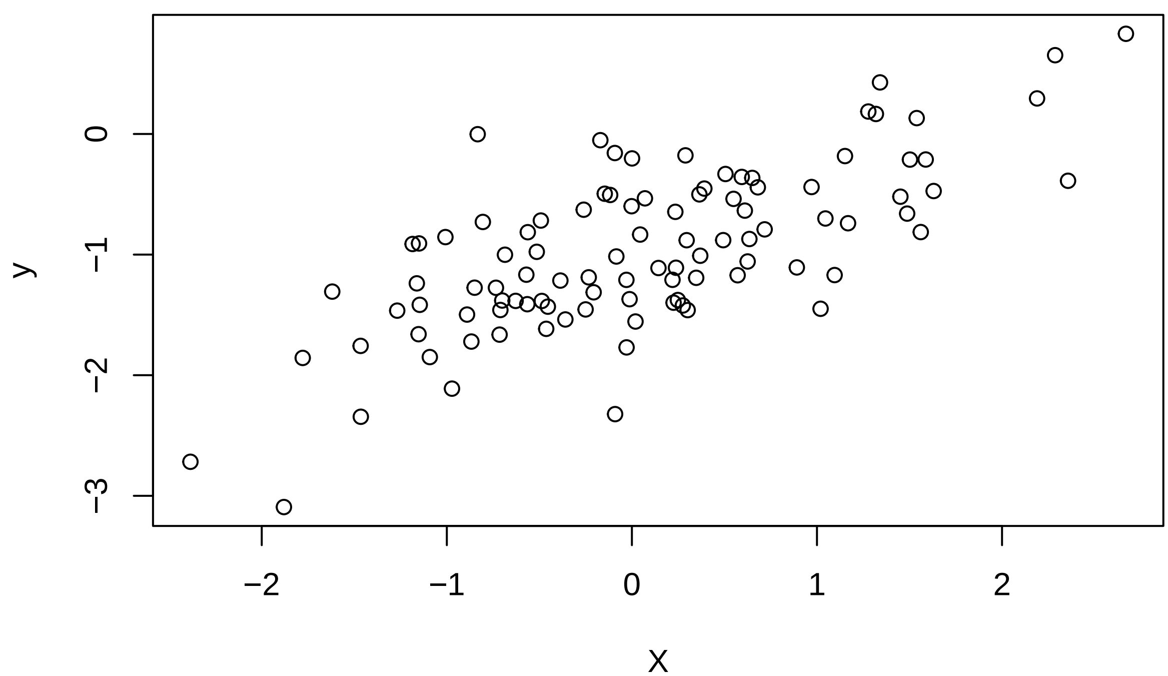
13c) Using x and eps, generate a vector y according to the model Y = -1 + 0.5X + eps. (3.39) What is the length of the vector y? What are the values of B0 and B1 in this linear model?

```
y = -1 + 0.5*X + eps
```

y is a vector of 100 elements, B0 is -1, B1 is 0.5

13d) Create a scatterplot displaying the relationship between x and y. Comment on what you observe.

```
plot(X, y)
```



There seems to be a moderately strong positive linear association between Y and X

13e) Fit a least squares linear model to predict y using x. Comment on the model obtained. How do B0 and B1 compare to B0 and B1?

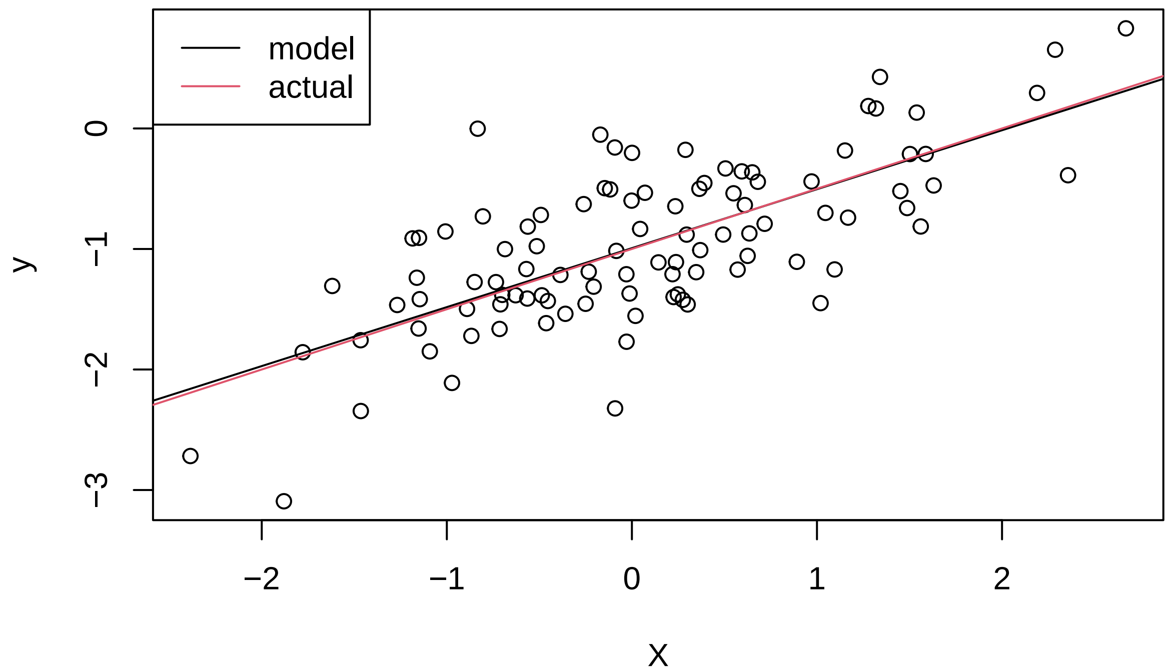
```r
reg = lm(y~X)
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28523 -0.32273 -0.03906  0.34713  1.39894
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.99310    0.04801  -20.68   <2e-16 ***
## X            0.48920    0.04864   10.06   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4791 on 98 degrees of freedom
## Multiple R-squared:  0.5079, Adjusted R-squared:  0.5029
## F-statistic: 101.2 on 1 and 98 DF,  p-value: < 2.2e-16
```

Both of the predicted variables are generally close to -1 and 0.5, both are statistically significant

13f) Display the least squares line on the scatterplot obtained in (d). Draw the population regression line on the plot, in a different color. Use the legend() command to create an appropriate legend.

```r
plot(X, y)
abline(reg, col=1)
abline(-1, 0.5, col=2)
legend("topleft",legend=c("model","actual"), col=1:2, lwd = 1)
```



13g) Now fit a polynomial regression model that predicts y using x and x2. Is there evidence that the quadratic term improves the model fit? Explain your answer.

```
squared = lm(y ~ X + I(X^2))
summary(squared)
```

```
##
## Call:
## lm(formula = y ~ X + I(X^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29365 -0.32485 -0.04037  0.33927  1.39942
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.984292   0.059463 -16.553  < 2e-16 ***
## X            0.492621   0.050702   9.716  5.5e-16 ***
## I(X^2)      -0.009268   0.036574  -0.253      0.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4814 on 97 degrees of freedom
## Multiple R-squared:  0.5083, Adjusted R-squared:  0.4981
## F-statistic: 50.13 on 2 and 97 DF,  p-value: 1.121e-15
```

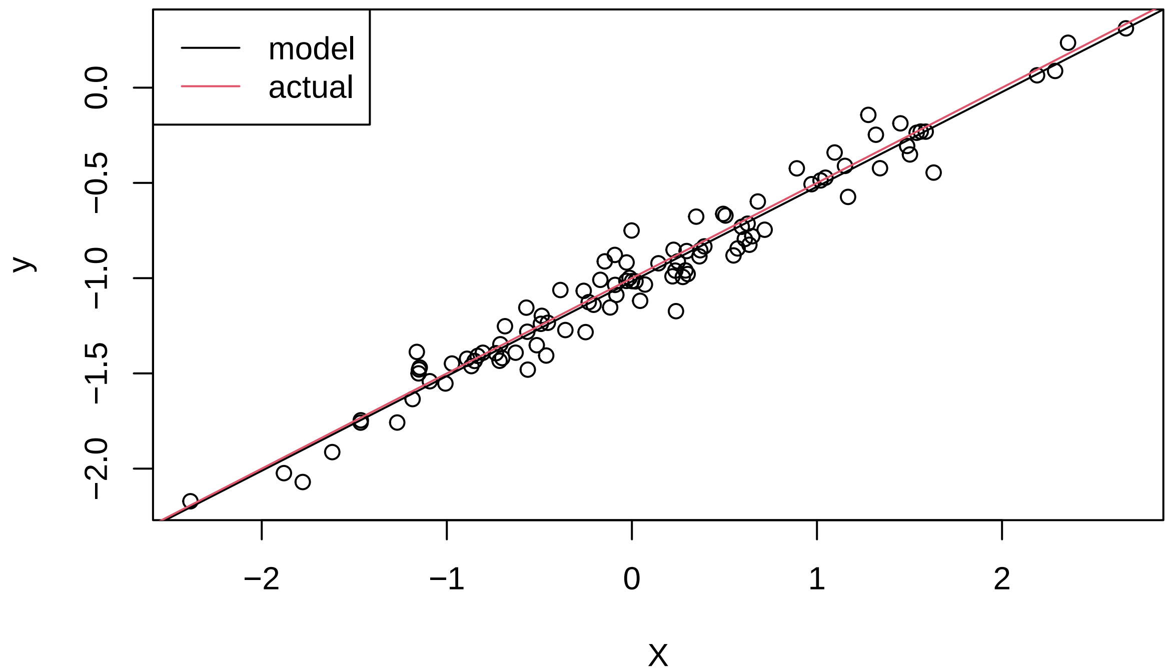
The R^2 value is every so slightly stronger than the previous one, but not by much

13h) Repeat (a)–(f) after modifying the data generation process in such a way that there is less noise in the data. The model (3.39) should remain the same. You can do this by decreasing the variance of the normal distribution used to generate the error term ϱ in (b). Describe your results.

```
eps = rnorm(100, 0, 0.1)
y = -1 + 0.5*X + eps
lessNoise = lm(y~X)
summary(lessNoise)
```

```
##
## Call:
## lm(formula = y ~ X)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.274689 -0.072071  0.000086  0.052510  0.268024
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.016990   0.009943 -102.28   <2e-16 ***
## X            0.497147   0.010073   49.36   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09921 on 98 degrees of freedom
## Multiple R-squared:  0.9613, Adjusted R-squared:  0.9609
## F-statistic:  2436 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
plot(X, y)
abline(lessNoise, col=1)
abline(-1, 0.5, col=2)
legend("topleft",legend=c("model","actual"), col=1:2, lwd = 1)
```



The R2 value is significantly stronger and there is a smaller gap between the trend lines

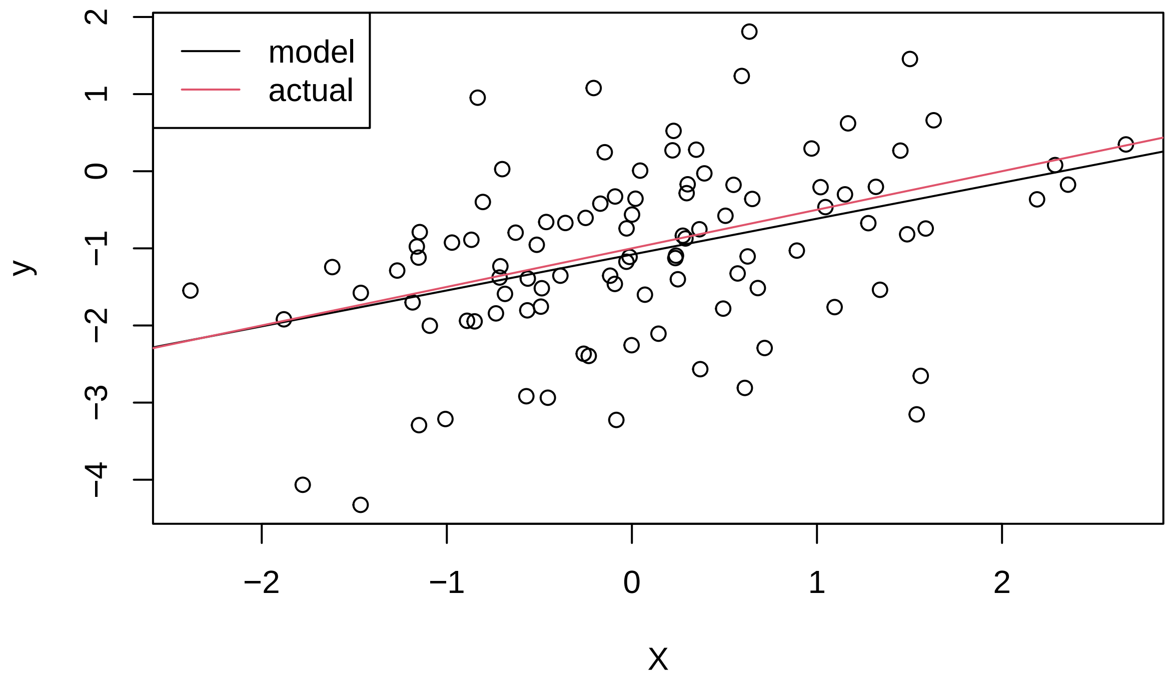
13i) Repeat (a)–(f) after modifying the data generation process in such a way that there is more noise in the data. The model (3.39) should remain the same. You can do this by increasing the variance of the normal distribution used to generate the error term ϱ in (b). Describe your results.

```
eps = rnorm(100, 0, 1)
y = -1 + 0.5*X + eps
moreNoise = lm(y~X)
summary(moreNoise)
```

```
##
## Call:
## lm(formula = y ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.78808 -0.44463  0.08497  0.64349  2.59540
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.0805     0.1066 -10.136  < 2e-16 ***
## X             0.4652     0.1080   4.308 3.92e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.064 on 98 degrees of freedom
```

```
## Multiple R-squared:  0.1592, Adjusted R-squared:  0.1506
## F-statistic: 18.56 on 1 and 98 DF,  p-value: 3.919e-05
```

```
plot(X, y)
abline(moreNoise, col=1)
abline(-1, 0.5, col=2)
legend("topleft",legend=c("model","actual"), col=1:2, lwd = 1)
```



The R2 value is significanly weeker adn there is a bigger gap between the trend lines

13j) What are the confidence intervals for B0 and B1 based on the original data set, the noisier data set, and the less noisy data set? Comment on your results.

```
confint(lessNoise)
```

```
##                  2.5 %     97.5 %
## (Intercept) -1.0367219 -0.9972589
## X            0.4771586  0.5171357
```

```
confint(reg)
```

```
##                  2.5 %     97.5 %
## (Intercept) -1.0883823 -0.8978176
## X            0.3926781  0.5857257
```

```r
confint(moreNoise)
```

```
##                  2.5 %     97.5 %
## (Intercept) -1.2921082 -0.8689838
## X            0.2509304  0.6795675
```

15) This problem involves the Boston data set, which we saw in the lab for this chapter. We will now try to predict per capita crime rate using the other variables in this data set. In other words, per capita crime rate is the response, and the other variables are the predictors.

15a) For each predictor, fit a simple linear regression model to predict the response. Describe your results. In which of the models is there a statistically significant association between the predictor and the response? Create some plots to back up your assertions.

```r
detach(Auto)
library(MASS)
attach(Boston)
reg = lm(crim ~ ., data=Boston)
ZN = lm(crim ~ zn)
summary(ZN)
```

```
##
## Call:
## lm(formula = crim ~ zn)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.429 -4.222 -2.620  1.250 84.523
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.45369    0.41722  10.675  < 2e-16 ***
## zn          -0.07393    0.01609  -4.594 5.51e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.435 on 504 degrees of freedom
## Multiple R-squared:  0.04019,    Adjusted R-squared:  0.03828
## F-statistic:  21.1 on 1 and 504 DF,  p-value: 5.506e-06
```

```r
INDUS = lm(crim ~ indus)
summary(INDUS)
```

```
##
## Call:
## lm(formula = crim ~ indus)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.972  -2.698  -0.736   0.712  81.813
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.06374    0.66723  -3.093  0.00209 **
## indus        0.50978    0.05102   9.991  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.866 on 504 degrees of freedom
## Multiple R-squared:  0.1653, Adjusted R-squared:  0.1637
## F-statistic: 99.82 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
CHAS = lm(crim ~ chas)
summary(CHAS)
```

```
##
## Call:
## lm(formula = crim ~ chas)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.738 -3.661 -3.435  0.018 85.232
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.7444     0.3961   9.453   <2e-16 ***
## chas         -1.8928     1.5061  -1.257    0.209
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.597 on 504 degrees of freedom
## Multiple R-squared:  0.003124,   Adjusted R-squared:  0.001146
## F-statistic: 1.579 on 1 and 504 DF,  p-value: 0.2094
```

```
NOX = lm(crim ~ nox)
summary(NOX)
```

```
##
## Call:
## lm(formula = crim ~ nox)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.371  -2.738  -0.974   0.559  81.728
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -13.720      1.699  -8.073 5.08e-15 ***
## nox           31.249      2.999  10.419  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.81 on 504 degrees of freedom
## Multiple R-squared:  0.1772, Adjusted R-squared:  0.1756
## F-statistic: 108.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

```r
RM = lm(crim ~ rm)
summary(RM)
```

```
##
## Call:
## lm(formula = crim ~ rm)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.604 -3.952 -2.654  0.989 87.197
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   20.482      3.365   6.088 2.27e-09 ***
## rm            -2.684      0.532  -5.045 6.35e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.401 on 504 degrees of freedom
## Multiple R-squared:  0.04807,    Adjusted R-squared:  0.04618
## F-statistic: 25.45 on 1 and 504 DF,  p-value: 6.347e-07
```

```r
AGE = lm(crim ~ age)
summary(AGE)
```

```
##
## Call:
## lm(formula = crim ~ age)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.789 -4.257 -1.230  1.527 82.849
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.77791    0.94398  -4.002 7.22e-05 ***
## age          0.10779    0.01274   8.463 2.85e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.057 on 504 degrees of freedom
## Multiple R-squared:  0.1244, Adjusted R-squared:  0.1227
## F-statistic: 71.62 on 1 and 504 DF,  p-value: 2.855e-16
```

```r
DIS = lm(crim ~ dis)
summary(DIS)
```

```
##
## Call:
## lm(formula = crim ~ dis)
##
```

```
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.708 -4.134 -1.527  1.516 81.674
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.4993     0.7304  13.006   <2e-16 ***
## dis          -1.5509     0.1683  -9.213   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.965 on 504 degrees of freedom
## Multiple R-squared:  0.1441, Adjusted R-squared:  0.1425
## F-statistic: 84.89 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
RAD = lm(crim ~ rad)
summary(RAD)
```

```
##
## Call:
## lm(formula = crim ~ rad)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.164  -1.381  -0.141   0.660  76.433
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.28716    0.44348  -5.157 3.61e-07 ***
## rad          0.61791    0.03433  17.998  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.718 on 504 degrees of freedom
## Multiple R-squared:  0.3913, Adjusted R-squared:   0.39
## F-statistic: 323.9 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
TAX = lm(crim ~ tax)
summary(TAX)
```

```
##
## Call:
## lm(formula = crim ~ tax)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.513  -2.738  -0.194   1.065  77.696
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.528369   0.815809  -10.45   <2e-16 ***
## tax          0.029742   0.001847   16.10   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.997 on 504 degrees of freedom
## Multiple R-squared:  0.3396, Adjusted R-squared:  0.3383
## F-statistic: 259.2 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
PTRATIO = lm(crim ~ ptratio)
summary(PTRATIO)
```

```
##
## Call:
## lm(formula = crim ~ ptratio)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.654 -3.985 -1.912  1.825 83.353
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.6469     3.1473  -5.607 3.40e-08 ***
## ptratio       1.1520     0.1694   6.801 2.94e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.24 on 504 degrees of freedom
## Multiple R-squared:  0.08407,    Adjusted R-squared:  0.08225
## F-statistic: 46.26 on 1 and 504 DF,  p-value: 2.943e-11
```

```
BLACK = lm(crim ~ black)
summary(BLACK)
```

```
##
## Call:
## lm(formula = crim ~ black)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.756  -2.299  -2.095  -1.296  86.822
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.553529   1.425903  11.609   <2e-16 ***
## black       -0.036280   0.003873  -9.367   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.946 on 504 degrees of freedom
## Multiple R-squared:  0.1483, Adjusted R-squared:  0.1466
## F-statistic: 87.74 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
LSTAT = lm(crim ~ lstat)
summary(LSTAT)
```

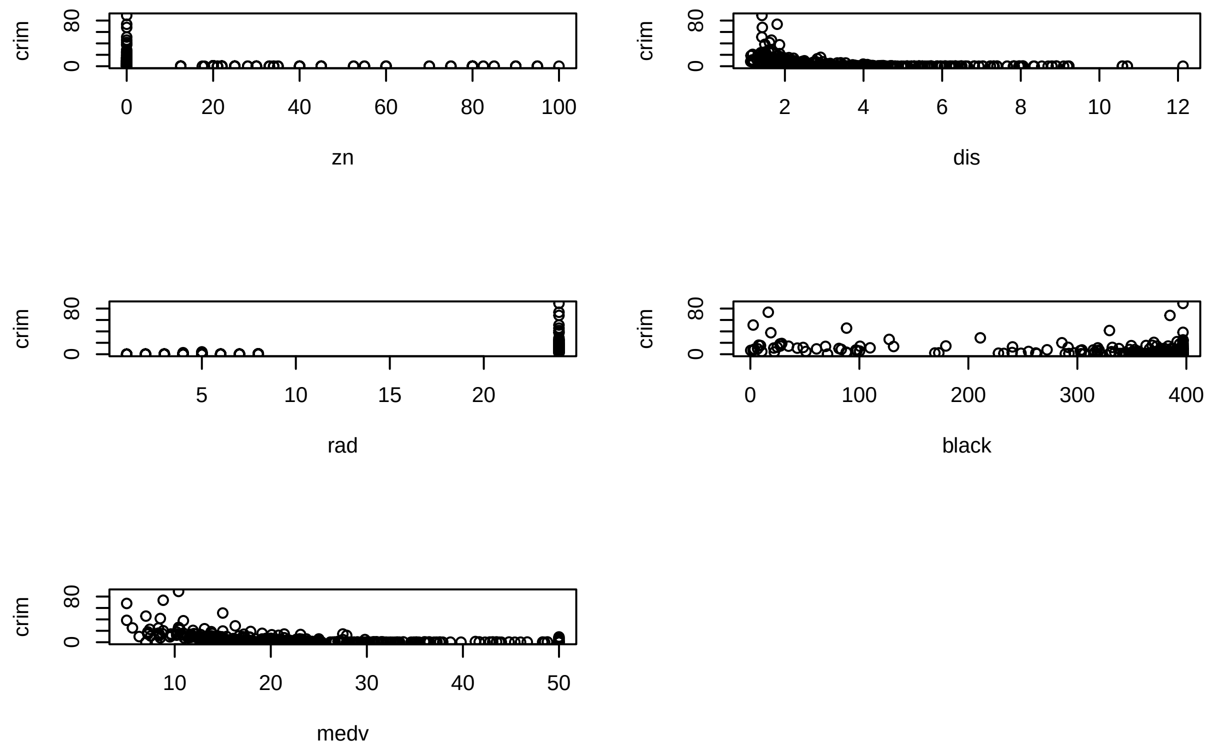
```
##
## Call:
## lm(formula = crim ~ lstat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.925  -2.822  -0.664   1.079  82.862
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.33054    0.69376  -4.801 2.09e-06 ***
## lstat        0.54880    0.04776  11.491  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.664 on 504 degrees of freedom
## Multiple R-squared:  0.2076, Adjusted R-squared:  0.206
## F-statistic:   132 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
MEDV = lm(crim ~ medv)
summary(MEDV)
```

```
##
## Call:
## lm(formula = crim ~ medv)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.071 -4.022 -2.343  1.298 80.957
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.79654    0.93419   12.63   <2e-16 ***
## medv        -0.36316    0.03839   -9.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.934 on 504 degrees of freedom
## Multiple R-squared:  0.1508, Adjusted R-squared:  0.1491
## F-statistic: 89.49 on 1 and 504 DF,  p-value: < 2.2e-16
```

Of all the variables, zn, dis, rad, black, and medev are significant

```
par(mfrow=c(3,2))
plot(zn, crim)
plot(dis, crim)
plot(rad, crim)
plot(black, crim)
plot(medv, crim)
```



15b) Fit a multiple regression model to predict the response using all of the predictors. Describe your results. For which predictors can we reject the null hypothesis H0 : Bj = 0?

```
reg = lm(crim ~ ., data=Boston)
summary(reg)
```

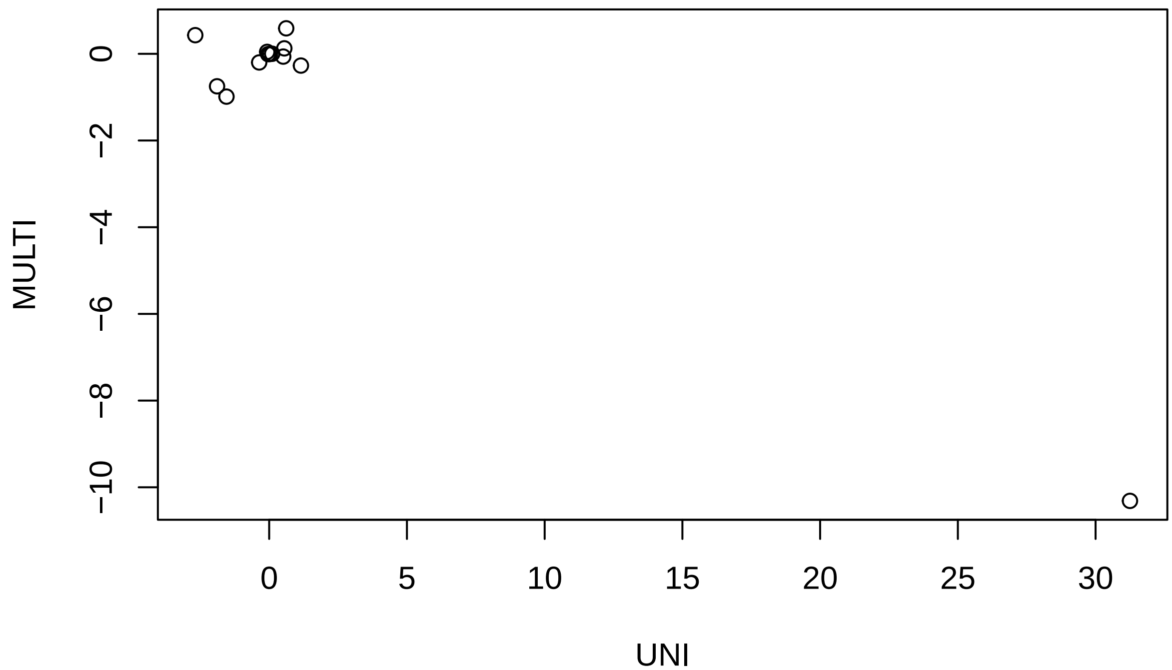
```
##
## Call:
## lm(formula = crim ~ ., data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.924 -2.120 -0.353  1.019 75.051
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.033228   7.234903   2.354 0.018949 *
## zn            0.044855   0.018734   2.394 0.017025 *
## indus        -0.063855   0.083407  -0.766 0.444294
## chas         -0.749134   1.180147  -0.635 0.525867
## nox         -10.313535   5.275536  -1.955 0.051152 .
## rm            0.430131   0.612830   0.702 0.483089
## age           0.001452   0.017925   0.081 0.935488
## dis          -0.987176   0.281817  -3.503 0.000502 ***
## rad           0.588209   0.088049   6.680 6.46e-11 ***
## tax          -0.003780   0.005156  -0.733 0.463793
```

```
## ptratio      -0.271081   0.186450  -1.454 0.146611
## black        -0.007538   0.003673  -2.052 0.040702 *
## lstat         0.126211   0.075725   1.667 0.096208 .
## medv         -0.198887   0.060516  -3.287 0.001087 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.439 on 492 degrees of freedom
## Multiple R-squared:  0.454,  Adjusted R-squared:  0.4396
## F-statistic: 31.47 on 13 and 492 DF,  p-value: < 2.2e-16
```

Based on the data at all standard significance levels we reject the null hypothesis for dis and rad. At the 0.01 level we reject med, and at the 0.05 level we reject zn, and black. We fail to reject the null for all other regressors

15c) How do your results from (a) compare to your results from (b)? Create a plot displaying the univariate regression coefficients from (a) on the x-axis, and the multiple regression coefficients from (b) on the y-axis. That is, each predictor is displayed as a single point in the plot. Its coefficient in a simple linear regression model is shown on the x-axis, and its coefficient estimate in the multiple linear regression model is shown on the y-axis.

```
coeff = c(ZN$coefficients[2], INDUS$coefficients[2],
          CHAS$coefficients[2], NOX$coefficients[2],
          RM$coefficients[2], AGE$coefficients[2],
          DIS$coefficients[2], RAD$coefficients[2],
          TAX$coefficients[2], PTRATIO$coefficients[2],
          BLACK$coefficients[2], LSTAT$coefficients[2],
          MEDV$coefficients[2])
plot(coeff, reg$coefficients[2:14], xlab = "UNI", ylab = "MULTI")
```



15d) Is there evidence of non-linear association between any of the predictors and the response?

```r
ZN = lm(crim ~ zn + I(zn^2) + I(zn^3))
summary(ZN)
```

```
##
## Call:
## lm(formula = crim ~ zn + I(zn^2) + I(zn^3))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.821 -4.614 -1.294  0.473 84.130
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.846e+00  4.330e-01  11.192  < 2e-16 ***
## zn          -3.322e-01  1.098e-01  -3.025  0.00261 **
## I(zn^2)      6.483e-03  3.861e-03   1.679  0.09375 .
## I(zn^3)     -3.776e-05  3.139e-05  -1.203  0.22954
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.372 on 502 degrees of freedom
## Multiple R-squared:  0.05824,    Adjusted R-squared:  0.05261
## F-statistic: 10.35 on 3 and 502 DF,  p-value: 1.281e-06
```

```r
INDUS = lm(crim ~ indus + I(indus^2) + I(indus^3))
summary(INDUS)
```

```
##
## Call:
## lm(formula = crim ~ indus + I(indus^2) + I(indus^3))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.278 -2.514  0.054  0.764 79.713
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.6625683  1.5739833   2.327   0.0204 *
## indus       -1.9652129  0.4819901  -4.077 5.30e-05 ***
## I(indus^2)   0.2519373  0.0393221   6.407 3.42e-10 ***
## I(indus^3)  -0.0069760  0.0009567  -7.292 1.20e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.423 on 502 degrees of freedom
## Multiple R-squared:  0.2597, Adjusted R-squared:  0.2552
## F-statistic: 58.69 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
CHAS = lm(crim ~ chas + I(chas^2) + I(chas^3))
summary(CHAS)
```

```
##
## Call:
## lm(formula = crim ~ chas + I(chas^2) + I(chas^3))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.738 -3.661 -3.435  0.018 85.232
##
## Coefficients: (2 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.7444     0.3961   9.453   <2e-16 ***
## chas         -1.8928     1.5061  -1.257    0.209
## I(chas^2)         NA         NA      NA       NA
## I(chas^3)         NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.597 on 504 degrees of freedom
## Multiple R-squared:  0.003124,   Adjusted R-squared:  0.001146
## F-statistic: 1.579 on 1 and 504 DF,  p-value: 0.2094
```

```r
NOX = lm(crim ~ nox + I(nox^2) + I(nox^3))
summary(NOX)
```

```
##
```

```
## Call:
## lm(formula = crim ~ nox + I(nox^2) + I(nox^3))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.110 -2.068 -0.255  0.739 78.302
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   233.09      33.64   6.928 1.31e-11 ***
## nox         -1279.37     170.40  -7.508 2.76e-13 ***
## I(nox^2)     2248.54     279.90   8.033 6.81e-15 ***
## I(nox^3)    -1245.70     149.28  -8.345 6.96e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.234 on 502 degrees of freedom
## Multiple R-squared:  0.297,  Adjusted R-squared:  0.2928
## F-statistic: 70.69 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
RM = lm(crim ~ rm + I(rm^2) + I(rm^3))
summary(RM)
```

```
##
## Call:
## lm(formula = crim ~ rm + I(rm^2) + I(rm^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.485  -3.468  -2.221  -0.015  87.219
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 112.6246    64.5172   1.746   0.0815 .
## rm          -39.1501    31.3115  -1.250   0.2118
## I(rm^2)       4.5509     5.0099   0.908   0.3641
## I(rm^3)      -0.1745     0.2637  -0.662   0.5086
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.33 on 502 degrees of freedom
## Multiple R-squared:  0.06779,    Adjusted R-squared:  0.06222
## F-statistic: 12.17 on 3 and 502 DF,  p-value: 1.067e-07
```

```r
AGE = lm(crim ~ age + I(age^2) + I(age^3))
summary(AGE)
```

```
##
## Call:
## lm(formula = crim ~ age + I(age^2) + I(age^3))
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -9.762 -2.673 -0.516  0.019 82.842
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.549e+00  2.769e+00  -0.920  0.35780
## age          2.737e-01  1.864e-01   1.468  0.14266
## I(age^2)    -7.230e-03  3.637e-03  -1.988  0.04738 *
## I(age^3)     5.745e-05  2.109e-05   2.724  0.00668 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.84 on 502 degrees of freedom
## Multiple R-squared:  0.1742, Adjusted R-squared:  0.1693
## F-statistic: 35.31 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
DIS = lm(crim ~ dis + I(dis^2) + I(dis^3))
summary(DIS)
```

```
##
## Call:
## lm(formula = crim ~ dis + I(dis^2) + I(dis^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.757  -2.588   0.031   1.267  76.378
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  30.0476     2.4459  12.285  < 2e-16 ***
## dis         -15.5543     1.7360  -8.960  < 2e-16 ***
## I(dis^2)      2.4521     0.3464   7.078 4.94e-12 ***
## I(dis^3)     -0.1186     0.0204  -5.814 1.09e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.331 on 502 degrees of freedom
## Multiple R-squared:  0.2778, Adjusted R-squared:  0.2735
## F-statistic: 64.37 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
RAD = lm(crim ~ rad + I(rad^2) + I(rad^3))
summary(RAD)
```

```
##
## Call:
## lm(formula = crim ~ rad + I(rad^2) + I(rad^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.381  -0.412  -0.269   0.179  76.217
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.605545   2.050108  -0.295    0.768
```

```
## rad          0.512736   1.043597   0.491    0.623
## I(rad^2)    -0.075177   0.148543  -0.506    0.613
## I(rad^3)     0.003209   0.004564   0.703    0.482
##
## Residual standard error: 6.682 on 502 degrees of freedom
## Multiple R-squared:    0.4,  Adjusted R-squared:  0.3965
## F-statistic: 111.6 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
TAX = lm(crim ~ tax + I(tax^2) + I(tax^3))
summary(TAX)
```

```
##
## Call:
## lm(formula = crim ~ tax + I(tax^2) + I(tax^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.273  -1.389   0.046   0.536  76.950
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.918e+01  1.180e+01   1.626    0.105
## tax         -1.533e-01  9.568e-02  -1.602    0.110
## I(tax^2)     3.608e-04  2.425e-04   1.488    0.137
## I(tax^3)    -2.204e-07  1.889e-07  -1.167    0.244
##
## Residual standard error: 6.854 on 502 degrees of freedom
## Multiple R-squared:  0.3689, Adjusted R-squared:  0.3651
## F-statistic:  97.8 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
PTRATIO = lm(crim ~ ptratio + I(ptratio^2) + I(ptratio^3))
summary(PTRATIO)
```

```
##
## Call:
## lm(formula = crim ~ ptratio + I(ptratio^2) + I(ptratio^3))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.833 -4.146 -1.655  1.408 82.697
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  477.18405  156.79498   3.043  0.00246 **
## ptratio      -82.36054   27.64394  -2.979  0.00303 **
## I(ptratio^2)   4.63535    1.60832   2.882  0.00412 **
## I(ptratio^3)  -0.08476    0.03090  -2.743  0.00630 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.122 on 502 degrees of freedom
## Multiple R-squared:  0.1138, Adjusted R-squared:  0.1085
## F-statistic: 21.48 on 3 and 502 DF,  p-value: 4.171e-13
```

```r
BLACK = lm(crim ~ black + I(black^2) + I(black^3))
summary(BLACK)
```

```
##
## Call:
## lm(formula = crim ~ black + I(black^2) + I(black^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.096  -2.343  -2.128  -1.439  86.790
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.826e+01  2.305e+00   7.924  1.5e-14 ***
## black       -8.356e-02  5.633e-02  -1.483    0.139
## I(black^2)   2.137e-04  2.984e-04   0.716    0.474
## I(black^3)  -2.652e-07  4.364e-07  -0.608    0.544
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.955 on 502 degrees of freedom
## Multiple R-squared:  0.1498, Adjusted R-squared:  0.1448
## F-statistic: 29.49 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
LSTAT = lm(crim ~ lstat + I(lstat^2) + I(lstat^3))
summary(LSTAT)
```

```
##
## Call:
## lm(formula = crim ~ lstat + I(lstat^2) + I(lstat^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.234  -2.151  -0.486   0.066  83.353
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.2009656  2.0286452   0.592   0.5541
## lstat       -0.4490656  0.4648911  -0.966   0.3345
## I(lstat^2)   0.0557794  0.0301156   1.852   0.0646 .
## I(lstat^3)  -0.0008574  0.0005652  -1.517   0.1299
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.629 on 502 degrees of freedom
## Multiple R-squared:  0.2179, Adjusted R-squared:  0.2133
## F-statistic: 46.63 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
MEDV = lm(crim ~ medv + I(medv^2) + I(medv^3))
summary(MEDV)
```

```
##
```

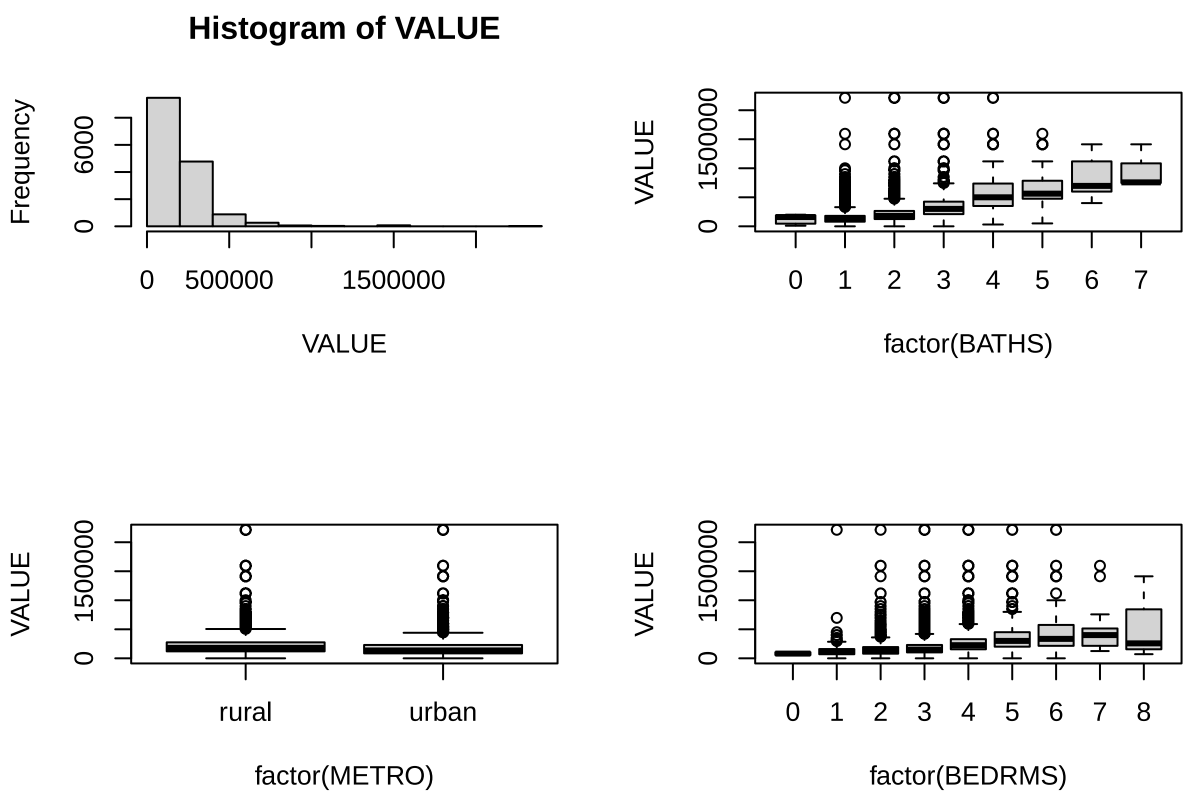
```
## Call:
## lm(formula = crim ~ medv + I(medv^2) + I(medv^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.427  -1.976  -0.437   0.439  73.655
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 53.1655381  3.3563105  15.840  < 2e-16 ***
## medv        -5.0948305  0.4338321 -11.744  < 2e-16 ***
## I(medv^2)    0.1554965  0.0171904   9.046  < 2e-16 ***
## I(medv^3)   -0.0014901  0.0002038  -7.312 1.05e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.569 on 502 degrees of freedom
## Multiple R-squared:  0.4202, Adjusted R-squared:  0.4167
## F-statistic: 121.3 on 3 and 502 DF,  p-value: < 2.2e-16
```

From this output it can be concluded that indux, nox, dis, ptratio, and medv have a relationship that maybe non-linear

Part 3

1) Plot some relationships and tell a story

```
homes = read.csv("homes2004.csv")
detach(Boston)
attach(homes)
par(mfrow=c(2,2))
hist(VALUE)
plot(VALUE ~ factor(BATHS))
plot(VALUE ~ factor(METRO))
plot(VALUE ~ factor(BEDRMS))
```



Most home are of lower value, and as the number of bathroom increases home value seems to increase. The Median value of rural homes also seems to be every so slightly higher than the urban values. Interestingly after adding more Bedrooms it seems as if home value levels off rather than increasing. There also seems to be more range in the price of home with bedrooms in between 2 and 7

2) Regress log value onto all but mortgage and purchase $.

```
logPrice = glm(log(VALUE) ~ . -AMMORT -LPRICE, data = homes)
summary(logPrice)
```

```
##
## Call:
## glm(formula = log(VALUE) ~ . - AMMORT - LPRICE, data = homes)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -13.2738   -0.1572    0.0574    0.2756    2.4649
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.159e+01  6.226e-02 186.232  < 2e-16 ***
## EAPTBLY          -4.347e-02  2.346e-02  -1.853  0.06390 .
## ECOM1Y           -2.568e-02  1.924e-02  -1.335  0.18202
## ECOM2Y           -8.645e-02  4.805e-02  -1.799  0.07205 .
## EGREENY           9.391e-03  1.400e-02   0.671  0.50249
## EJUNKY           -1.265e-01  5.105e-02  -2.478  0.01324 *
```

```
## ELOW1Y           2.870e-02  2.312e-02   1.241  0.21454
## ESFDY            2.945e-01  2.956e-02   9.963  < 2e-16 ***
## ETRANSY         -1.515e-02  2.532e-02  -0.598  0.54953
## EABANY          -1.621e-01  3.598e-02  -4.506 6.67e-06 ***
## HOWHgood         1.295e-01  2.632e-02   4.922 8.65e-07 ***
## HOWNgood         1.193e-01  2.190e-02   5.445 5.26e-08 ***
## ODORAY           1.026e-02  3.312e-02   0.310  0.75685
## STRNAY          -3.618e-02  1.607e-02  -2.251  0.02437 *
## ZINC2            6.244e-07  5.538e-08  11.273  < 2e-16 ***
## PER              9.651e-03  6.253e-03   1.543  0.12277
## ZADULT          -1.864e-02  1.088e-02  -1.714  0.08649 .
## HHGRADBach       1.321e-01  2.292e-02   5.766 8.28e-09 ***
## HHGRADGrad       1.973e-01  2.578e-02   7.652 2.09e-14 ***
## HHGRADHS Grad   -6.061e-02  2.171e-02  -2.792  0.00524 **
## HHGRADNo HS     -1.945e-01  3.183e-02  -6.112 1.01e-09 ***
## NUNITS          -9.324e-04  5.203e-04  -1.792  0.07314 .
## INTW            -4.637e-02  4.408e-03 -10.518  < 2e-16 ***
## METROurban       8.610e-02  1.807e-02   4.764 1.92e-06 ***
## STATECO         -2.921e-01  2.921e-02 -10.001  < 2e-16 ***
## STATECT         -3.464e-01  3.125e-02 -11.084  < 2e-16 ***
## STATEGA         -6.551e-01  3.108e-02 -21.077  < 2e-16 ***
## STATEIL         -8.618e-01  5.768e-02 -14.940  < 2e-16 ***
## STATEIN         -7.792e-01  3.070e-02 -25.379  < 2e-16 ***
## STATELA         -7.196e-01  3.688e-02 -19.511  < 2e-16 ***
## STATEMO         -6.645e-01  3.343e-02 -19.875  < 2e-16 ***
## STATEOH         -6.737e-01  3.269e-02 -20.610  < 2e-16 ***
## STATEOK         -9.982e-01  3.281e-02 -30.425  < 2e-16 ***
## STATEPA         -8.716e-01  3.389e-02 -25.722  < 2e-16 ***
## STATETX         -1.049e+00  3.431e-02 -30.575  < 2e-16 ***
## STATEWA         -1.228e-01  3.094e-02  -3.970 7.23e-05 ***
## BATHS            2.117e-01  1.159e-02  18.271  < 2e-16 ***
## BEDRMS           8.740e-02  1.006e-02   8.690  < 2e-16 ***
## MATBUYY         -2.966e-02  1.368e-02  -2.168  0.03015 *
## DWNPAYprev home  1.209e-01  1.785e-02   6.775 1.29e-11 ***
## FRSTHOY         -8.398e-02  1.724e-02  -4.870 1.12e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.6676521)
##
##     Null deviance: 14920  on 15564  degrees of freedom
## Residual deviance: 10365  on 15524  degrees of freedom
## AIC: 37926
##
## Number of Fisher Scoring iterations: 2
```

2i)How many coefficients are jointly significant at 10%? 34 of the coefficient are statically significant at the 0.1 alpha level

2ii)Re-run regression with only the significant covariates, and compare R2 to the full model.

```
p = summary(logPrice)$coefficients[-1,4]
names(p)[p > 0.1]
```

```
## [1] "ECOM1Y"  "EGREENY" "ELOW1Y"  "ETRANSY" "ODORAY"  "PER"
```

```r
logPrice2 = glm(log(VALUE) ~ . -AMMORT -LPRICE -ECOM1 -EGREEN -ELOW1 -ODORA -PER, data = homes)
summary(logPrice2)
```

```
##
## Call:
## glm(formula = log(VALUE) ~ . - AMMORT - LPRICE - ECOM1 - EGREEN -
##     ELOW1 - ODORA - PER, data = homes)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -13.2803   -0.1571    0.0576    0.2764    2.4799
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.160e+01  6.127e-02 189.341  < 2e-16 ***
## EAPTBLY         -4.346e-02  2.221e-02  -1.956  0.05043 .
## ECOM2Y          -9.215e-02  4.759e-02  -1.937  0.05282 .
## EJUNKY          -1.255e-01  5.088e-02  -2.466  0.01367 *
## ESFDY            2.888e-01  2.924e-02   9.877  < 2e-16 ***
## ETRANSY         -1.691e-02  2.510e-02  -0.674  0.50058
## EABANY          -1.622e-01  3.595e-02  -4.513 6.44e-06 ***
## HOWHgood         1.290e-01  2.630e-02   4.907 9.36e-07 ***
## HOWNgood         1.196e-01  2.186e-02   5.470 4.57e-08 ***
## STRNAY          -3.781e-02  1.590e-02  -2.378  0.01742 *
## ZINC2            6.240e-07  5.536e-08  11.272  < 2e-16 ***
## ZADULT          -9.286e-03  8.887e-03  -1.045  0.29609
## HHGRADBach       1.324e-01  2.291e-02   5.780 7.60e-09 ***
## HHGRADGrad       1.966e-01  2.575e-02   7.632 2.44e-14 ***
## HHGRADHS Grad   -6.127e-02  2.170e-02  -2.823  0.00477 **
## HHGRADNo HS     -1.955e-01  3.182e-02  -6.143 8.28e-10 ***
## NUNITS          -9.822e-04  5.197e-04  -1.890  0.05879 .
## INTW            -4.667e-02  4.406e-03 -10.593  < 2e-16 ***
## METROurban       8.339e-02  1.793e-02   4.651 3.33e-06 ***
## STATECO         -2.884e-01  2.905e-02  -9.930  < 2e-16 ***
## STATECT         -3.446e-01  3.121e-02 -11.042  < 2e-16 ***
## STATEGA         -6.565e-01  3.094e-02 -21.218  < 2e-16 ***
## STATEIL         -8.620e-01  5.763e-02 -14.958  < 2e-16 ***
## STATEIN         -7.779e-01  3.068e-02 -25.350  < 2e-16 ***
## STATELA         -7.209e-01  3.679e-02 -19.593  < 2e-16 ***
## STATEMO         -6.647e-01  3.342e-02 -19.886  < 2e-16 ***
## STATEOH         -6.747e-01  3.266e-02 -20.657  < 2e-16 ***
## STATEOK         -9.978e-01  3.279e-02 -30.431  < 2e-16 ***
## STATEPA         -8.680e-01  3.383e-02 -25.662  < 2e-16 ***
## STATETX         -1.049e+00  3.427e-02 -30.596  < 2e-16 ***
## STATEWA         -1.208e-01  3.091e-02  -3.908 9.33e-05 ***
## BATHS            2.133e-01  1.156e-02  18.452  < 2e-16 ***
## BEDRMS           9.003e-02  9.675e-03   9.306  < 2e-16 ***
## MATBUYY         -2.833e-02  1.365e-02  -2.075  0.03801 *
## DWNPAYprev home  1.220e-01  1.784e-02   6.837 8.41e-12 ***
## FRSTHOY         -8.269e-02  1.719e-02  -4.811 1.51e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for gaussian family taken to be 0.6676959)
##
##     Null deviance: 14920  on 15564  degrees of freedom
## Residual deviance: 10369  on 15529  degrees of freedom
## AIC: 37922
##
## Number of Fisher Scoring iterations: 2
```

3) Fit a regression for whether the buyer had >= 20% down (again, onto everything but AMMORT and LPRICE).

```r
homes$twentyDown = factor((LPRICE-AMMORT)/AMMORT > 0.2)
twenty = glm(twentyDown ~ . -AMMORT -LPRICE, data=homes, family='binomial')
summary(twenty)
```

```
##
## Call:
## glm(formula = twentyDown ~ . - AMMORT - LPRICE, family = "binomial",
##     data = homes)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6540  -0.8673  -0.6282   1.0747   2.3985
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -1.072e+00  1.768e-01  -6.063 1.34e-09 ***
## EAPTBLY          3.492e-02  6.682e-02   0.523 0.601269
## ECOM1Y          -1.046e-01  5.507e-02  -1.900 0.057477 .
## ECOM2Y          -3.586e-01  1.517e-01  -2.364 0.018086 *
## EGREENY         -8.876e-03  3.858e-02  -0.230 0.818070
## EJUNKY          -2.824e-02  1.539e-01  -0.184 0.854393
## ELOW1Y           5.808e-02  6.362e-02   0.913 0.361275
## ESFDY           -3.039e-01  8.013e-02  -3.793 0.000149 ***
## ETRANSY         -7.018e-02  7.274e-02  -0.965 0.334607
## EABANY          -1.827e-01  1.117e-01  -1.636 0.101829
## HOWHgood        -7.610e-02  7.631e-02  -0.997 0.318668
## HOWNgood         1.312e-01  6.391e-02   2.053 0.040106 *
## ODORAY           8.174e-02  9.493e-02   0.861 0.389215
## STRNAY          -9.464e-02  4.545e-02  -2.082 0.037322 *
## ZINC2           -4.767e-08  1.664e-07  -0.286 0.774500
## PER             -1.156e-01  1.783e-02  -6.487 8.77e-11 ***
## ZADULT           8.097e-03  3.070e-02   0.264 0.792002
## HHGRADBach       2.517e-01  6.362e-02   3.956 7.63e-05 ***
## HHGRADGrad       3.694e-01  7.071e-02   5.224 1.75e-07 ***
## HHGRADHS Grad    8.321e-03  6.130e-02   0.136 0.892033
## HHGRADNo HS     -1.139e-01  9.466e-02  -1.203 0.229103
## NUNITS           2.304e-03  1.473e-03   1.564 0.117763
## INTW            -7.718e-02  1.320e-02  -5.848 4.97e-09 ***
## METROurban      -7.224e-02  5.157e-02  -1.401 0.161269
## STATECO         -6.453e-02  8.217e-02  -0.785 0.432266
## STATECT          7.746e-01  8.605e-02   9.002  < 2e-16 ***
## STATEGA         -2.141e-01  9.065e-02  -2.361 0.018217 *
```

```
## STATEIL          4.493e-01  1.607e-01   2.797 0.005163 **
## STATEIN          1.966e-01  9.061e-02   2.170 0.029977 *
## STATELA          5.544e-01  1.043e-01   5.314 1.07e-07 ***
## STATEMO          4.572e-01  9.506e-02   4.809 1.51e-06 ***
## STATEOH          7.443e-01  9.238e-02   8.057 7.81e-16 ***
## STATEOK          5.709e-02  9.931e-02   0.575 0.565421
## STATEPA          5.405e-01  9.793e-02   5.519 3.41e-08 ***
## STATETX          2.581e-01  1.033e-01   2.498 0.012482 *
## STATEWA          2.207e-01  8.485e-02   2.601 0.009294 **
## BATHS            2.372e-01  3.347e-02   7.085 1.39e-12 ***
## BEDRMS          -1.223e-02  2.815e-02  -0.435 0.663830
## MATBUYY          3.809e-01  3.812e-02   9.991  < 2e-16 ***
## DWNPAYprev home  7.383e-01  4.717e-02  15.652  < 2e-16 ***
## VALUE            1.579e-06  1.512e-07  10.441  < 2e-16 ***
## FRSTHOY         -3.865e-01  4.887e-02  -7.909 2.60e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 20081  on 15564  degrees of freedom
## Residual deviance: 17896  on 15523  degrees of freedom
## AIC: 17980
##
## Number of Fisher Scoring iterations: 4
```

3i) Interpret effects for 1st home buyers and # of bathrooms.

```
firstTime = glm(twentyDown ~ . -AMMORT -LPRICE +FRSTHO*BATHS, data=homes, family='binomial')
summary(firstTime)
```

```
##
## Call:
## glm(formula = twentyDown ~ . - AMMORT - LPRICE + FRSTHO * BATHS,
##     family = "binomial", data = homes)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6500  -0.8639  -0.6268   1.0735   2.4069
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -1.145e+00  1.791e-01  -6.395 1.60e-10 ***
## EAPTBLY          3.257e-02  6.677e-02   0.488 0.625666
## ECOM1Y          -1.037e-01  5.504e-02  -1.883 0.059660 .
## ECOM2Y          -3.620e-01  1.515e-01  -2.389 0.016906 *
## EGREENY         -9.598e-03  3.861e-02  -0.249 0.803670
## EJUNKY          -2.462e-02  1.537e-01  -0.160 0.872764
## ELOW1Y           6.048e-02  6.361e-02   0.951 0.341759
## ESFDY           -3.072e-01  8.011e-02  -3.834 0.000126 ***
## ETRANSY         -6.924e-02  7.270e-02  -0.953 0.340841
## EABANY          -1.902e-01  1.115e-01  -1.706 0.088087 .
## HOWHgood        -7.277e-02  7.623e-02  -0.955 0.339750
```

```
## HOWNgood         1.339e-01  6.389e-02   2.096 0.036065 *
## ODORAY           8.028e-02  9.487e-02   0.846 0.397435
## STRNAY          -9.486e-02  4.544e-02  -2.088 0.036826 *
## ZINC2           -6.297e-08  1.686e-07  -0.373 0.708855
## PER             -1.167e-01  1.785e-02  -6.537 6.29e-11 ***
## ZADULT           1.011e-02  3.074e-02   0.329 0.742171
## HHGRADBach       2.532e-01  6.362e-02   3.980 6.89e-05 ***
## HHGRADGrad       3.727e-01  7.075e-02   5.268 1.38e-07 ***
## HHGRADHS Grad    9.253e-03  6.128e-02   0.151 0.879988
## HHGRADNo HS     -1.175e-01  9.460e-02  -1.242 0.214291
## NUNITS           2.223e-03  1.462e-03   1.520 0.128536
## INTW            -7.790e-02  1.319e-02  -5.905 3.53e-09 ***
## METROurban      -7.549e-02  5.157e-02  -1.464 0.143261
## STATECO         -7.244e-02  8.237e-02  -0.879 0.379173
## STATECT          7.637e-01  8.615e-02   8.865  < 2e-16 ***
## STATEGA         -2.216e-01  9.092e-02  -2.437 0.014812 *
## STATEIL          4.387e-01  1.607e-01   2.730 0.006333 **
## STATEIN          1.914e-01  9.074e-02   2.110 0.034883 *
## STATELA          5.507e-01  1.044e-01   5.274 1.34e-07 ***
## STATEMO          4.477e-01  9.521e-02   4.702 2.58e-06 ***
## STATEOH          7.330e-01  9.248e-02   7.926 2.26e-15 ***
## STATEOK          4.730e-02  9.947e-02   0.476 0.634401
## STATEPA          5.248e-01  9.813e-02   5.348 8.89e-08 ***
## STATETX          2.527e-01  1.034e-01   2.443 0.014554 *
## STATEWA          2.212e-01  8.492e-02   2.605 0.009200 **
## BATHS            2.843e-01  3.794e-02   7.494 6.66e-14 ***
## BEDRMS          -1.280e-02  2.818e-02  -0.454 0.649600
## MATBUYY          3.813e-01  3.814e-02   9.997  < 2e-16 ***
## DWNPAYprev home  7.316e-01  4.728e-02  15.473  < 2e-16 ***
## VALUE            1.543e-06  1.518e-07  10.168  < 2e-16 ***
## FRSTHOY         -1.130e-01  1.129e-01  -1.001 0.316992
## BATHS:FRSTHOY   -1.581e-01  5.898e-02  -2.680 0.007369 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 20081  on 15564  degrees of freedom
## Residual deviance: 17889  on 15522  degrees of freedom
## AIC: 17975
##
## Number of Fisher Scoring iterations: 4
```

If the house has more bathroom then a first time buyer seems to be less likely to pay the 20% downpayment as opposed to someone who is not a first time buyer

4) Re-fit your model from Q3 for only homes worth > 100k. Compare in-sample fit to R2 for predicting homes worth < 100k.

```
greaterThan100 = glm(twentyDown ~ . -AMMORT -LPRICE, data=homes, subset = VALUE > 100000, family = 'bin
summary(greaterThan100)
```

```
##
```

```
## Call:
## glm(formula = twentyDown ~ . - AMMORT - LPRICE, family = "binomial",
##     data = homes, subset = VALUE > 1e+05)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7509  -0.9058  -0.6354   1.0866   2.5029
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.251e+00  2.061e-01  -6.068 1.30e-09 ***
## EAPTBLY           9.379e-02  7.827e-02   1.198 0.230850
## ECOM1Y           -3.698e-02  6.368e-02  -0.581 0.561445
## ECOM2Y           -3.414e-01  1.914e-01  -1.783 0.074536 .
## EGREENY          -9.157e-03  4.260e-02  -0.215 0.829808
## EJUNKY           -2.157e-01  2.042e-01  -1.056 0.290832
## ELOW1Y            5.298e-02  7.097e-02   0.746 0.455375
## ESFDY            -3.966e-01  9.603e-02  -4.130 3.63e-05 ***
## ETRANSY          -1.207e-01  8.468e-02  -1.426 0.153879
## EABANY           -3.325e-01  1.553e-01  -2.141 0.032266 *
## HOWHgood          7.075e-03  9.360e-02   0.076 0.939748
## HOWNgood          1.536e-01  7.663e-02   2.004 0.045065 *
## ODORAY            1.138e-01  1.125e-01   1.011 0.311797
## STRNAY           -1.119e-01  5.221e-02  -2.143 0.032145 *
## ZINC2            -1.789e-07  2.037e-07  -0.878 0.379711
## PER              -1.194e-01  1.967e-02  -6.068 1.29e-09 ***
## ZADULT           -6.276e-03  3.446e-02  -0.182 0.855482
## HHGRADBach        2.695e-01  7.039e-02   3.830 0.000128 ***
## HHGRADGrad        3.666e-01  7.743e-02   4.735 2.19e-06 ***
## HHGRADHS Grad     3.243e-02  6.963e-02   0.466 0.641424
## HHGRADNo HS      -2.303e-01  1.192e-01  -1.932 0.053299 .
## NUNITS            1.718e-03  1.942e-03   0.885 0.376306
## INTW             -7.717e-02  1.651e-02  -4.675 2.93e-06 ***
## METROurban       -1.147e-01  6.034e-02  -1.900 0.057385 .
## STATECO          -1.679e-02  8.447e-02  -0.199 0.842446
## STATECT           7.573e-01  8.965e-02   8.447  < 2e-16 ***
## STATEGA          -2.320e-01  9.517e-02  -2.438 0.014761 *
## STATEIL           4.134e-01  1.944e-01   2.127 0.033448 *
## STATEIN           2.968e-01  9.849e-02   3.014 0.002582 **
## STATELA           6.438e-01  1.178e-01   5.467 4.59e-08 ***
## STATEMO           5.498e-01  1.029e-01   5.344 9.08e-08 ***
## STATEOH           7.532e-01  9.994e-02   7.536 4.83e-14 ***
## STATEOK           1.944e-01  1.181e-01   1.646 0.099684 .
## STATEPA           6.840e-01  1.144e-01   5.981 2.21e-09 ***
## STATETX           4.285e-01  1.255e-01   3.414 0.000640 ***
## STATEWA           2.397e-01  8.728e-02   2.746 0.006039 **
## BATHS             2.267e-01  3.683e-02   6.156 7.48e-10 ***
## BEDRMS            1.423e-03  3.135e-02   0.045 0.963798
## MATBUYY           4.713e-01  4.214e-02  11.183  < 2e-16 ***
## DWNPAYprev home   7.882e-01  5.165e-02  15.259  < 2e-16 ***
## VALUE             1.796e-06  1.674e-07  10.728  < 2e-16 ***
## FRSTHOY          -3.363e-01  5.606e-02  -6.000 1.97e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 16099  on 12143  degrees of freedom
## Residual deviance: 14320  on 12102  degrees of freedom
## AIC: 14404
##
## Number of Fisher Scoring iterations: 4
```

```
lessThan100 = glm(twentyDown ~ . -AMMORT -LPRICE, data=homes, subset = VALUE < 100000, family = 'binomia
summary(lessThan100)
```

```
##
## Call:
## glm(formula = twentyDown ~ . - AMMORT - LPRICE, family = "binomial",
##     data = homes, subset = VALUE < 1e+05)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9206  -0.7197  -0.5720  -0.3986   2.3975
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       8.708e-01  5.117e-01   1.702 0.088816 .
## EAPTBLY          -1.608e-01  1.391e-01  -1.155 0.247899
## ECOM1Y           -2.590e-01  1.193e-01  -2.171 0.029907 *
## ECOM2Y           -5.894e-01  2.743e-01  -2.149 0.031658 *
## EGREENY          -6.348e-02  1.021e-01  -0.621 0.534315
## EJUNKY            1.155e-01  2.436e-01   0.474 0.635468
## ELOW1Y            1.795e-01  1.627e-01   1.103 0.269926
## ESFDY             2.781e-02  1.635e-01   0.170 0.864965
## ETRANSY           1.892e-01  1.505e-01   1.257 0.208820
## EABANY           -5.284e-02  1.693e-01  -0.312 0.754928
## HOWHgood         -2.460e-01  1.414e-01  -1.740 0.081811 .
## HOWNgood          5.157e-02  1.258e-01   0.410 0.681781
## ODORAY           -5.340e-02  1.947e-01  -0.274 0.783846
## STRNAY           -4.406e-02  1.012e-01  -0.435 0.663282
## ZINC2             1.299e-06  8.688e-07   1.495 0.134850
## PER              -8.111e-02  4.677e-02  -1.734 0.082897 .
## ZADULT            4.036e-02  7.632e-02   0.529 0.596919
## HHGRADBach        1.626e-01  1.767e-01   0.920 0.357555
## HHGRADGrad        3.854e-01  2.119e-01   1.819 0.068969 .
## HHGRADHS Grad     4.266e-03  1.429e-01   0.030 0.976181
## HHGRADNo HS       7.843e-02  1.809e-01   0.434 0.664545
## NUNITS            2.674e-03  2.909e-03   0.919 0.357958
## INTW             -6.479e-02  2.326e-02  -2.786 0.005335 **
## METROurban        4.077e-02  1.113e-01   0.366 0.714276
## STATECO          -1.266e+00  5.074e-01  -2.495 0.012582 *
## STATECT           2.304e-01  4.164e-01   0.553 0.580061
## STATEGA          -1.067e+00  4.250e-01  -2.511 0.012043 *
## STATEIL          -5.512e-01  4.451e-01  -1.238 0.215580
## STATEIN          -1.189e+00  3.924e-01  -3.031 0.002440 **
## STATELA          -6.636e-01  4.000e-01  -1.659 0.097139 .
## STATEMO          -8.389e-01  4.024e-01  -2.085 0.037064 *
```

```
## STATEOH         -1.791e-01  3.966e-01  -0.452 0.651623
## STATEOK         -1.323e+00  3.843e-01  -3.441 0.000579 ***
## STATEPA         -8.440e-01  3.896e-01  -2.166 0.030310 *
## STATETX         -1.034e+00  3.838e-01  -2.694 0.007065 **
## STATEWA         -6.715e-01  4.706e-01  -1.427 0.153630
## BATHS            1.655e-01  9.732e-02   1.700 0.089102 .
## BEDRMS          -8.695e-02  6.971e-02  -1.247 0.212334
## MATBUYY         -5.501e-02  9.800e-02  -0.561 0.574579
## DWNPAYprev home  5.440e-01  1.348e-01   4.036 5.44e-05 ***
## VALUE           -5.156e-06  1.881e-06  -2.742 0.006109 **
## FRSTHOY         -5.042e-01  1.094e-01  -4.607 4.08e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3322.7  on 3087  degrees of freedom
## Residual deviance: 3062.9  on 3046  degrees of freedom
## AIC: 3146.9
##
## Number of Fisher Scoring iterations: 4
```

R^2 for greater than 100k: 1 - (14320/16099) = 0.110 R^2 for less than 100k: 1 - (3062.9/3322.7) = 0.078 The R^2 for greater than 100k is stronger than that of R^2 for less thank 100k