SDS323 HW1

Cherry Sun, Andrew Chen, Cristian Sigala, Bailey Brady

3/2/2021

Section 2.4 Question 8

a)

```
college <- read.csv(file = 'College.csv')
head(college)
```

```
##                              X Private Apps Accept Enroll Top10perc
Top25perc
## 1 Abilene Christian University     Yes 1660   1232    721        23
52
## 2           Adelphi University     Yes 2186   1924    512        16
29
## 3               Adrian College     Yes 1428   1097    336        22
50
## 4          Agnes Scott College     Yes  417    349    137        60
89
## 5    Alaska Pacific University     Yes  193    146     55        16
44
## 6            Albertson College     Yes  587    479    158        38
62
##   F.Undergrad P.Undergrad Outstate Room.Board Books Personal PhD Terminal
## 1        2885         537     7440       3300   450     2200  70       78
## 2        2683        1227    12280       6450   750     1500  29       30
## 3        1036          99    11250       3750   400     1165  53       66
## 4         510          63    12960       5450   450      875  92       97
## 5         249         869     7560       4120   800     1500  76       72
## 6         678          41    13500       3335   500      675  67       73
##   S.F.Ratio perc.alumni Expend Grad.Rate
## 1      18.1          12   7041        60
## 2      12.2          16  10527        56
## 3      12.9          30   8735        54
## 4       7.7          37  19016        59
## 5      11.9           2  10922        15
## 6       9.4          11   9727        55
```

b)

```
#rownames(college) = college[,1]
#fix(college)
college = college[,-1]
#fix(college)
```

```
head(college)
```

```
##   Private Apps Accept Enroll Top10perc Top25perc F.Undergrad P.Undergrad
## 1     Yes 1660   1232    721        23        52        2885         537
## 2     Yes 2186   1924    512        16        29        2683        1227
## 3     Yes 1428   1097    336        22        50        1036          99
## 4     Yes  417    349    137        60        89         510          63
## 5     Yes  193    146     55        16        44         249         869
## 6     Yes  587    479    158        38        62         678          41
##   Outstate Room.Board Books Personal PhD Terminal S.F.Ratio perc.alumni
Expend
## 1     7440       3300   450     2200  70       78      18.1          12
7041
## 2    12280       6450   750     1500  29       30      12.2          16
10527
## 3    11250       3750   400     1165  53       66      12.9          30
8735
## 4    12960       5450   450      875  92       97       7.7          37
19016
## 5     7560       4120   800     1500  76       72      11.9           2
10922
## 6    13500       3335   500      675  67       73       9.4          11
9727
##   Grad.Rate
## 1        60
## 2        56
## 3        54
## 4        59
## 5        15
## 6        55
```

c)

i.

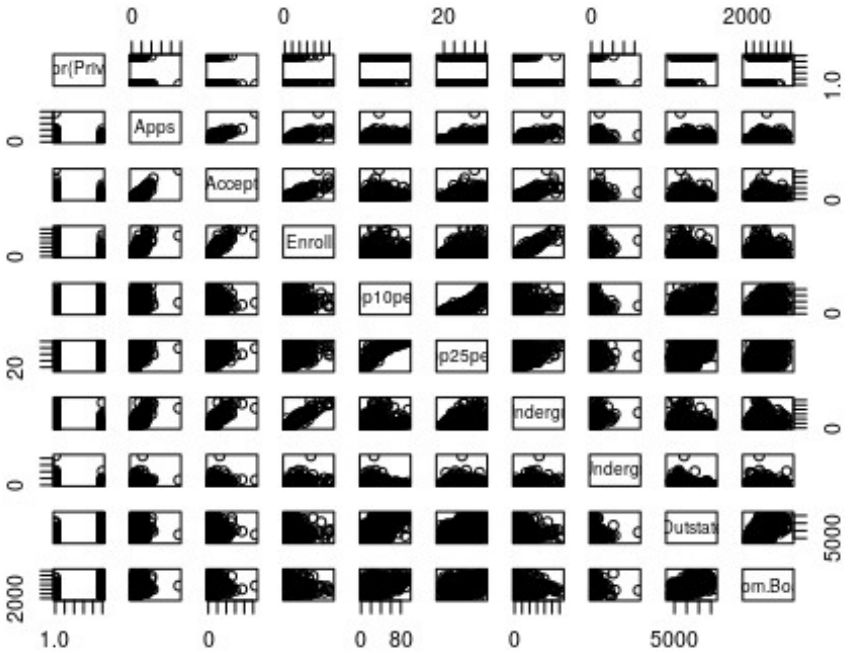
```
summary(college)
```

```
##  Private        Apps           Accept          Enroll       Top10perc
##  No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00
##  Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00
##            Median : 1558   Median : 1110   Median : 434   Median :23.00
##            Mean   : 3002   Mean   : 2019   Mean   : 780   Mean   :27.56
##            3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.00
##            Max.   :48094   Max.   :26330   Max.   :6392   Max.   :96.00
##    Top25perc      F.Undergrad     P.Undergrad         Outstate
##  Min.   :  9.0   Min.   :  139   Min.   :    1.0   Min.   : 2340
##  1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320
##  Median : 54.0   Median : 1707   Median :  353.0   Median : 9990
##  Mean   : 55.8   Mean   : 3700   Mean   :  855.3   Mean   :10441
##  3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925
```

```
##  Max.   :100.0   Max.   :31643   Max.   :21836.0   Max.   :21700
##    Room.Board       Books           Personal         PhD
##  Min.   :1780   Min.   :  96.0   Min.   : 250   Min.   :  8.00
##  1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850   1st Qu.: 62.00
##  Median :4200   Median : 500.0   Median :1200   Median : 75.00
##  Mean   :4358   Mean   : 549.4   Mean   :1341   Mean   : 72.66
##  3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.: 85.00
##  Max.   :8124   Max.   :2340.0   Max.   :6800   Max.   :103.00
##     Terminal       S.F.Ratio      perc.alumni        Expend
##  Min.   : 24.0   Min.   : 2.50   Min.   : 0.00   Min.   : 3186
##  1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00   1st Qu.: 6751
##  Median : 82.0   Median :13.60   Median :21.00   Median : 8377
##  Mean   : 79.7   Mean   :14.09   Mean   :22.74   Mean   : 9660
##  3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00   3rd Qu.:10830
##  Max.   :100.0   Max.   :39.80   Max.   :64.00   Max.   :56233
##    Grad.Rate
##  Min.   : 10.00
##  1st Qu.: 53.00
##  Median : 65.00
##  Mean   : 65.46
##  3rd Qu.: 78.00
##  Max.   :118.00
```

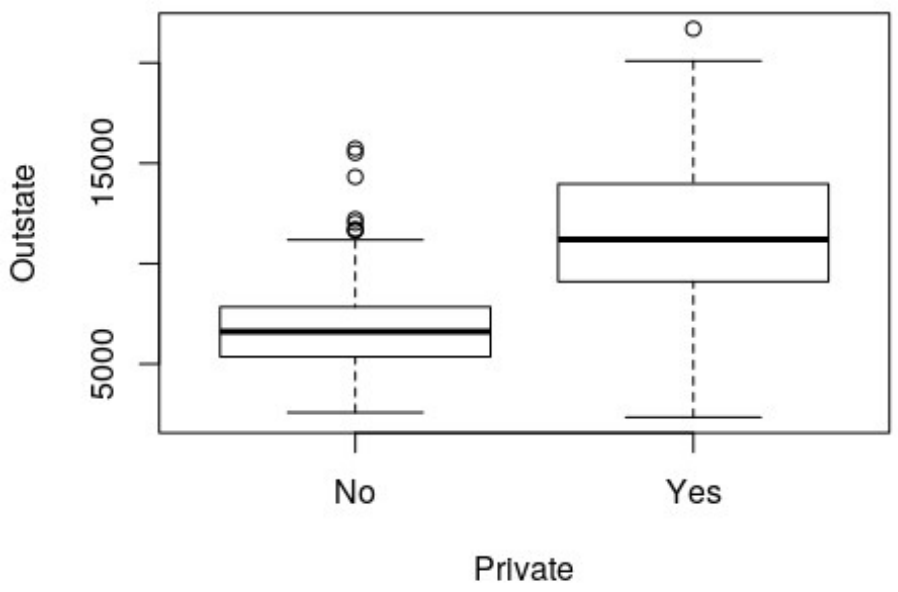
ii.

```
pairs(~ factor(Private) + Apps + Accept + Enroll + Top10perc + Top25perc +
F.Undergrad + P.Undergrad + Outstate + Room.Board, data = college)
```



iii.

```
plot(factor(college$Private), college$Outstate, xlab = "Private", ylab =
"Outstate")
```



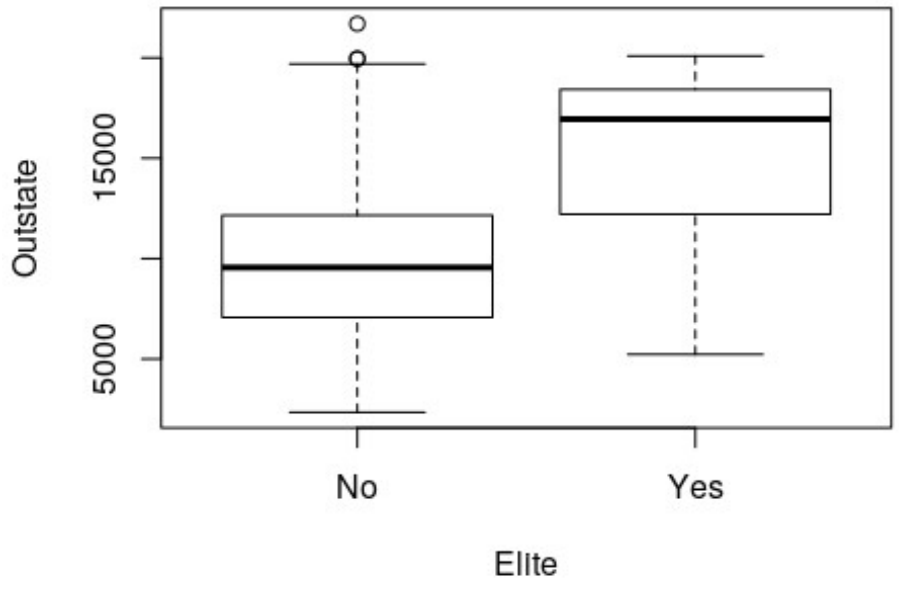
iv.

```
Elite = rep("No", nrow(college))
Elite[college$Top10perc > 50] = "Yes"
Elite = as.factor(Elite)
college = data.frame(college, Elite)

summary(college)
```

```
##  Private        Apps           Accept          Enroll       Top10perc
##  No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00
##  Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00
##            Median : 1558   Median : 1110   Median : 434   Median :23.00
##            Mean   : 3002   Mean   : 2019   Mean   : 780   Mean   :27.56
##            3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.00
##            Max.   :48094   Max.   :26330   Max.   :6392   Max.   :96.00
##    Top25perc      F.Undergrad     P.Undergrad         Outstate
##  Min.   :  9.0   Min.   :  139   Min.   :    1.0   Min.   : 2340
##  1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320
##  Median : 54.0   Median : 1707   Median :  353.0   Median : 9990
##  Mean   : 55.8   Mean   : 3700   Mean   :  855.3   Mean   :10441
##  3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925
##  Max.   :100.0   Max.   :31643   Max.   :21836.0   Max.   :21700
```

```
##    Room.Board       Books           Personal         PhD
##  Min.   :1780   Min.   :  96.0   Min.   : 250   Min.   :  8.00
##  1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850   1st Qu.: 62.00
##  Median :4200   Median : 500.0   Median :1200   Median : 75.00
##  Mean   :4358   Mean   : 549.4   Mean   :1341   Mean   : 72.66
##  3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.: 85.00
##  Max.   :8124   Max.   :2340.0   Max.   :6800   Max.   :103.00
##     Terminal       S.F.Ratio      perc.alumni        Expend
##  Min.   : 24.0   Min.   : 2.50   Min.   : 0.00   Min.   : 3186
##  1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00   1st Qu.: 6751
##  Median : 82.0   Median :13.60   Median :21.00   Median : 8377
##  Mean   : 79.7   Mean   :14.09   Mean   :22.74   Mean   : 9660
##  3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00   3rd Qu.:10830
##  Max.   :100.0   Max.   :39.80   Max.   :64.00   Max.   :56233
##    Grad.Rate      Elite
##  Min.   : 10.00   No :699
##  1st Qu.: 53.00   Yes: 78
##  Median : 65.00
##  Mean   : 65.46
##  3rd Qu.: 78.00
##  Max.   :118.00
```

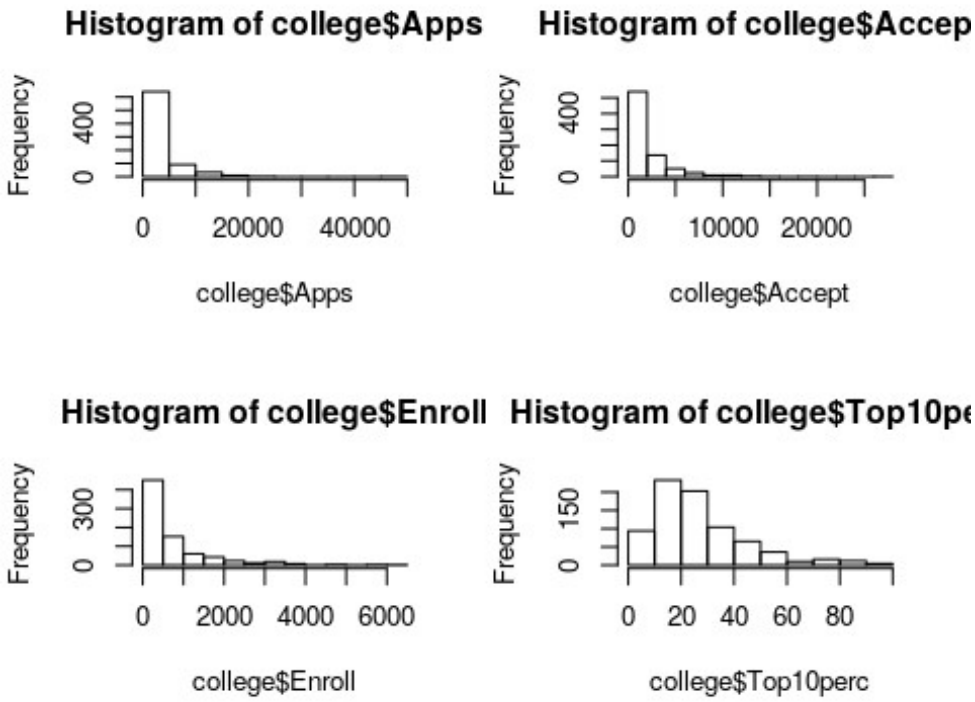
```r
plot(college$Elite, college$Outstate, xlab = "Elite", ylab = "Outstate")
```



v.

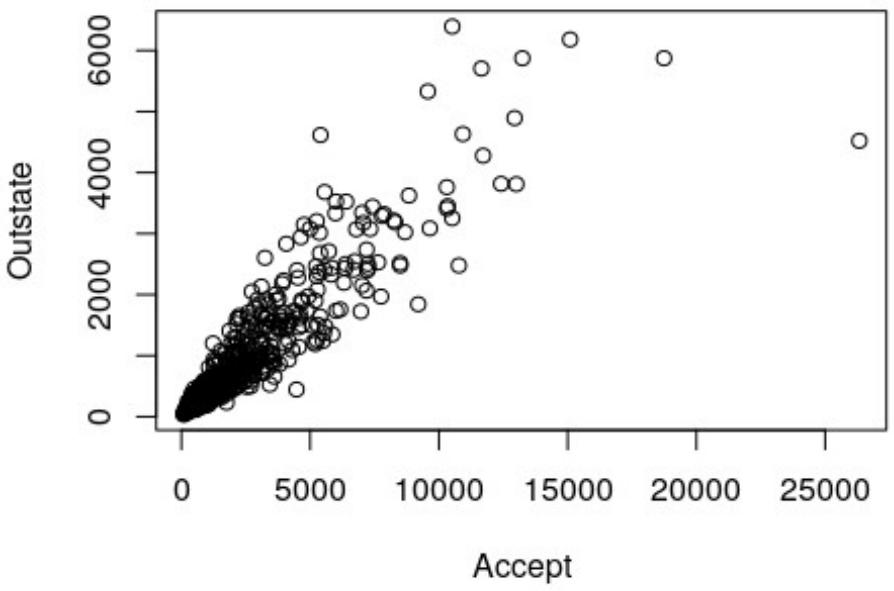
```r
par(mfrow = c(2,2))
hist(college$Apps)
```

```
hist(college$Accept)
hist(college$Enroll)
hist(college$Top10perc)
```



vi.

```
par(mfrow = c(1,1))
plot(college$Accept, college$Enroll, xlab = "Accept", ylab = "Outstate")
```



Here looks to be a positive correlation between people accepting the college offer and people enrolling in the college.

Section 2.4 Question 10

a)

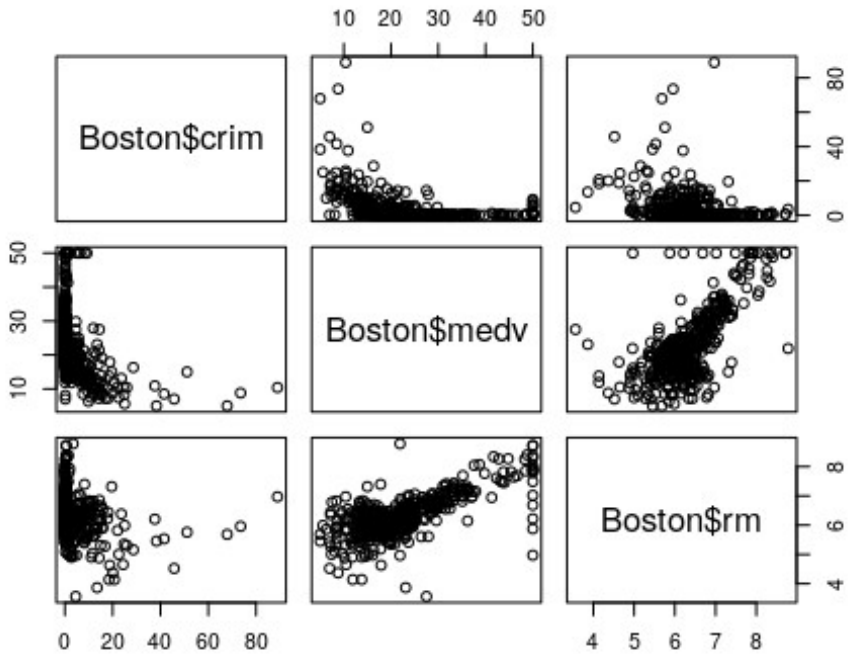
```
library(MASS)

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

head(Boston)

##      crim zn indus chas   nox    rm  age    dis rad tax ptratio  black
lstat
## 1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90
4.98
## 2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8 396.90
9.14
## 3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242    17.8 392.83
4.03
```

```
## 4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222    18.7 394.63
2.94
## 5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7 396.90
5.33
## 6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222    18.7 394.12
5.21
##   medv
## 1 24.0
## 2 21.6
## 3 34.7
## 4 33.4
## 5 36.2
## 6 28.7

#?Boston
dim(Boston)

## [1] 506  14
```

There are 506 rows and 14 columns in this data set. The rows represent each town/suburb within Boston. The columms represent the different values that we can use to identify the town/suburb with. For example, crim represents the per capita crime rate by town, indus represents the proportion of non-retail business acres per town, etc.

b)

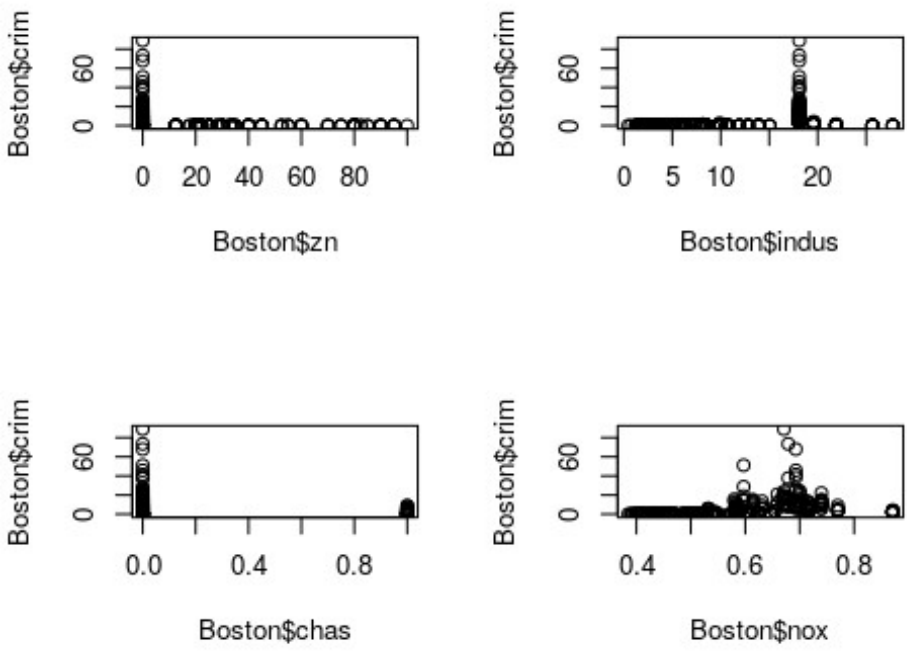
```
pairs(~ Boston$crim + Boston$medv + Boston$rm, data = Boston)
```



Looking at the different plots created, we can see that there seems to be a negative correlation between crime rate and median value, so as the median value increases the crime rate decreases as well. When comparing crime rate and number of rooms there doesn't seem to be a strong correlation between them. WHen comparing the median value to the number of rooms there is a strong positive correlation showing that as the number of rooms increases the median value also increases.

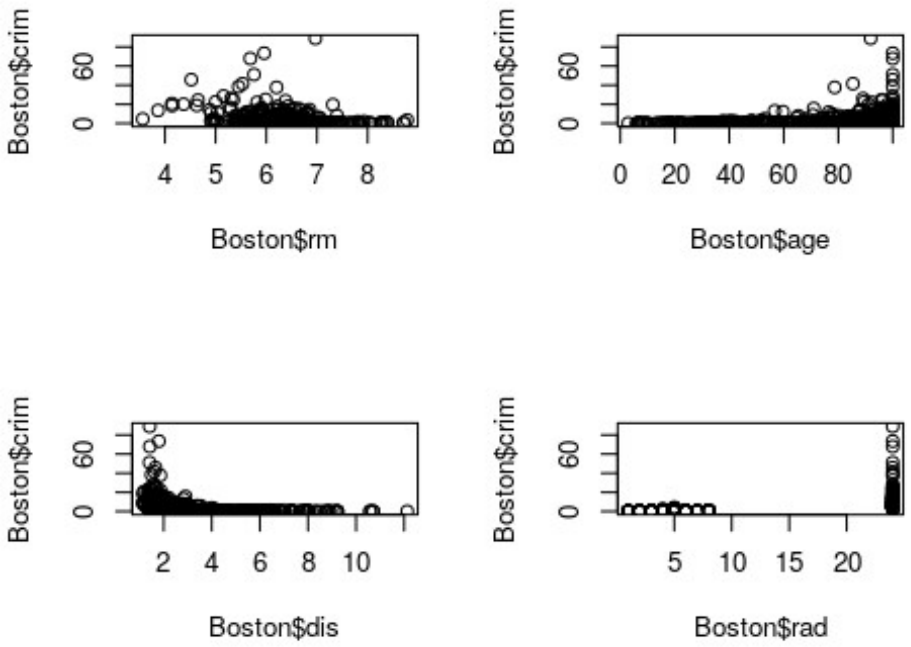
c)

```
par(mfrow = c(2,2))
plot(Boston$zn, Boston$crim)
plot(Boston$indus, Boston$crim)
plot(Boston$chas, Boston$crim)
plot(Boston$nox, Boston$crim)
```



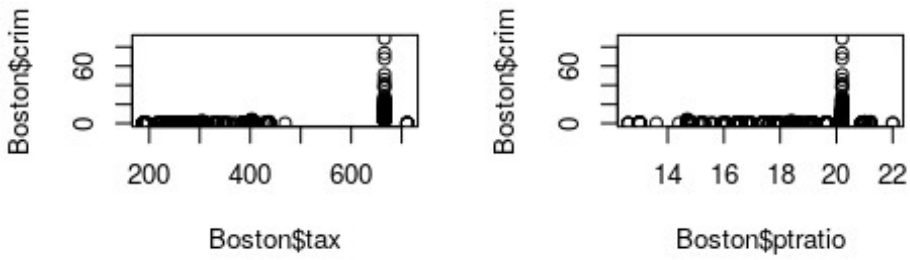
When testing zn, indus, chas, and nox none of these four predictors do not seem to be strongly associated with per capita crime rate.

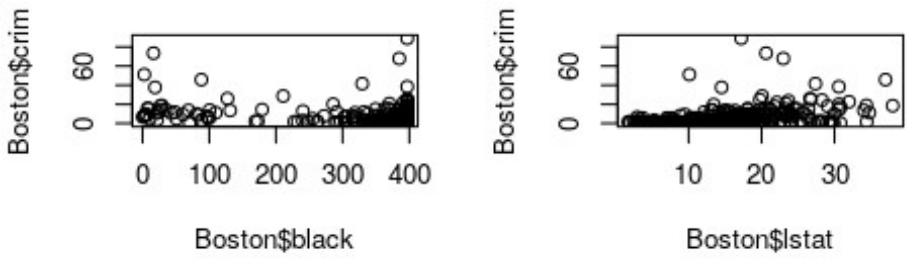
```
par(mfrow = c(2,2))
plot(Boston$rm, Boston$crim)
plot(Boston$age, Boston$crim)
plot(Boston$dis, Boston$crim)
plot(Boston$rad, Boston$crim)
```



When testing rm, age, dis, and rad there seems to be a stronger association for age and dis when comparing to crime rate. For age and crime, as age increases there seems to be a higher crime rate.For distance and crime rate as distance increases there seems to be a lower crime rate.

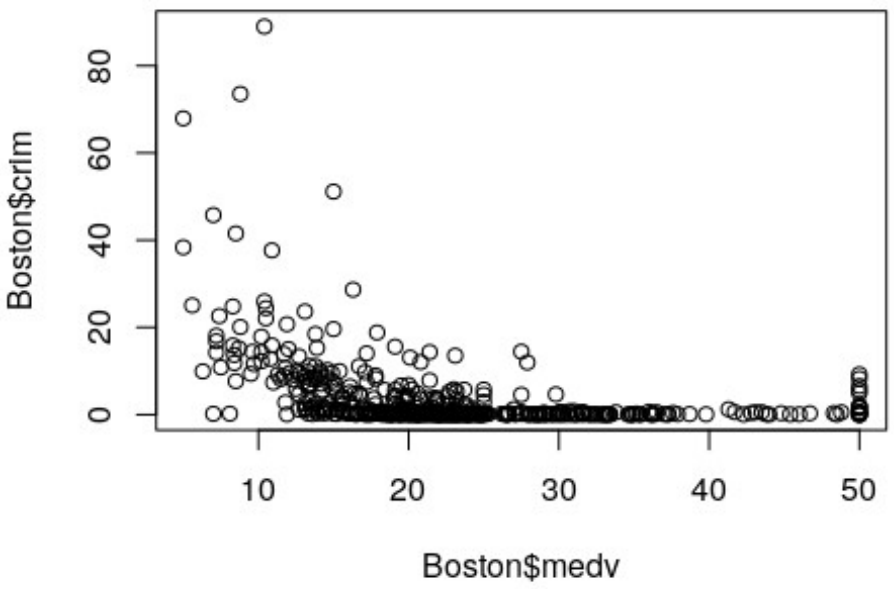
```
par(mfrow = c(2,2))
plot(Boston$tax, Boston$crim)
plot(Boston$ptratio, Boston$crim)
plot(Boston$black, Boston$crim)
plot(Boston$lstat, Boston$crim)
```





When testing tax, ptratio, black, and lstat there seems to be an association with lstat and crime when comparing to crime rate. For percent lower status and crime rate, as lower stats percent increases the crime rate also increases.

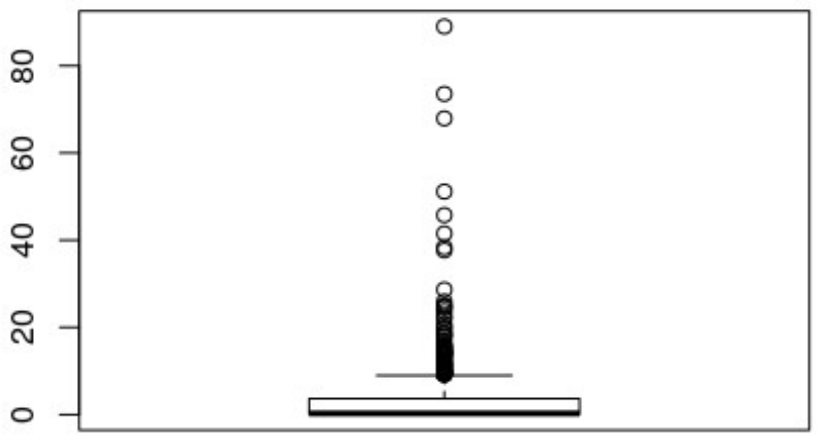
```
plot(Boston$medv, Boston$crim)
```



When testing crime rate and medv there seems to be a negative association. As median value increases the crime rate seems to decrease.

d)

```
par(mfrow = c(1,1))
boxplot(Boston$crim)
```




```
range(Boston$crim)
```

```
## [1]  0.00632 88.97620
```

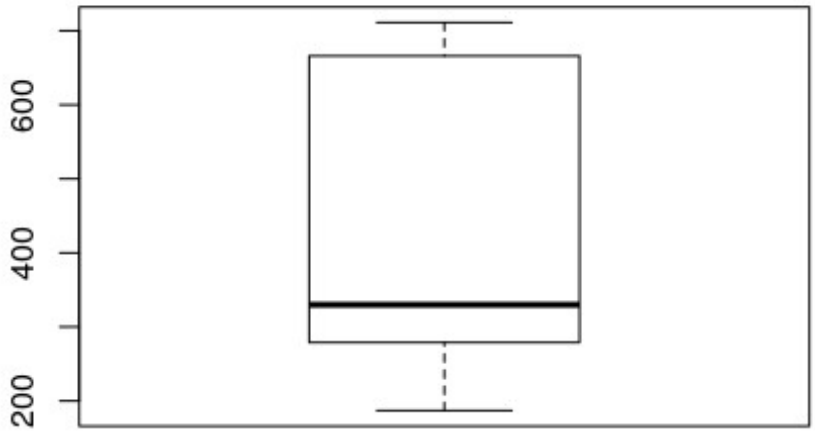
```
crimrange <- max(Boston$crim) - min(Boston$crim)
crimrange
```

```
## [1] 88.96988
```

According to the box plot there are suburbs in Boston that have high crime rates. The max value for crime rate is 88.9762 and the min value is 0.00632.

The range for crime rate in the suburbs is about 88.96988.

```
boxplot(Boston$tax)
```




```
range(Boston$tax)
```

```
## [1] 187 711
```

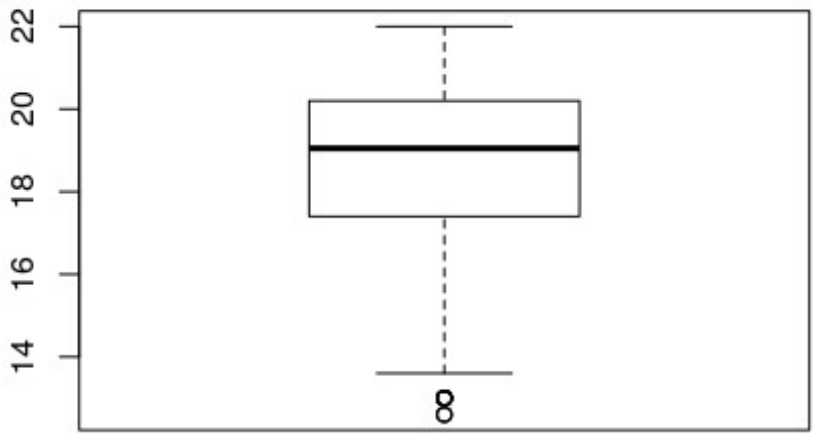
```
taxrange <- max(Boston$tax) - min(Boston$tax)
taxrange
```

```
## [1] 524
```

According to the box plot there does not seem to be any suburbs that have a particularly high tax rate as all of them fall within the quartile ranges. We also did not see any outliers.

The max value of the tax rate is 711 and the min value is 187. The range of the tax rate is 524.

```
boxplot(Boston$ptratio)
```




```
range(Boston$ptratio)
```

```
## [1] 12.6 22.0
```

```
ptratiorange <- max(Boston$ptratio) - min(Boston$ptratio)
ptratiorange
```

```
## [1] 9.4
```

According to the box plot there does not seem to be any suburbs with a high pupil-teacher ratio, but there are some with a very low pupil-teacher ratio.

The max value for the pupil-teacher ratio is 22 and the min value is 12.6. The range of the pupil-teacher ratio is 9.4.

e)

```
Charlescount <- sum(Boston$chas == 1)
Charlescount
```

```
## [1] 35
```

There are 35 suburbs in this data set bound to the Charles River.

f)

```
median(Boston$ptratio)
```

```
## [1] 19.05
```

The median for the pupil-teacher ratio in this data set is 19.05.

Section 3.7 Question 8

a)

```
#Auto = read.table("Auto.data")
#fix(Auto)
head(Auto)
```

```
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1  18         8          307        130   3504         12.0   70      1
## 2  15         8          350        165   3693         11.5   70      1
## 3  18         8          318        150   3436         11.0   70      1
## 4  16         8          304        150   3433         12.0   70      1
## 5  17         8          302        140   3449         10.5   70      1
## 6  15         8          429        198   4341         10.0   70      1
##                       name
## 1 chevrolet chevelle malibu
## 2         buick skylark 320
## 3        plymouth satellite
## 4             amc rebel sst
## 5               ford torino
## 6          ford galaxie 500
```

```
lm8 <- lm(formula = mpg~horsepower, data = Auto)
summary(lm8)
```

```
##
## 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
```

i.

Yes there is a relationship between mpg and horsepower as the p-value for the f-stat is small.

ii.

The relationship between mpg and horsepower is moderately strong with a 0.6059 R-squared value.

iii.

The relationship between mpg and horsepower is negative. As horsepower increases mpg will decrease.

```
predicted_val <- lm8$coef[1] + lm8$coef[2] * 98
predicted_val
```

```
## (Intercept)
##    24.46708
```

```
predict(lm8, data.frame(horsepower = 98), interval = "confidence")
```

```
##        fit      lwr      upr
## 1 24.46708 23.97308 24.96108
```

```
predict(lm8, data.frame(horsepower = 98), interval = "prediction")
```

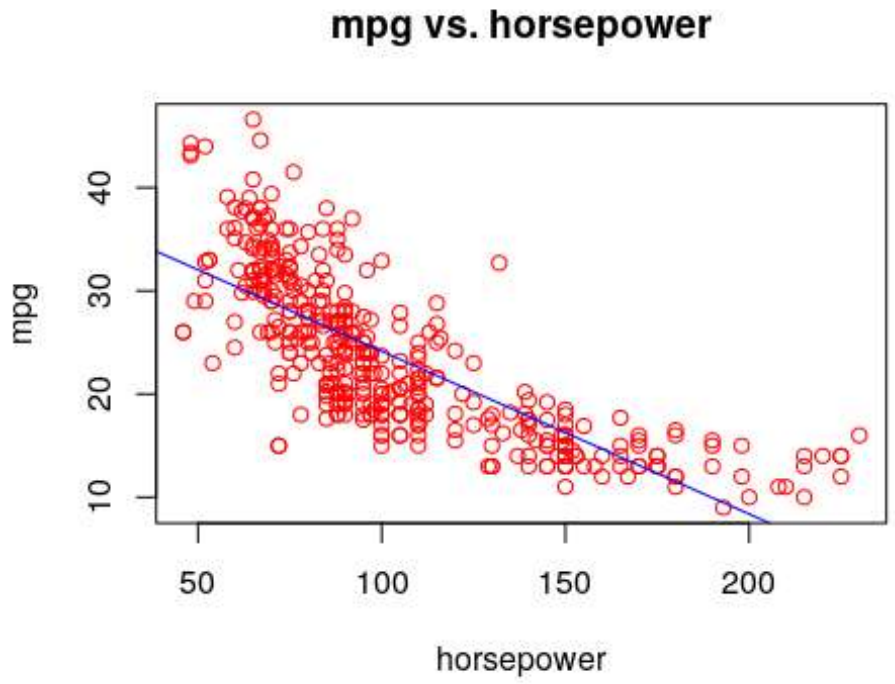
```
##        fit     lwr      upr
## 1 24.46708 14.8094 34.12476
```

iv.

The predicted mpg is 24.46708. The associated 95% confidence interval is from 23.97308 to 24.96108 and the associated prediction interval is from 14.8094 to 34.12476.

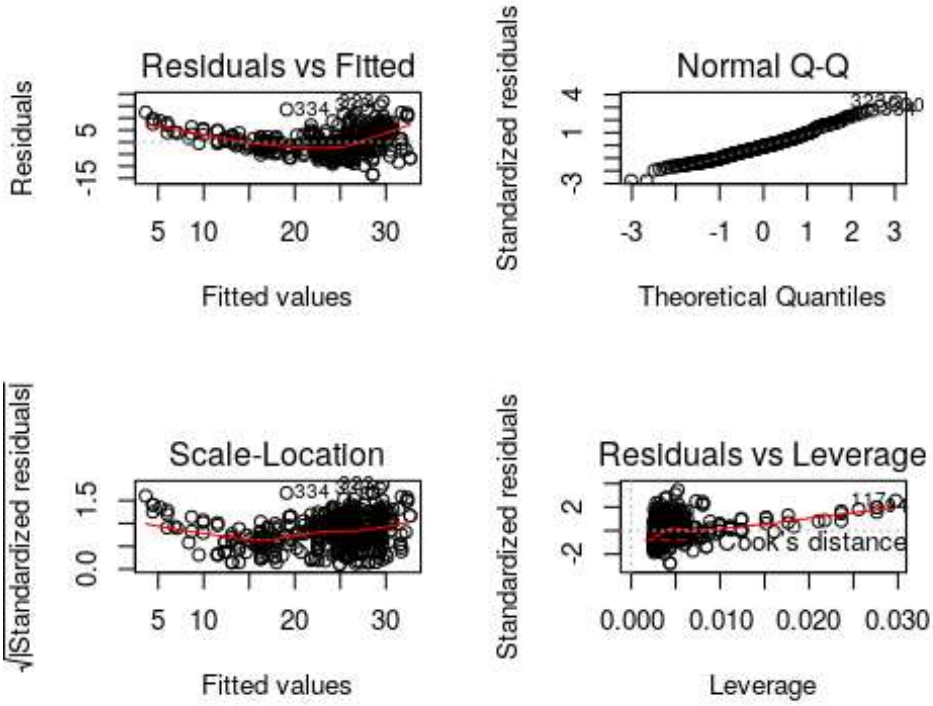
b)

```
plot(Auto$horsepower, Auto$mpg, main = "mpg vs. horsepower", xlab =
"horsepower", ylab = "mpg", col = "Red")
abline(lm8, col = "Blue")
```



c)

```
par(mfrow = c(2, 2))
plot(lm8)
```



A problem that we see is that the line in the residual plot is not linear rather it is curved. This can indicate that there is a non linear relationship within the model. Also, we see in the residuals vs leverage plot that there are some values that are above 2 and below -2 which can indicate potential outliers in the data.

Section 3.7 Question 9

a)

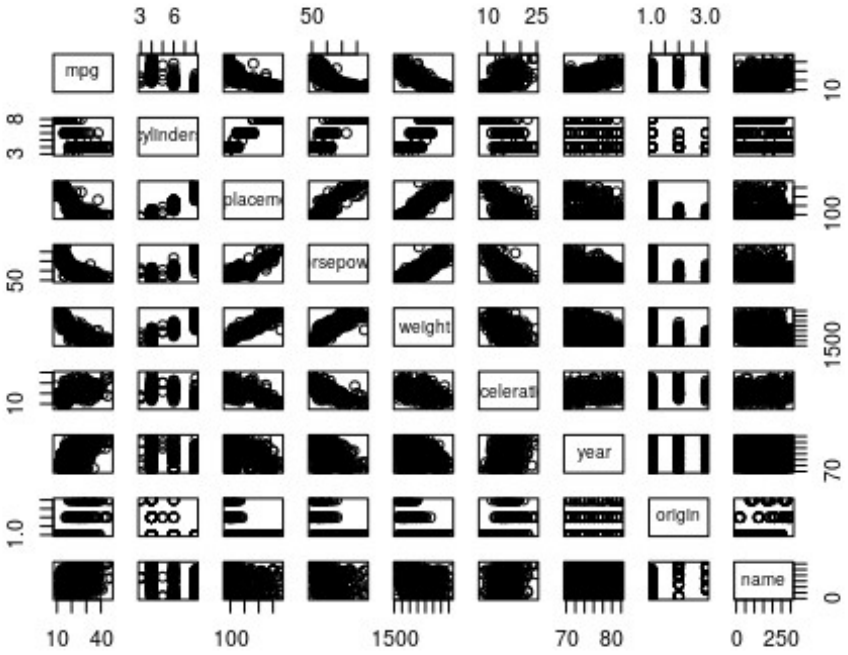
```
par(mfrow = c(1, 1))
names(Auto)
```

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

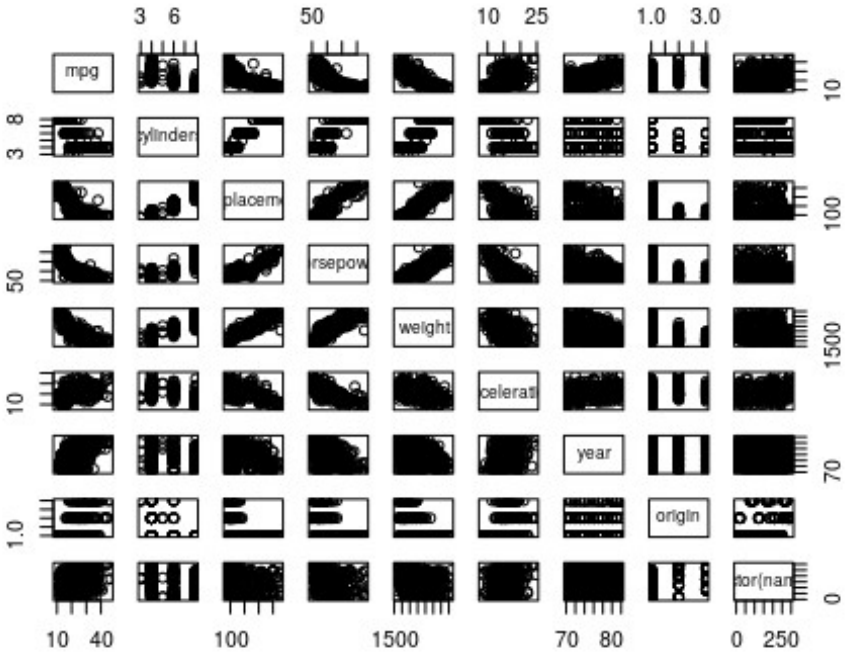
```
head(Auto)
```

```
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1  18         8          307        130   3504         12.0   70      1
## 2  15         8          350        165   3693         11.5   70      1
## 3  18         8          318        150   3436         11.0   70      1
## 4  16         8          304        150   3433         12.0   70      1
## 5  17         8          302        140   3449         10.5   70      1
## 6  15         8          429        198   4341         10.0   70      1
##                       name
## 1 chevrolet chevelle malibu
## 2         buick skylark 320
## 3        plymouth satellite
## 4             amc rebel sst
## 5               ford torino
## 6          ford galaxie 500
```

```
pairs(Auto)
```




```
pairs(~ mpg + cylinders + displacement + horsepower + weight + acceleration +
year + origin + factor(name), data = Auto)
```



b)

```r
cor(Auto[, names(Auto) != "name"])
```

```
##                     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
## origin        0.5652088 -0.5689316   -0.6145351 -0.4551715 -0.5850054
##              acceleration       year     origin
## mpg             0.4233285  0.5805410  0.5652088
## cylinders      -0.5046834 -0.3456474 -0.5689316
## displacement   -0.5438005 -0.3698552 -0.6145351
## horsepower     -0.6891955 -0.4163615 -0.4551715
## weight         -0.4168392 -0.3091199 -0.5850054
## acceleration    1.0000000  0.2903161  0.2127458
## year            0.2903161  1.0000000  0.1815277
## origin          0.2127458  0.1815277  1.0000000
```

c)

```r
lm9 <- lm(formula = mpg~cylinders + displacement + horsepower + weight +
acceleration + year + origin, data = Auto)
summary(lm9)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + origin, 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
```

i.

Yes, there is a relationship between the predictors and the response mpg given the f-stat has a very small p-value. The multiple R-squared is 0.8215 which is pretty high. This means that the 82% of the change is explained by the predictors.

ii.

The predictors that are significant are displacement, weight, year, and origin.

iii.

Whenever every other predictor is held constant, when the year increases by 1 the mpg also increased by 0.750773.

e)

```
lm9.1 <- lm(formula = mpg~cylinders + displacement + horsepower + weight +
acceleration + year + origin + displacement:weight, data = Auto)
summary(lm9.1)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + origin + displacement:weight, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.9027 -1.8092 -0.0946  1.5549 12.1687
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -5.389e+00  4.301e+00  -1.253   0.2109
## cylinders            1.175e-01  2.943e-01   0.399   0.6899
## displacement        -6.837e-02  1.104e-02  -6.193 1.52e-09 ***
## horsepower          -3.280e-02  1.238e-02  -2.649   0.0084 **
## weight              -1.064e-02  7.136e-04 -14.915  < 2e-16 ***
## acceleration         6.724e-02  8.805e-02   0.764   0.4455
## year                 7.852e-01  4.553e-02  17.246  < 2e-16 ***
## origin               5.610e-01  2.622e-01   2.139   0.0331 *
## displacement:weight  2.269e-05  2.257e-06  10.054  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.964 on 383 degrees of freedom
## Multiple R-squared:  0.8588, Adjusted R-squared:  0.8558
## F-statistic: 291.1 on 8 and 383 DF,  p-value: < 2.2e-16
```

```r
lm9.2 <- lm(formula = mpg~cylinders + displacement + horsepower + weight +
acceleration + year + origin + displacement:cylinders + displacement:weight +
year:origin + acceleration:horsepower, data = Auto)
summary(lm9.2)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + origin + displacement:cylinders +
displacement:weight +
##     year:origin + acceleration:horsepower, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.6504 -1.6476  0.0381  1.4254 12.7893
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)              5.287e+00  9.074e+00   0.583 0.560429
## cylinders                4.249e-01  6.079e-01   0.699 0.485011
## displacement            -7.322e-02  1.334e-02  -5.490 7.38e-08 ***
## horsepower               5.252e-02  2.586e-02   2.031 0.042913 *
## weight                  -8.689e-03  1.086e-03  -7.998 1.54e-14 ***
## acceleration             5.796e-01  1.582e-01   3.665 0.000283 ***
## year                     5.116e-01  9.976e-02   5.129 4.66e-07 ***
## origin                  -1.220e+01  4.161e+00  -2.933 0.003560 **
## cylinders:displacement  -4.368e-04  2.712e-03  -0.161 0.872156
## displacement:weight      1.992e-05  3.608e-06   5.522 6.21e-08 ***
## year:origin              1.630e-01  5.341e-02   3.051 0.002440 **
## horsepower:acceleration -6.735e-03  1.781e-03  -3.781 0.000181 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.874 on 380 degrees of freedom
## Multiple R-squared:  0.8683, Adjusted R-squared:  0.8644
## F-statistic: 227.7 on 11 and 380 DF,  p-value: < 2.2e-16
```

```r
lm9.3 <- lm(formula = mpg~cylinders + displacement + horsepower + weight +
acceleration + year + origin + displacement:cylinders + displacement*weight +
acceleration*horsepower, data = Auto)
summary(lm9.3)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + origin + displacement:cylinders + displacement *
##     weight + acceleration * horsepower, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -9.3344 -1.6333  0.0188  1.4740 11.9723
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.725e+01  5.328e+00  -3.237  0.00131 **
## cylinders                 6.354e-01  6.106e-01   1.041  0.29870
## displacement             -6.805e-02  1.337e-02  -5.088 5.68e-07 ***
## horsepower                6.026e-02  2.601e-02   2.317  0.02105 *
## weight                   -8.864e-03  1.097e-03  -8.084 8.43e-15 ***
## acceleration              6.257e-01  1.592e-01   3.931  0.00010 ***
## year                      7.845e-01  4.470e-02  17.549  < 2e-16 ***
## origin                    4.668e-01  2.595e-01   1.799  0.07284 .
## cylinders:displacement   -1.337e-03  2.726e-03  -0.490  0.62415
## displacement:weight       2.071e-05  3.638e-06   5.694 2.49e-08 ***
## horsepower:acceleration -7.467e-03  1.784e-03  -4.185 3.55e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.905 on 381 degrees of freedom
## Multiple R-squared:  0.865,  Adjusted R-squared:  0.8615
## F-statistic: 244.2 on 10 and 381 DF,  p-value: < 2.2e-16
```

```
lm9.4 <- lm(formula = mpg~cylinders + displacement + horsepower + weight +
acceleration + year + origin + weight*acceleration*horsepower, data = Auto)
summary(lm9.4)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + origin + weight * acceleration * horsepower,
##     data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8859 -1.5686 -0.0118  1.3821 11.9505
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -1.151e+01  1.058e+01  -1.088  0.27741
## cylinders                         1.650e-02  3.280e-01   0.050  0.95990
## displacement                     -4.407e-03  7.662e-03  -0.575  0.56553
## horsepower                        4.844e-02  1.035e-01   0.468  0.63998
## weight                           -9.218e-03  3.431e-03  -2.687  0.00753 **
## acceleration                      8.681e-01  6.018e-01   1.442  0.14999
## year                              7.647e-01  4.448e-02  17.191  < 2e-16 ***
## origin                            7.166e-01  2.489e-01   2.879  0.00422 **
## weight:acceleration              -4.540e-05  2.036e-04  -0.223  0.82369
## horsepower:weight                 9.235e-06  2.717e-05   0.340  0.73411
## horsepower:acceleration          -2.007e-02  7.234e-03  -2.774  0.00581 **
## horsepower:weight:acceleration    3.140e-06  1.883e-06   1.668  0.09620 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.878 on 380 degrees of freedom
## Multiple R-squared:  0.8679, Adjusted R-squared:  0.8641
## F-statistic: 226.9 on 11 and 380 DF,  p-value: < 2.2e-16
```

```
lm9.5 <- lm(formula = mpg~displacement + horsepower + weight + acceleration +
weight:acceleration + year + origin + year:origin + displacement:weight +
acceleration:horsepower, data = Auto)
summary(lm9.5)
```

```
##
## Call:
## lm(formula = mpg ~ displacement + horsepower + weight + acceleration +
##     weight:acceleration + year + origin + year:origin +
displacement:weight +
##     acceleration:horsepower, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3653 -1.6289  0.0849  1.4952 12.7142
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.582e+01  9.919e+00   1.595 0.111602
## displacement               -7.578e-02  1.015e-02  -7.468 5.62e-13 ***
## horsepower                  8.346e-02  3.269e-02   2.553 0.011063 *
## weight                     -1.251e-02  2.481e-03  -5.042 7.13e-07 ***
## acceleration                1.318e-01  2.819e-01   0.467 0.640518
## year                        4.951e-01  9.872e-02   5.015 8.15e-07 ***
## origin                     -1.252e+01  4.119e+00  -3.041 0.002521 **
## weight:acceleration         2.262e-04  1.324e-04   1.708 0.088398 .
## year:origin                 1.674e-01  5.288e-02   3.166 0.001673 **
## displacement:weight         2.174e-05  2.755e-06   7.892 3.19e-14 ***
## horsepower:acceleration -8.900e-03  2.294e-03  -3.879 0.000124 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.864 on 381 degrees of freedom
## Multiple R-squared:  0.8688, Adjusted R-squared:  0.8653
## F-statistic: 252.3 on 10 and 381 DF,  p-value: < 2.2e-16
```

```
lm9.6 <- lm(formula = mpg~displacement + horsepower + weight +
weight:acceleration + year + origin + year:origin + displacement:weight +
acceleration:horsepower, data = Auto)
summary(lm9.6)
```

```
##
## Call:
## lm(formula = mpg ~ displacement + horsepower + weight +
```

```
weight:acceleration +
##     year + origin + year:origin + displacement:weight +
acceleration:horsepower,
##     data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5074 -1.6324  0.0599  1.4577 12.7376
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.868e+01  7.796e+00   2.396 0.017051 *
## displacement               -7.794e-02  9.026e-03  -8.636  < 2e-16 ***
## horsepower                  8.719e-02  3.167e-02   2.753 0.006183 **
## weight                     -1.350e-02  1.287e-03 -10.490  < 2e-16 ***
## year                        4.911e-01  9.825e-02   4.998 8.83e-07 ***
## origin                     -1.262e+01  4.109e+00  -3.071 0.002288 **
## weight:acceleration         2.784e-04  7.087e-05   3.929 0.000101 ***
## year:origin                 1.686e-01  5.277e-02   3.195 0.001516 **
## displacement:weight         2.253e-05  2.184e-06  10.312  < 2e-16 ***
## horsepower:acceleration -9.164e-03  2.222e-03  -4.125 4.56e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.861 on 382 degrees of freedom
## Multiple R-squared:  0.8687, Adjusted R-squared:  0.8656
## F-statistic: 280.8 on 9 and 382 DF,  p-value: < 2.2e-16
```

For lm9.1, the R-squared value is 0.8588. This is pretty high, but it isn't a high as some of the other models we did. For lm9.2 the R-squared value was a little higher with 0.8683, but the predictors cylindersand cylinders:displacement was found to not be statistically significant.

For lm9.3, the R-squared value was slightly smaller at 0.865. This model had a couple predictors that were not significant as well such as cylinders and cylinders:displacement.

For lm9.4, the R-squared value was somewhat in the middle with 0.8679. This model has a lot of non-significant predictors and we would not recommend using this model.

For lm9.5, this model got us the highest R-squared value with 0.8688. We removed cylinders from this model since it was non-significant in previous models. We also found that acceleration seemed to be a non-significant predictors in most of them, therefore we tried the last model lm9.6 without both cylinders and acceleration. The R-squared value is slightly less in 0.8687, but every single predictor was found to be significant. Therefore, it seems that this model is the best out of them and that weight:acceleration, year:origin, displacement:weight, and horsepower:acceleration seem to be significant interactions in the models.