Math642_HW1_Fyona

Fyona Sun

1/18/2020

Question 2.8

(a) Use the read.csv() function to read the data into R. Call the loaded data college. Make sure that you have the directory set to the correct location for the data.

```
college<- read.csv(file = '~/Math642_FyonaSun/College.csv',header = TRUE, sep =',')
summary(college)
```

```
##                               X          Private        Apps
##  Abilene Christian University:  1    No :212    Min.   :   81
##  Adelphi University          :  1    Yes:565    1st Qu.:  776
##  Adrian College              :  1               Median : 1558
##  Agnes Scott College         :  1               Mean   : 3002
##  Alaska Pacific University   :  1               3rd Qu.: 3624
##  Albertson College           :  1               Max.   :48094
##  (Other)                     :771
##      Accept          Enroll        Top10perc       Top25perc
##  Min.   :   72   Min.   :  35   Min.   : 1.00   Min.   :  9.0
##  1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00   1st Qu.: 41.0
##  Median : 1110   Median : 434   Median :23.00   Median : 54.0
##  Mean   : 2019   Mean   : 780   Mean   :27.56   Mean   : 55.8
##  3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.00   3rd Qu.: 69.0
##  Max.   :26330   Max.   :6392   Max.   :96.00   Max.   :100.0
##
##   F.Undergrad     P.Undergrad         Outstate       Room.Board
##  Min.   :  139   Min.   :    1.0   Min.   : 2340   Min.   :1780
##  1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320   1st Qu.:3597
##  Median : 1707   Median :  353.0   Median : 9990   Median :4200
##  Mean   : 3700   Mean   :  855.3   Mean   :10441   Mean   :4358
##  3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925   3rd Qu.:5050
##  Max.   :31643   Max.   :21836.0   Max.   :21700   Max.   :8124
##
##      Books           Personal         PhD            Terminal
##  Min.   :  96.0   Min.   : 250   Min.   :  8.00   Min.   : 24.0
##  1st Qu.: 470.0   1st Qu.: 850   1st Qu.: 62.00   1st Qu.: 71.0
##  Median : 500.0   Median :1200   Median : 75.00   Median : 82.0
##  Mean   : 549.4   Mean   :1341   Mean   : 72.66   Mean   : 79.7
##  3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.: 85.00   3rd Qu.: 92.0
##  Max.   :2340.0   Max.   :6800   Max.   :103.00   Max.   :100.0
##
##    S.F.Ratio      perc.alumni        Expend         Grad.Rate
##  Min.   : 2.50   Min.   : 0.00   Min.   : 3186   Min.   : 10.00
##  1st Qu.:11.50   1st Qu.:13.00   1st Qu.: 6751   1st Qu.: 53.00
##  Median :13.60   Median :21.00   Median : 8377   Median : 65.00
##  Mean   :14.09   Mean   :22.74   Mean   : 9660   Mean   : 65.46
##  3rd Qu.:16.50   3rd Qu.:31.00   3rd Qu.:10830   3rd Qu.: 78.00
##  Max.   :39.80   Max.   :64.00   Max.   :56233   Max.   :118.00
```

```
##
```

(b) Look at the data using the fix() function. You should notice that the first column is just the name of each university. We don't really want R to treat this as data. However, it may be handy to have these names for later. Try the following commands:

```r
rownames(college)<- college[,1]
college<- college[,-1]
```

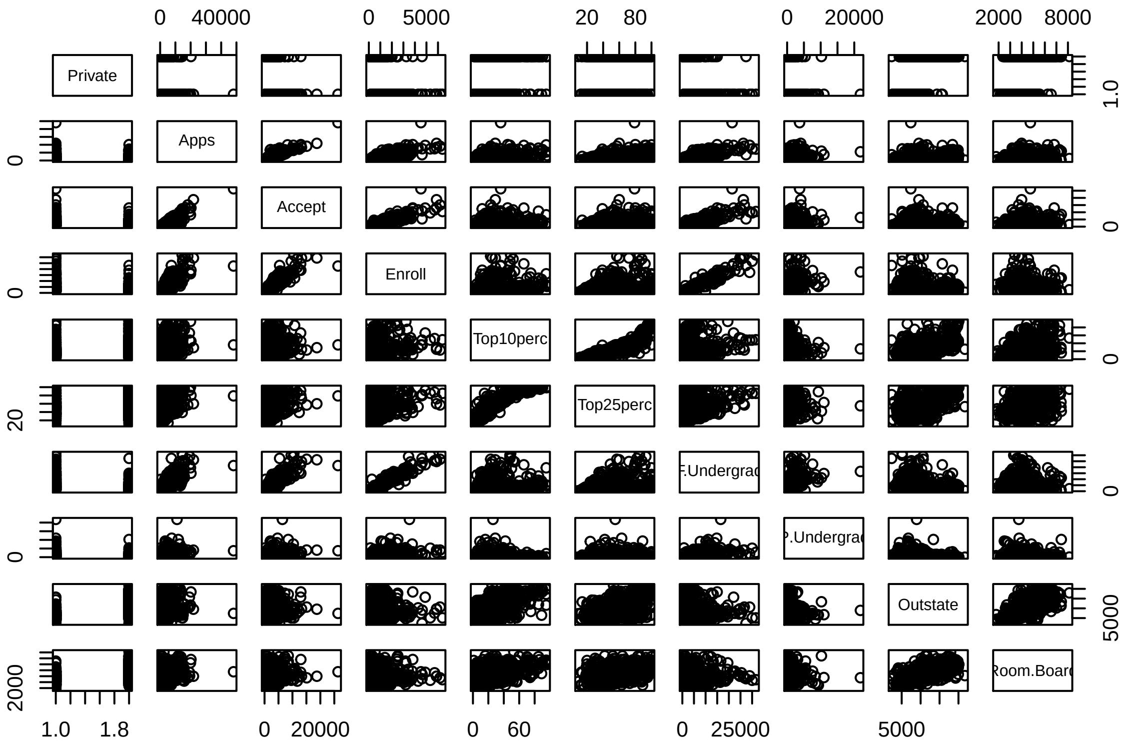
i. Use the summary() function to produce a numerical summary of the variables in the data set.
ii. Use the pairs() function to produce a scatterplot matrix of the first ten columns or variables of the data. Recall that you can reference the first ten columns of a matrix A using A[,1:10].
iii. Use the plot() function to produce side-by-side boxplots of Outstate versus Private.
iv. Create a new qualitative variable, called Elite, by binning the Top10perc variable. We are going to divide universities into two groups based on whether or not the proportion of students coming from the top 10% of their high school classes exceeds 50 %. Use the summary() function to see how many elite univer- sities there are. Now use the plot() function to produce side-by-side boxplots of Outstate versus Elite.
v. Use the hist() function to produce some histograms with differing numbers of bins for a few of the quantitative variables. You may find the command par(mfrow=c(2,2)) useful: it will divide the print window into four regions so that four plots can be made simultaneously. Modifying the arguments to this function will divide the screen in other ways.
vi. Continue exploring the data, and provide a brief summary of what you discover.

```r
#Summary of the data
summary(college)
```

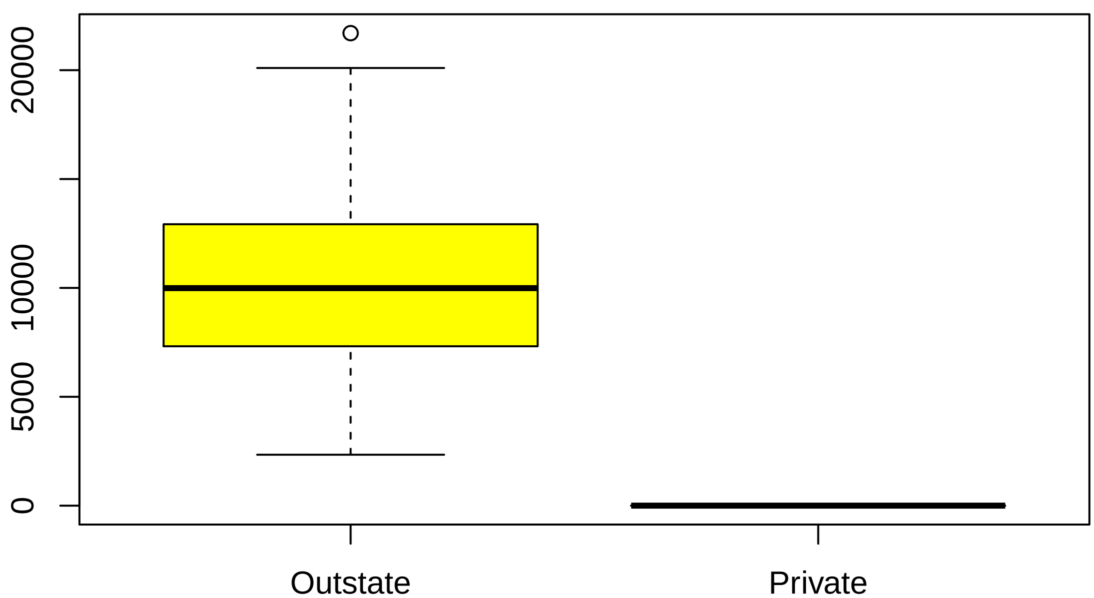
```
##  Private        Apps           Accept          Enroll       Top10perc
##  No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00
##  Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00
##            Median : 1558   Median : 1110   Median : 434   Median :23.00
##            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
```

```r
#Scatterplot matrix of the variables
pairs(college[,1:10])
```



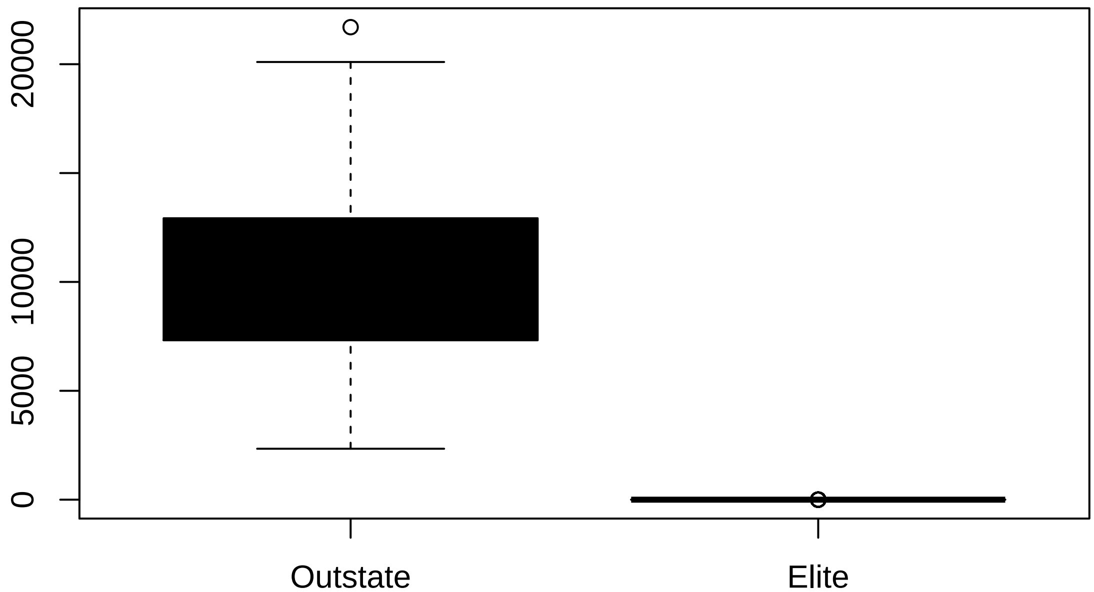

```r
#Boxplot of Outstate versus Private
boxplot(college$Outstate, college$Private,names=c('Outstate','Private'),col=23:24)
```



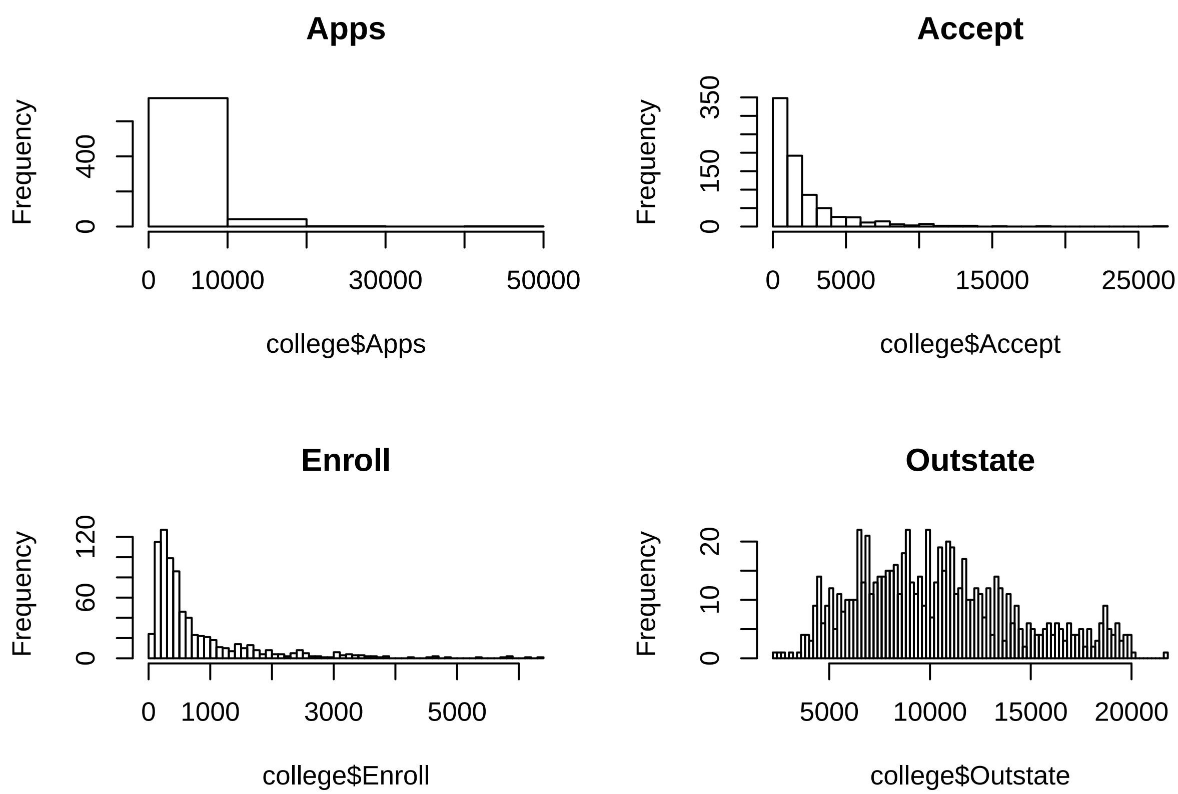
```r
#Create a new qualitative variable
Elite<- rep("No", nrow(college))
Elite[college$Top10perc>50]='Yes'
Elite<- as.factor(Elite)
college<- data.frame(college,Elite)
summary(college)
```

```
##  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
```

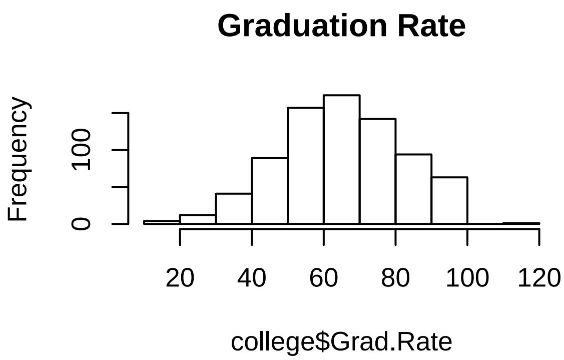
```r
#Boxplot of Outstate versus Elite
boxplot(college$Outstate, college$Elite,names=c('Outstate','Elite'),col=25:26)
```




```r
#Histograms with differing numbers of bins
par(mfrow=c(2,2))
hist(college$Apps,breaks = 4,main='Apps')
hist(college$Accept,breaks=25,main='Accept')
hist(college$Enroll,breaks=50,main='Enroll')
hist(college$Outstate, breaks = 100,main='Outstate')
```




```r
hist(college$Grad.Rate,main='Graduation Rate')
```



Exploring more information from the data 1. What schools have the most students in the top 10 percent of the class? What schools have the most students in the top 25 percent of the class?

```
#schools have the most top10 percent students
row.names(head(college[order(college$Top10perc,decreasing = TRUE),],10))
```

```
##  [1] "Massachusetts Institute of Technology"
##  [2] "Harvey Mudd College"
##  [3] "University of California at Berkeley"
##  [4] "Yale University"
##  [5] "Duke University"
##  [6] "Harvard University"
##  [7] "Princeton University"
##  [8] "Georgia Institute of Technology"
##  [9] "Brown University"
## [10] "Dartmouth College"
```

```
#schools have the most top25 percent students
row.names(head(college[order(college$Top25perc,decreasing = TRUE),],10))
```

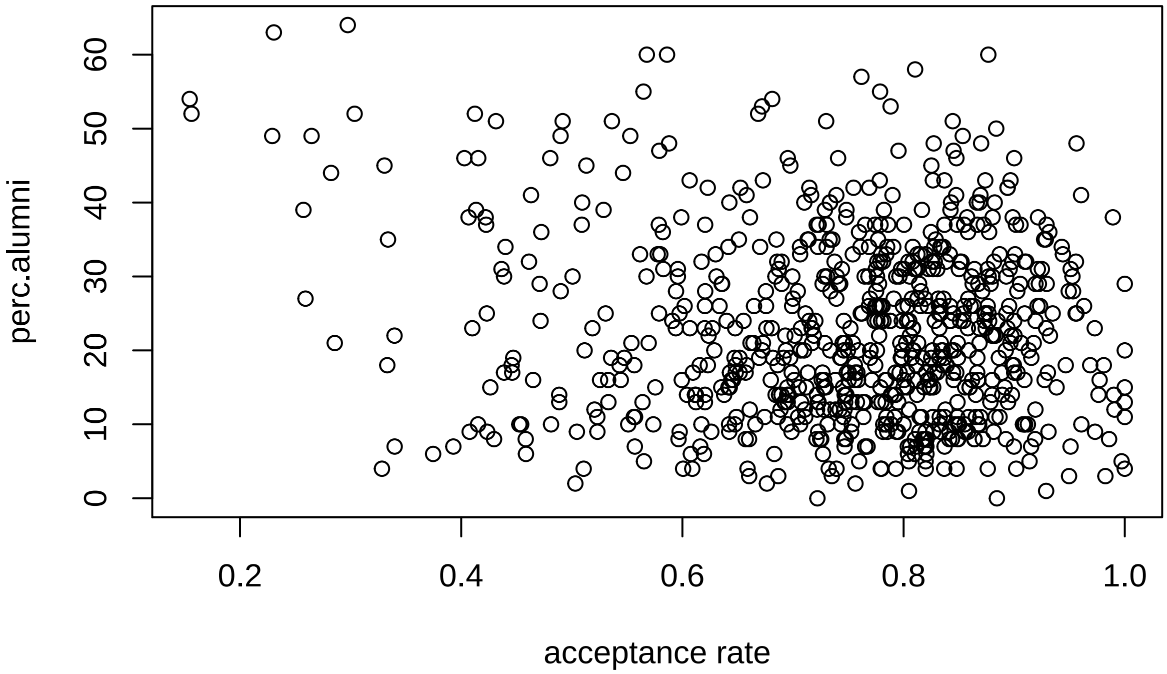
```
##  [1] "Bowdoin College"
##  [2] "Harvard University"
##  [3] "Harvey Mudd College"
##  [4] "SUNY at Buffalo"
##  [5] "University of California at Berkeley"
##  [6] "University of California at Irvine"
##  [7] "University of Pennsylvania"
##  [8] "Dartmouth College"
##  [9] "Georgia Institute of Technology"
## [10] "Massachusetts Institute of Technology"
```

```
school_with_top25<- head(college[order(college$Top25perc,decreasing = TRUE),],10)[,3:4]
enrollment_rate<- school_with_top25$Enroll / school_with_top25$Accept
school_with_top25<- data.frame(school_with_top25,enrollment_rate)
```

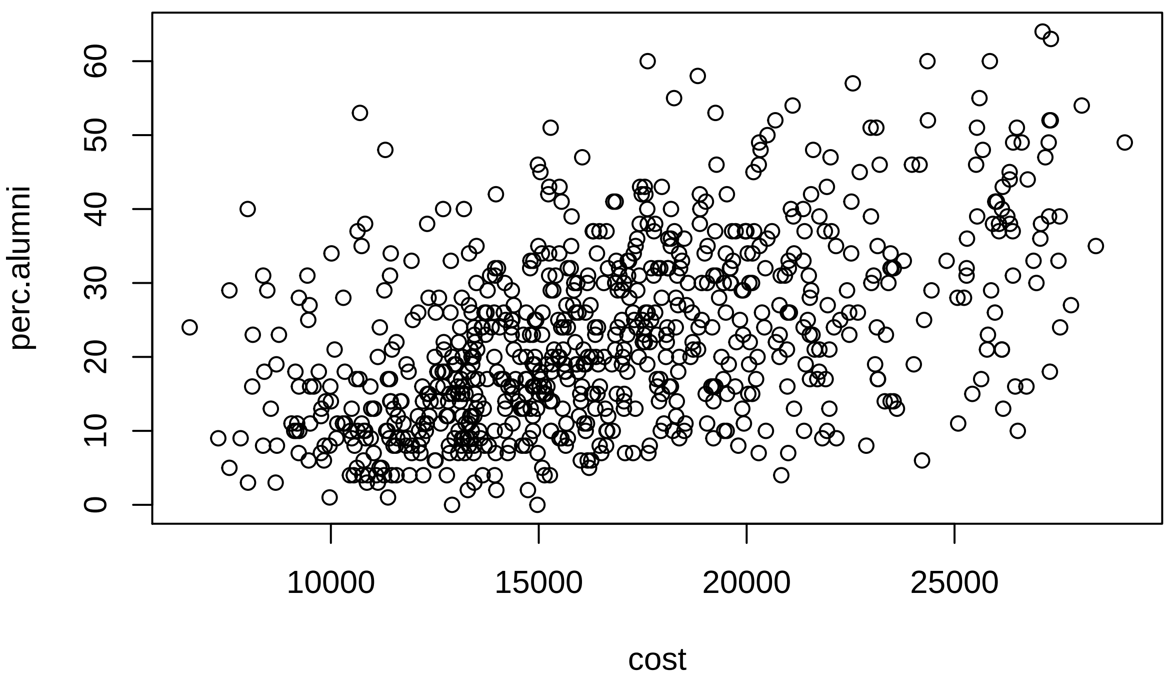
Comparing the top 10 school list, we can see that Yale, Duke, Princeton, and Brown University have high percentage of studetns from top 10% of high school class but not as many top 25% students. This might result from the school size or school cost. On the other hand, Bowdoin College, SUNY at Buffalo, UC Irvine has high percentage of top 25% student but not as many top 10% students which indicate that these are not top choices for top 10 students and it's possible that they are safe schools for some students. From the enrollment rate we can also see that among the 10 schools, University of California at Irvine and SUNY at Buffalo has a relatively lower enrollment rate.

2. What's the key factor for alumini donation?

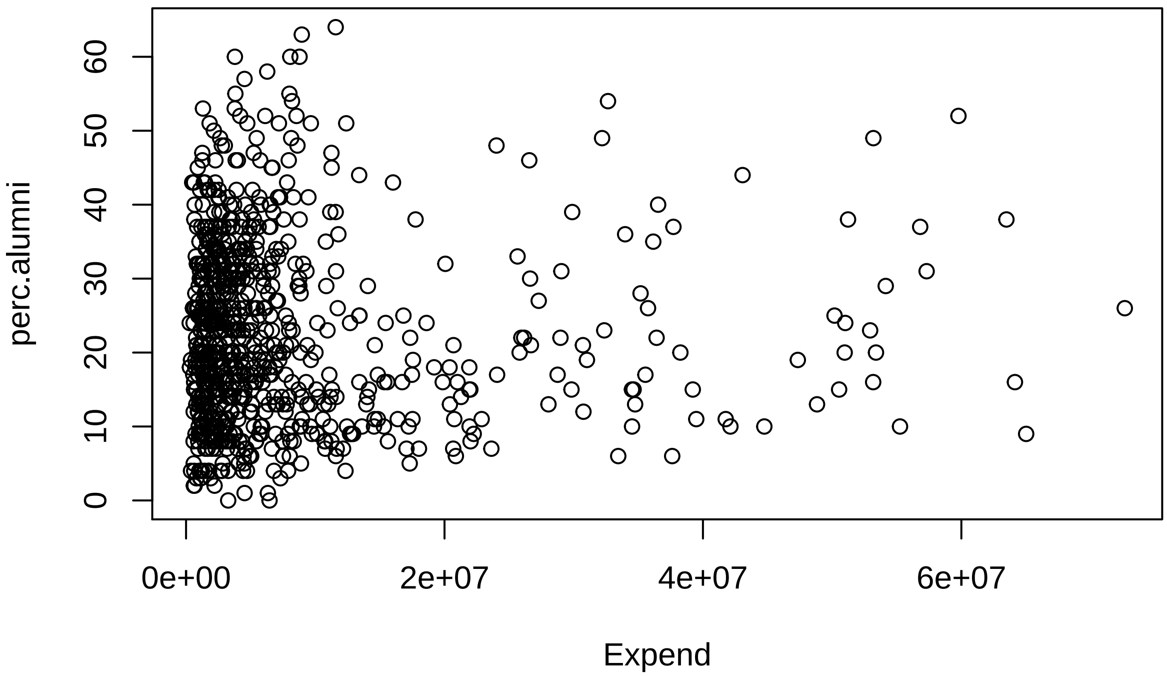
```
#the accptance rate
plot(perc.alumni ~ I(Accept/Apps), data = college, xlab='acceptance rate',ylab='perc.alumni')
```




```
#the cost of attending school
plot(perc.alumni ~ I(Personal+Books+Room.Board+Outstate), data = college, xlab='cost',ylab='perc.alumni
```




```
#the estimated overall school expend
plot(perc.alumni ~ I(Expend * Enroll), data = college, xlab='Expend',ylab='perc.alumni')
```



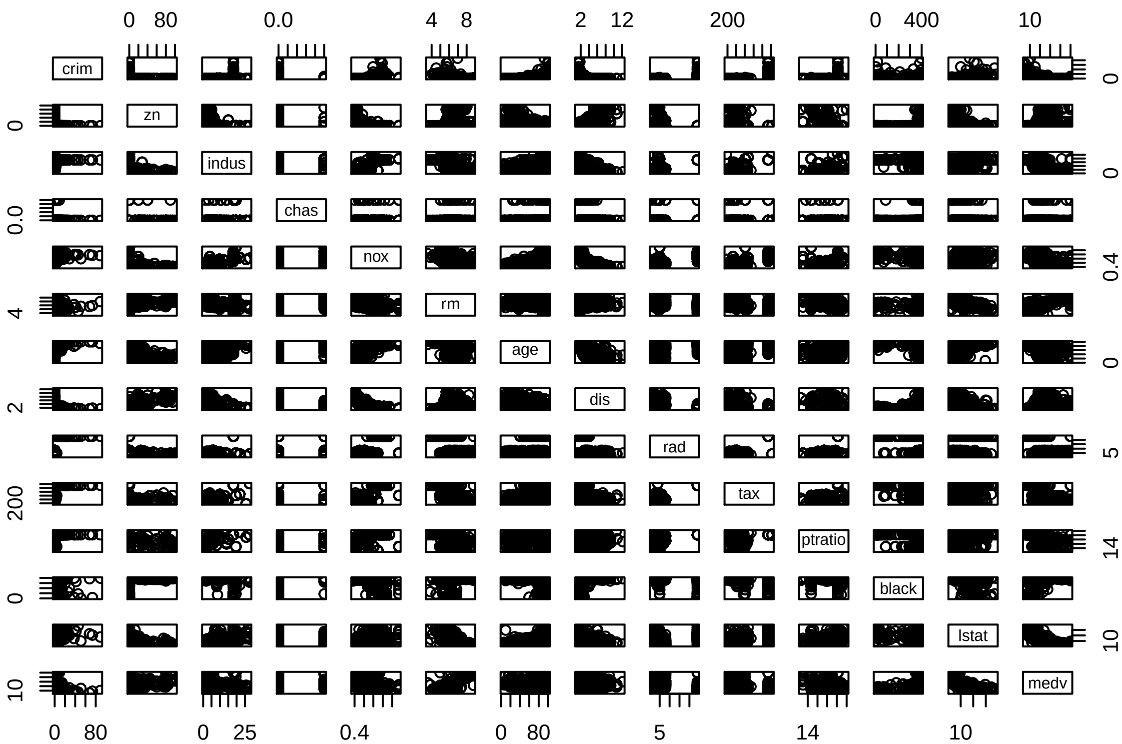
From the analysis, we found that some schools with lower acceptance rates tend to have higher donation rate, but the correlation between acceptance rate and percentage of alumini who donate is not significant. We cannot conclude that the lower accpetance rate leads to higher possible of donation. However, there is a possitive relationship between the student cost during college and the percentage of alumini donation. Students who spend more on college education are more likely to donate. Finally, there is no evidence to show that schools who spend more overall can get high percetange of donation. ## Question 2.10 (a) How many rows are in this data set? How many columns? What do the rows and columns represent?

```
library(MASS)
data(Boston)
# dimension of the dataset. 506 rows and 14 colums
dim(Boston)
```

```
## [1] 506  14
```

Each row represent an observation of a neiborhood. Each column represent a feature regarding that neiborhood. (b) Make some pairwise scatterplots of the predictors (columns) in this data set. Describe your findings.

```
pairs(Boston)
```

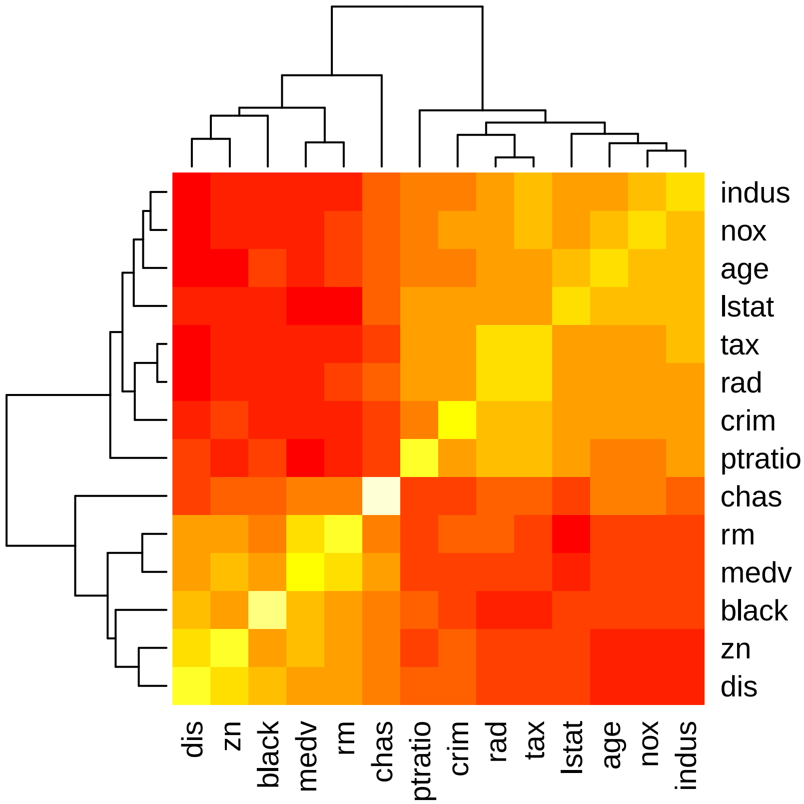


(c) Are any of the predictors associated with per capita crime rate? If so, explain the relationship.

```
Boston.corr = cor(Boston)
Boston.corr.crim = Boston.corr[-1,1]
print(
  Boston.corr.crim[order(abs(Boston.corr.crim), decreasing = T)]
)
```

```
##         rad         tax       lstat         nox       indus        medv
##  0.62550515  0.58276431  0.45562148  0.42097171  0.40658341 -0.38830461
##       black         dis         age     ptratio          rm          zn
## -0.38506394 -0.37967009  0.35273425  0.28994558 -0.21924670 -0.20046922
##        chas
## -0.05589158
```

```
#visulization
heatmap(cor(Boston, use="pairwise.complete.obs"))
```



From the correlation coefficients and the heatmap, the most importance factors are rad, tax, lstat, nox and indus.These 5 factors has positive relationship with the per capita crime rate.The medv, black, and dis has negative relationship with the crime rate, but the correlation is weak.

(d) Do any of the suburbs of Boston appear to have particularly high crime rates? Tax rates? Pupil-teacher ratios? Comment on the range of each predictor.

```
par(mfrow=c(2,2))
#crime rate
summary(Boston$crim)
```

```
##     Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
##  0.00632  0.08204  0.25651  3.61352  3.67708 88.97620
```

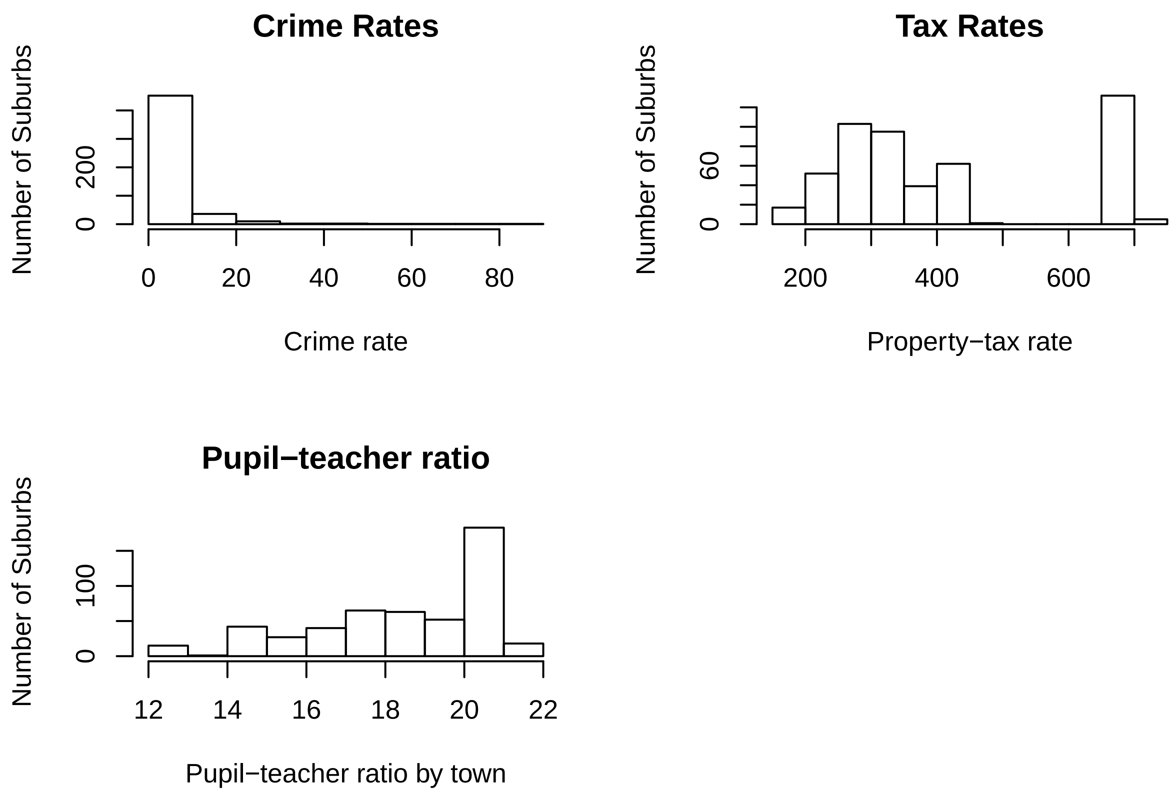
```
hist(Boston$crim, xlab = "Crime rate", ylab="Number of Suburbs",main="Crime Rates")
#tax rate
summary(Boston$tax)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   187.0   279.0   330.0   408.2   666.0   711.0
```

```
hist(Boston$tax, xlab = "Property-tax rate", ylab="Number of Suburbs",main="Tax Rates")
#pt ratio
summary(Boston$ptratio)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.60   17.40   19.05   18.46   20.20   22.00
```

```r
hist(Boston$ptratio, xlab ="Pupil-teacher ratio by town", ylab="Number of Suburbs",main="Pupil-teacher
```



The range of the crime rate is from 0.006% to 88.98% with a mean of 3.61% and a median of 0.26%. Thus there are 10.6% of the suburbs have crime rate above 10 and there are 0.79% of the suburbs have crime rate above 50. The histogram of the Tax rates shows that there are fewer neiborhoods has tax rates between 450-650, around 0.2% of the total number of the towns. Above 30% of the towns have the pupil-teacher ratio ratio between 20-22.

```r
#crime rate
above_10 <- subset(Boston, crim > 10)
nrow(above_10)/ nrow(Boston)
```

```
## [1] 0.1067194
```

```r
above_50<- subset(Boston, crim > 50)
nrow(above_50)/ nrow(Boston)
```

```
## [1] 0.007905138
```

```r
#tax rate
tax_rate_between<- subset(Boston, tax > 450 & tax < 650)
nrow(tax_rate_between)/ nrow(Boston)
```

```
## [1] 0.001976285
```

```r
#ptratio
ptratio_between<- subset(Boston, ptratio >=20)
nrow(ptratio_between)/ nrow(Boston)
```

```
## [1] 0.3972332
```

(e) How many of the suburbs in this data set bound the Charles river?

```r
summary(Boston$chas==1)
```

```
##    Mode   FALSE    TRUE
## logical     471      35
```

There are 35 suburbs bound the Charles river (f) What is the median pupil-teacher ratio among the towns in this data set?

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

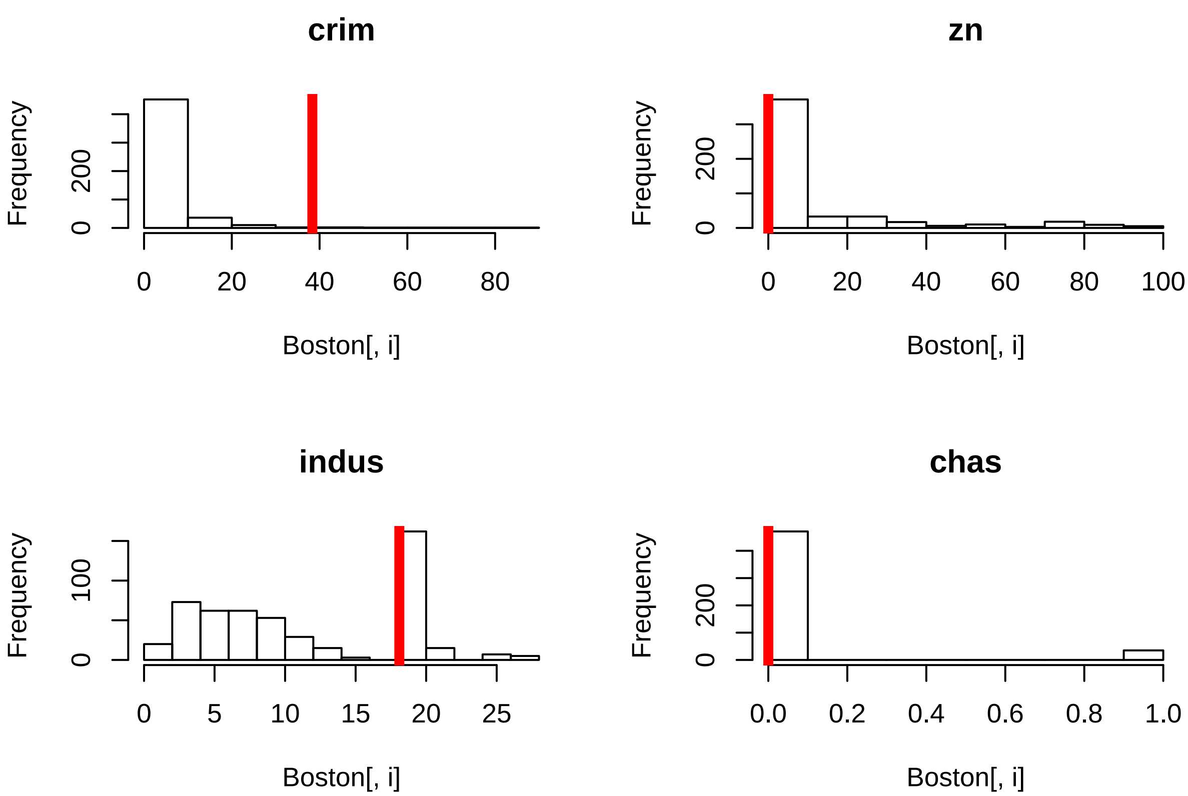
```
## [1] 19.05
```

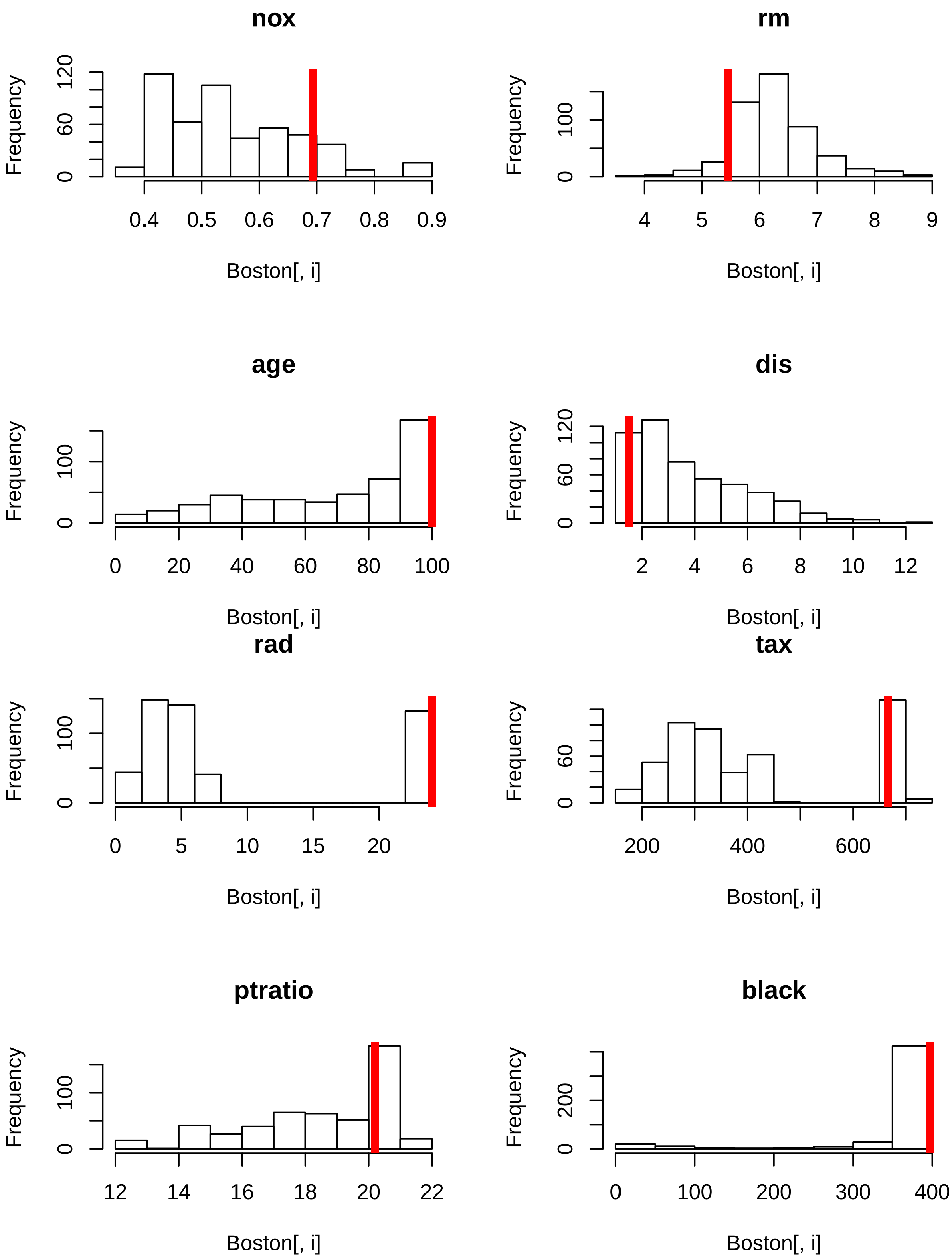
(g) Which suburb of Boston has lowest median value of owner- occupied homes? What are the values of the other predictors for that suburb, and how do those values compare to the overall ranges for those predictors? Comment on your findings.

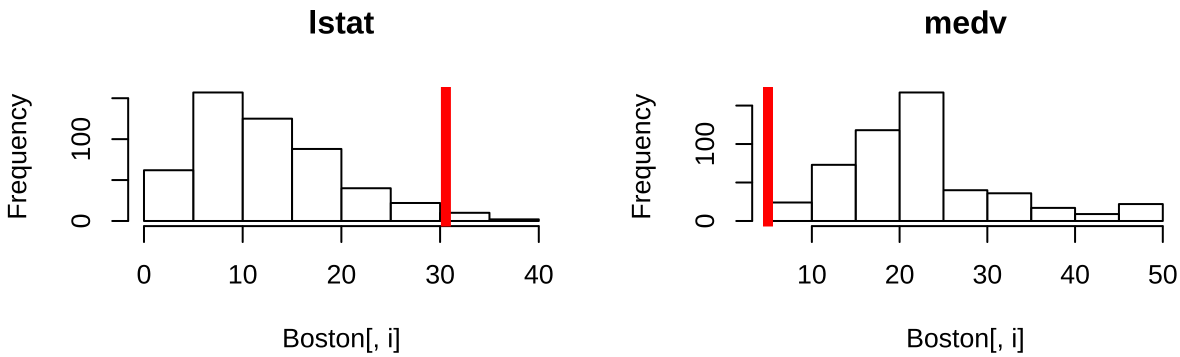
```r
Boston[which.min(Boston$medv),]
```

```
##        crim zn indus chas   nox    rm age    dis rad tax ptratio black
## 399 38.3518  0  18.1    0 0.693 5.453 100 1.4896  24 666    20.2 396.9
##     lstat medv
## 399 30.59    5
```

```r
par(mfrow=c(2,2))
for (i in 1:ncol(Boston)){
  hist(Boston[, i], main=colnames(Boston)[i])
  abline(v=Boston[399, i], col="red", lw=5)
}
```




nox
Frequency
0
60
120
0.4
0.5
0.6
0.7
0.8
0.9
Boston[, i]
rm
Frequency
0
100
4
5
6
7
8
9
Boston[, i]
age
Frequency
0
100
0
20
40
60
80
100
Boston[, i]
dis
Frequency
0
60
120
2
4
6
8
10
12
Boston[, i]
rad
Frequency
0
100
0
5
10
15
20
Boston[, i]
tax
Frequency
0
60
200
400
600
Boston[, i]
ptratio
Frequency
0
100
12
14
16
18
20
22
Boston[, i]
black
Frequency
0
200
0
100
200
300
400
Boston[, i]



The suburb of Boston that has lowest median value of owner-occupied homes is far away from Charles River, trafic(becuase of the low level in nitrogen oxides concentration and the distance from highway), and school (because of the low level of pupil-teacher ratio by town). The town has more people of lower status of the population and more aged population.

(h) In this data set, how many of the suburbs average more than seven rooms per dwelling? More than eight rooms per dwelling? Comment on the suburbs that average more than eight rooms per dwelling.

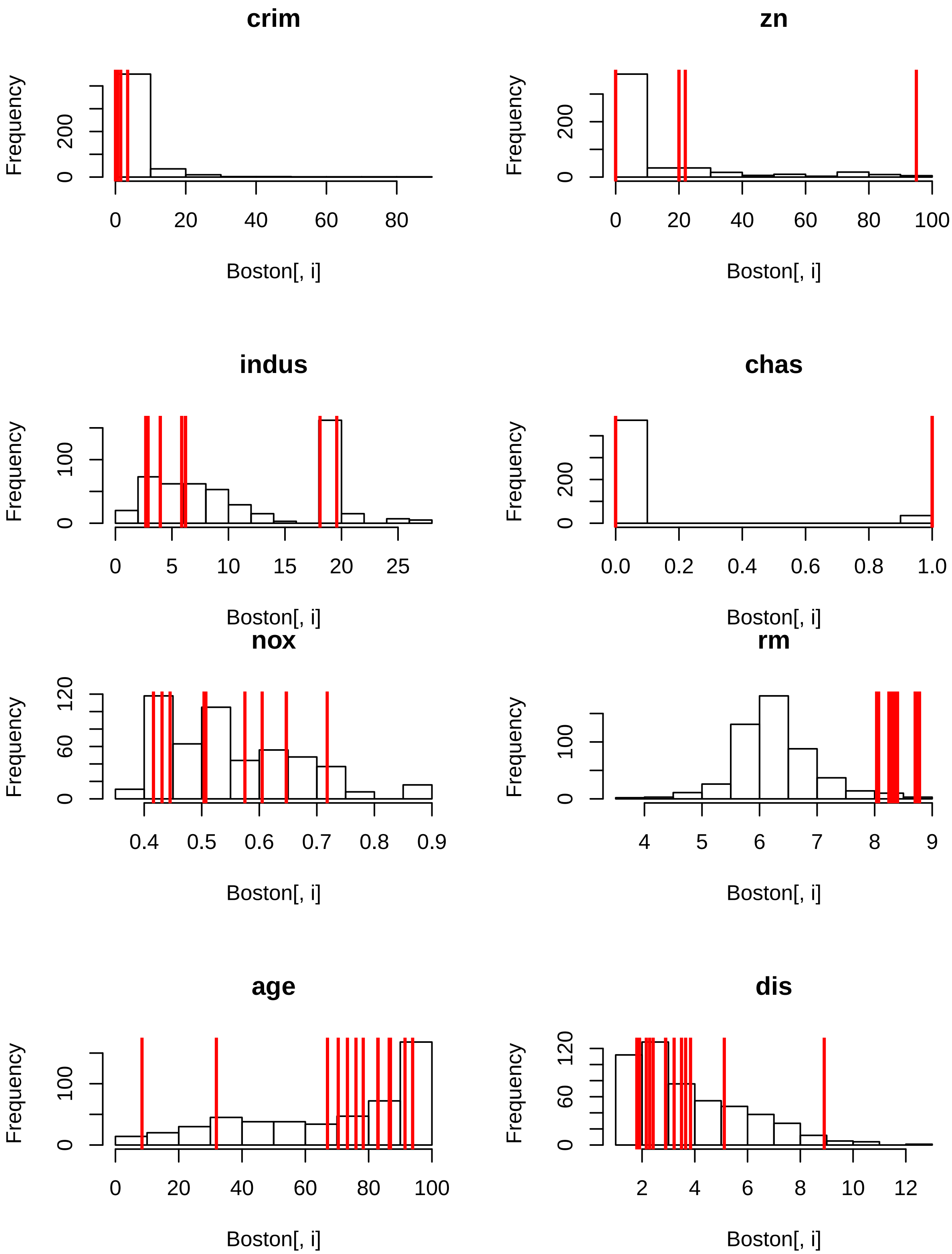
```
#more than 7 rooms per dwelling
summary(Boston$rm > 7)
```

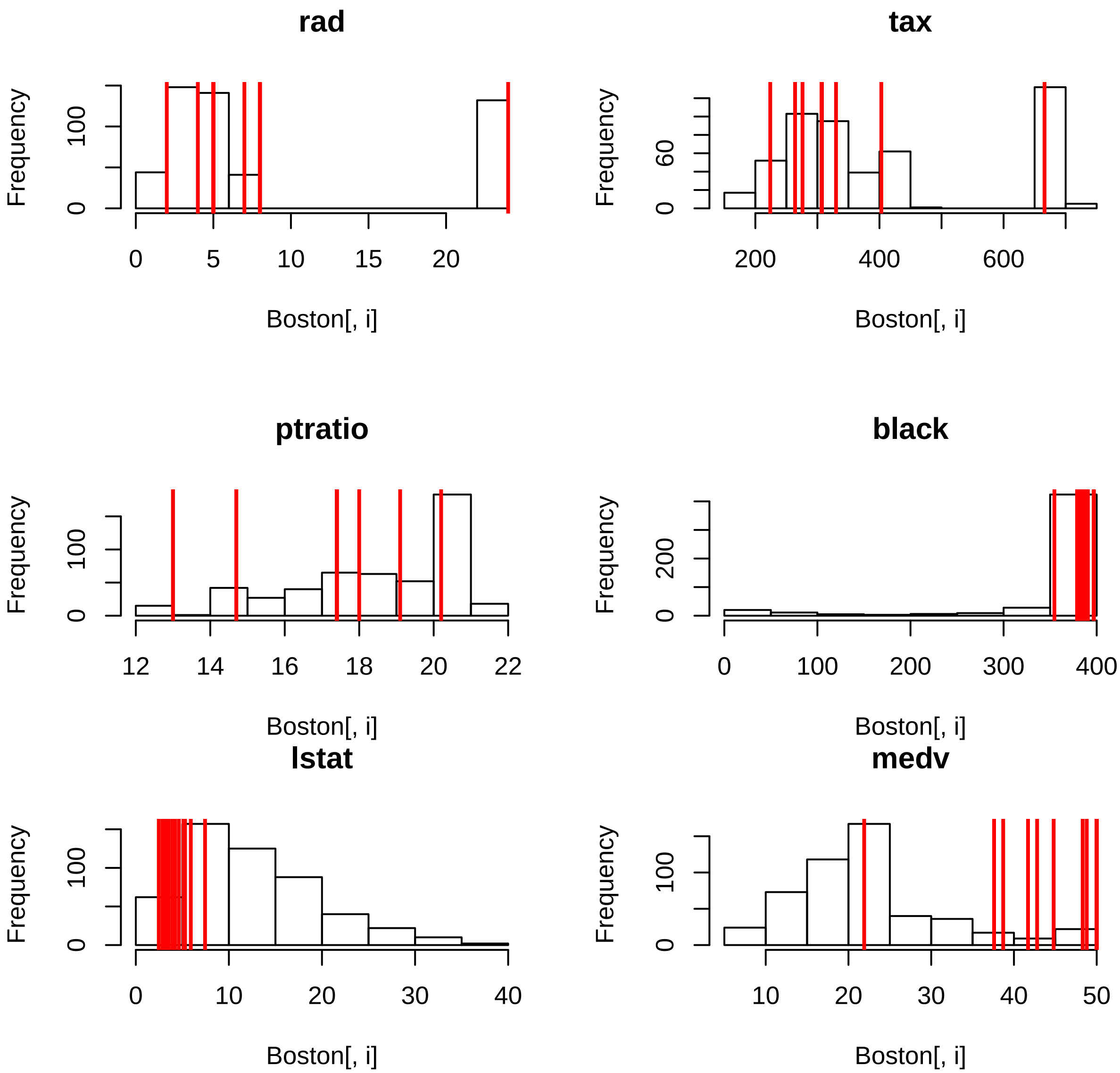
```
##    Mode   FALSE    TRUE
## logical     442      64
```

```
#more than 8 rooms per dwelling
summary(Boston$rm > 8)
```

```
##    Mode   FALSE    TRUE
## logical     493      13
```

```
idx <- Boston$rm > 8
par(mfrow=c(2,2))
for (i in 1:ncol(Boston)){
  hist(Boston[, i], main=colnames(Boston)[i])
  abline(v=Boston[idx, i], col="red", lw=2)
}
```


crim
Frequency
0
200
0
20
40
60
80
Boston[, i]
zn
Frequency
0
200
0
20
40
60
80
100
Boston[, i]
indus
Frequency
0
100
0
5
10
15
20
25
Boston[, i]
chas
Frequency
0
200
0.0
0.2
0.4
0.6
0.8
1.0
Boston[, i]
nox
Frequency
0
60
120
0.4
0.5
0.6
0.7
0.8
0.9
Boston[, i]
rm
Frequency
0
100
4
5
6
7
8
9
Boston[, i]
age
Frequency
0
100
0
20
40
60
80
100
Boston[, i]
dis
Frequency
0
60
120
2
4
6
8
10
12
Boston[, i]



##

Derive Equation 3.4

last page

Question 3.1

The null hypotheses of table 3.4 is that none of the advertising budgets of TV, radio, newspaper have an effect on sales. The corresponding p-values of TV a dn radio are highly significant, but the p-value of newspaper is not significant. Thus we reject H0 for TV and radio, but there is not significant evidence to reject H0 for nespaper. We conclude that newspaper advertising budget do not affect sales.

Question 3.3

a)

iii. For a fixed value of IQ and GPA, males earn more on average than females provided that the GPA is high enough.

Since from the regression result, if the observant is a male then, $\hat{y} = 50+20*GPA+0.07*IQ+0.01*GPA \times IQ$. If the observant is a felame then, $\hat{y} = 85 + 10 * GPA + 0.07 * IQ + 0.01 * GPA \times IQ$.

Therefore for a fixed value of IQ and GPA, if GPA greater or equal to 3.5, then males earn more on average than females.

b) For a female observant, $\hat{y} = 85 + 10 * GPA + 0.07 * IQ + 0.01 * GPA \times IQ$

```
IQ= 110
GPA=4.0

Yhat<- 85+10*GPA +0.07*IQ +0.01*GPA*IQ
Yhat
```

```
## [1] 137.1
```

Thus according to the model, the female most likely has a salary around $137,100

c) False. We should look at the p-value of the regression coefficient.

Question 3.8

(a) 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. For example:

i. Is there a relationship between the predictor and the response?
ii. How strong is the relationship between the predictor and the response?
iii. Is the relationship between the predictor and the response positive or negative?
iv. What is the predicted mpg associated with a horsepower of 98? What are the associated 95 % confidence and prediction intervals?

```
library(ISLR)
data(Auto)
fit <- lm(mpg ~ horsepower, data = Auto)
summary(fit)
```

```
##
## 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. We could perform a hypothesis testing. Suppose $H_0 : \beta = 0$, then the F-statistic is 599 with an associated p-value of 2.2e-16 which is small enough to reject the null hypothesis. Thus there is a clear evidence that shows the strong relationship between mpg and horsepower.

ii. The Adjusted R^2 value indicates that about 60.49% of the variation in the mpg is due to the horsepower.

iii. Since the coefficeint associated with horsepower is negative, the relationship between the predictor and the response is negative

iv.

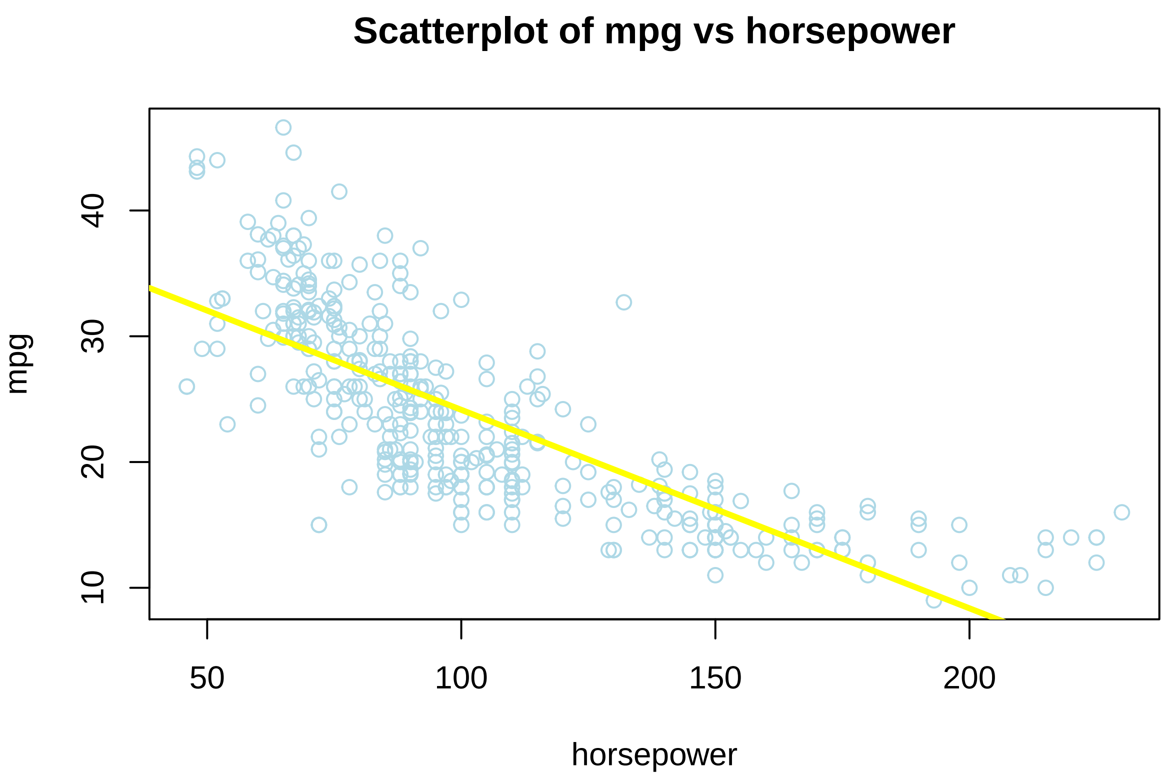
```
predict(fit, data.frame(horsepower = c(98)), interval ="confidence")
```

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

The predicted mpg associated with housepower=98 is 24.46708. The prediction interval associated with the 95% confidence is between 23.97308 and 24.96108.

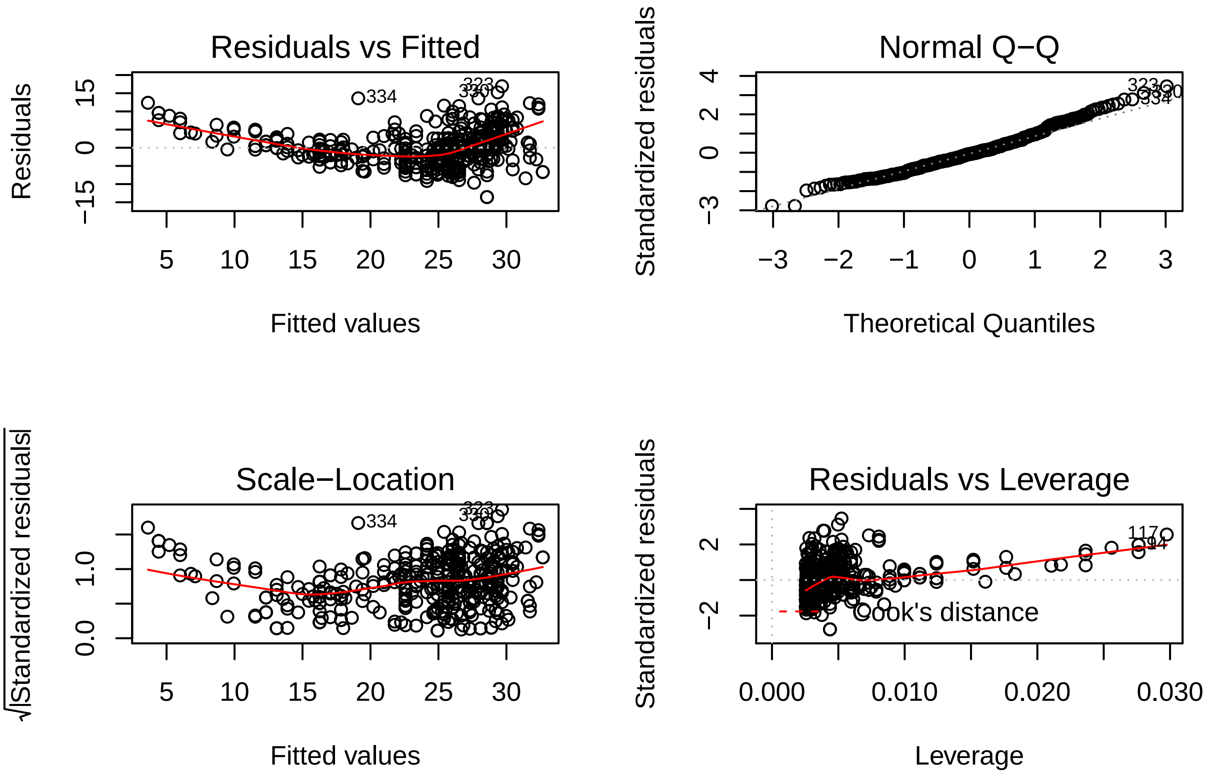
(b) Plot the response and the predictor. Use the abline() function to display the least squares regression line.

```
plot(Auto$horsepower, Auto$mpg, main = "Scatterplot of mpg vs horsepower", xlab = "horsepower", ylab =
abline(fit, col = "yellow",lw=3)
```



(c)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(fit)
```



The plot of residuals vs fitted values indicated that there could be some non linearity for the model. The plot of residuals vs leverage indicated that there could be some outliers and a few high leverage points in the data.

Question 3.15

(a) 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.

```
mod = list()
for (i in 2:ncol(Boston)){
  mod[[i]]=lm(Boston[,1]~Boston[,i])
  print(summary(mod[[i]]))
}
```

```
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 ***
## Boston[, i] -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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 **
## Boston[, i]  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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 ***
## Boston[, i]  -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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 ***
## Boston[, i]   31.249      2.999  10.419  < 2e-16 ***
```

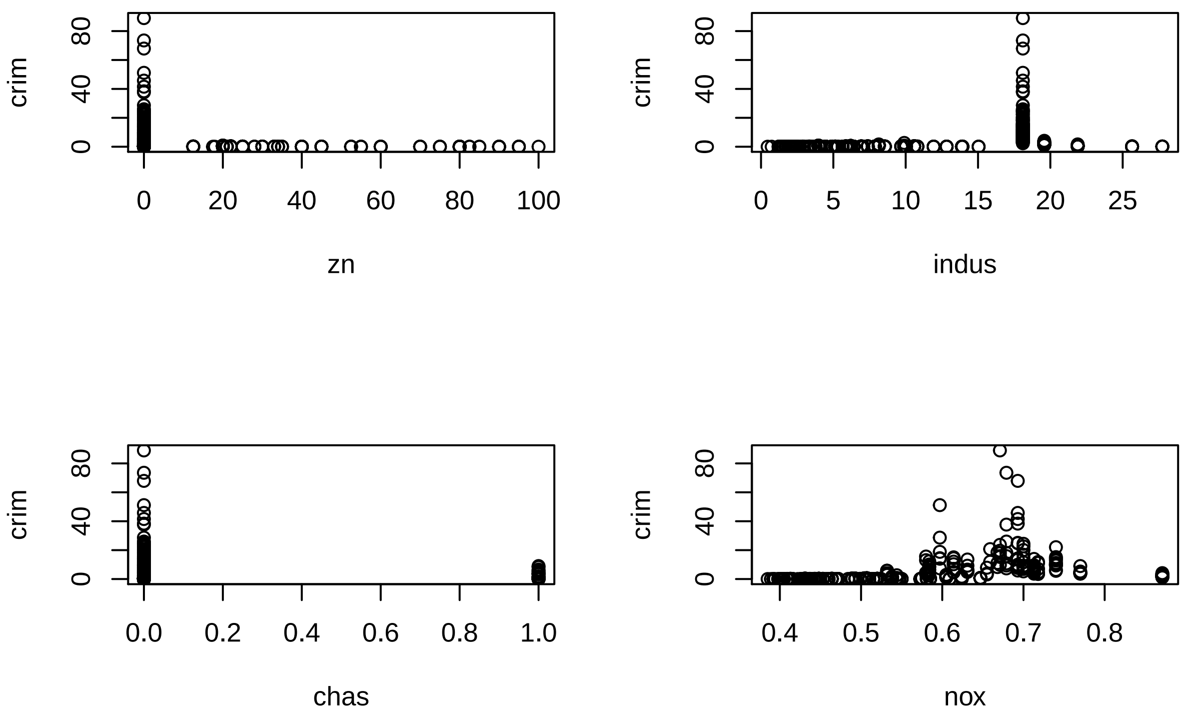
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.81 on 504 degrees of freedom
## Multiple R-squared:  0.1772, Adjusted R-squared:  0.1756
## F-statistic: 108.6 on 1 and 504 DF,  p-value: < 2.2e-16
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 ***
## Boston[, i]   -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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 ***
## Boston[, i]  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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.708 -4.134 -1.527  1.516 81.674
##
## Coefficients:
```

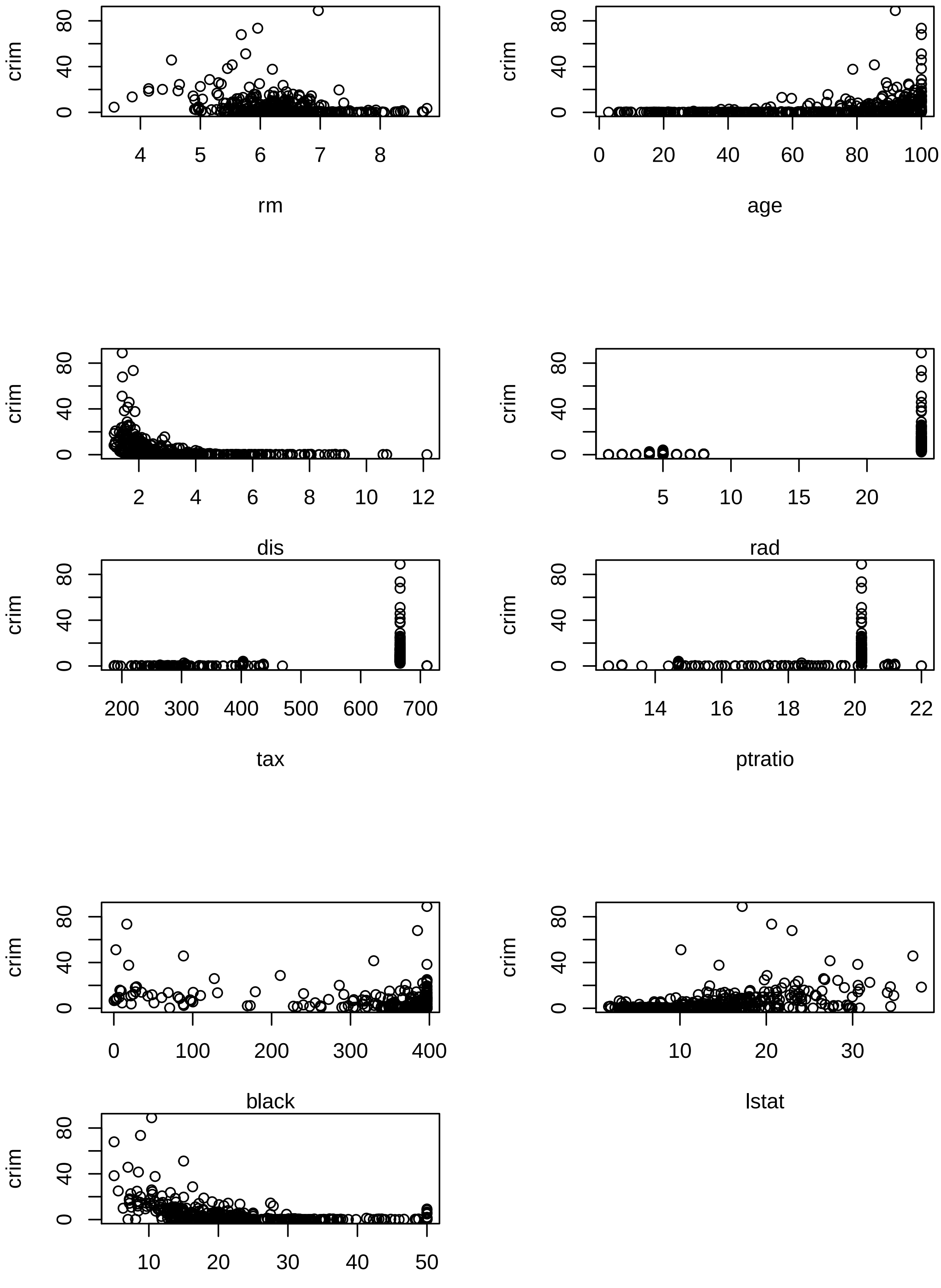
```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.4993     0.7304  13.006   <2e-16 ***
## Boston[, i]  -1.5509     0.1683  -9.213   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.965 on 504 degrees of freedom
## Multiple R-squared:  0.1441, Adjusted R-squared:  0.1425
## F-statistic: 84.89 on 1 and 504 DF,  p-value: < 2.2e-16
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 ***
## Boston[, i]  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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 ***
## Boston[, i]  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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 ***
## Boston[, i]   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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 ***
## Boston[, i] -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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
##
## 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 ***
## Boston[, i]  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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i])
```

```
##
## 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 ***
## Boston[, i] -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
```

```r
#plot the result
par(mfrow = c(2, 2))
plot(crim ~ .-crim, data = Boston)
```





All predictors have a relatively small p-value except for "chas", which represent Charles River dummy. We cany conclude that there is enough evidence to show that there are statistically significant association between each predictor and the response variable except for the "chas" variable.

(b) 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 : βj = 0?

```
mod_multi<- lm(crim ~ . - crim, data = Boston)
summary(mod_multi)
```

```
##
## Call:
## lm(formula = crim ~ . - 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
```

The p-value associated with "dis" and "rad" is less than 0.001 which are very small. The p-value associated with medv is 0.001087 which is also relatively small. Thus there is enough evidence to reject the null hypothesis for these three variables. For other variables, we fail to reject the null hypothesis. Compared to the single regression model, R-squared is higher when using a multiple regression model.

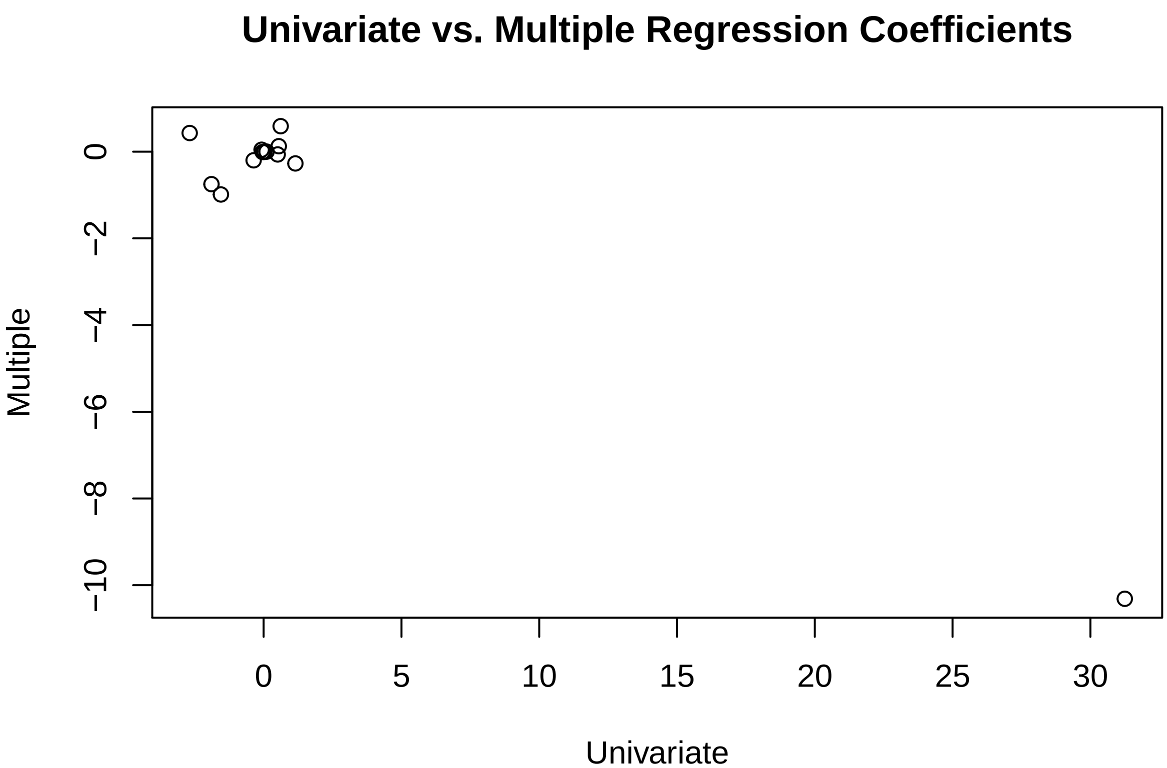
(c) How do your results from (a) compare to your results from (b)? Create a plot displaying the univariate regression coefficients from (a) on the x-axis, and the multiple regression coefficients from (b) on the y-axis. That is, each predictor is displayed as a single point in the plot. Its coefficient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

```
univ<- vector()
for (i in 2:ncol(Boston)){
  univ[i]<-mod[[i]]$coefficients[2]
}
names(univ)<- colnames(Boston)

#regression coefficients from (a)
```

```r
univ<- univ[-1]
#regression coefficients from (b)
multi<- mod_multi$coefficients[2:14]

plot(univ,multi,main = "Univariate vs. Multiple Regression Coefficients",
    xlab = "Univariate", ylab = "Multiple")
```



(d) Is there evidence of non-linear association between any of the predictors and the response? To answer this question, for each predictor X, fit a model of the form

```r
mod2 = list()
for (i in 2:ncol(Boston)){
  mod2[[i]]=lm(Boston[,1]~Boston[,i]+I(Boston[,i]^2)+I(Boston[,i]^3))
  print(summary(mod2[[i]]))
}
```

```
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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 ***
## Boston[, i]      -3.322e-01  1.098e-01  -3.025  0.00261 **
## I(Boston[, i]^2)  6.483e-03  3.861e-03   1.679  0.09375 .
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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 *
## Boston[, i]      -1.9652129  0.4819901  -4.077 5.30e-05 ***
## I(Boston[, i]^2)  0.2519373  0.0393221   6.407 3.42e-10 ***
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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 ***
## Boston[, i]       -1.8928     1.5061  -1.257    0.209
## I(Boston[, i]^2)       NA         NA      NA       NA
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
```

```
##     i]^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 ***
## Boston[, i]      -1279.37     170.40  -7.508 2.76e-13 ***
## I(Boston[, i]^2)  2248.54     279.90   8.033 6.81e-15 ***
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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 .
## Boston[, i]      -39.1501    31.3115  -1.250   0.2118
## I(Boston[, i]^2)   4.5509     5.0099   0.908   0.3641
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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
## Boston[, i]       2.737e-01  1.864e-01   1.468  0.14266
```

```
## I(Boston[, i]^2) -7.230e-03  3.637e-03  -1.988  0.04738 *
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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 ***
## Boston[, i]      -15.5543     1.7360  -8.960  < 2e-16 ***
## I(Boston[, i]^2)   2.4521     0.3464   7.078 4.94e-12 ***
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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
## Boston[, i]       0.512736   1.043597   0.491    0.623
## I(Boston[, i]^2) -0.075177   0.148543  -0.506    0.613
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
```

```
##     i]^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
## Boston[, i]      -1.533e-01  9.568e-02  -1.602    0.110
## I(Boston[, i]^2)  3.608e-04  2.425e-04   1.488    0.137
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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 **
## Boston[, i]      -82.36054   27.64394  -2.979  0.00303 **
## I(Boston[, i]^2)   4.63535    1.60832   2.882  0.00412 **
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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 ***
## Boston[, i]      -8.356e-02  5.633e-02  -1.483    0.139
## I(Boston[, i]^2)  2.137e-04  2.984e-04   0.716    0.474
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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
## Boston[, i]      -0.4490656  0.4648911  -0.966   0.3345
## I(Boston[, i]^2)  0.0557794  0.0301156   1.852   0.0646 .
## I(Boston[, i]^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
##
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^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 ***
## Boston[, i]      -5.0948305  0.4338321 -11.744  < 2e-16 ***
## I(Boston[, i]^2)  0.1554965  0.0171904   9.046  < 2e-16 ***
## I(Boston[, i]^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
```

```r
#model for "chas"
mod2[[4]]
```

```
##
```

```
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^3))
##
## Coefficients:
##      (Intercept)       Boston[, i]  I(Boston[, i]^2)  I(Boston[, i]^3)
##            3.744            -1.893                NA                NA
```

```
#model for "age"
mod2[[7]]
```

```
##
## Call:
## lm(formula = Boston[, 1] ~ Boston[, i] + I(Boston[, i]^2) + I(Boston[,
##     i]^3))
##
## Coefficients:
##      (Intercept)       Boston[, i]  I(Boston[, i]^2)  I(Boston[, i]^3)
##       -2.549e+00         2.737e-01        -7.230e-03         5.745e-05
```

For variables “indus”, “nox”, “dis”, “ptracio”, and “medv”, there is evidence of a non-linear relationship, as each of these variables squared and cubed terms is found to be statistically signficant. For the “chas” variable we get NA values for the squared and cubed term, since it’s a dummy variable which only contains 0s and 1s. These values will not change if they are squared or cubed. The linear term of variable “age” becomes statistically insignficant compared with the squared and cubed term. For other variables, there is no obvious non-linear relationship between the predictor and outcome variables.