2.4 Exercises

Luke Schanne

- Exercise 8
- Exercise 9
- Exercise 10

Exercise 8

This exercise relates to the **College** data set, which can be found in the file **College.csv**. It contains a number of variables for 777 different universities and colleges in the US. The variables are:

- **Private**: Public/private indicator
- **Apps**: Number of applications received
- **Accept**: Number of new students enrolled
- **Top10perc**: New students from top 10% of high school class
- **Top25perc**: New students from top 25% of high school class
- **F.Undergrad**: Number of full-time undergraduates
- **P.Undergrad**: Number of part-time undergraduates
- **Outstate**: Out-of-state tuition
- **Room.Board**: Room and board costs
- **Books**: Estimated book costs
- **Personal**: Estimated personal spending
- **PhD**: Percent of faculty with Ph.D.'s
- **Terminal**: Percent of faculty with terminal degree
- **S.F.Ratio**: Student/faculty ratio
- **perc.alumni**: Percent of alumni who donate
- **Expend**: Instructional expenditure per student
- **Grad.Rate**: Graduation rate

Before reading the data into R, it can be viewed in Excel or a text editor.

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.

```
fh='D:\\GoogleDrive\\Introduction to Statistical Learning with Applications in R\\data-s
ets\\College.csv'
college = read.csv(file=fh,header=TRUE)
```

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:

```
> rownames(college)=college[,1]
> fix(college)
```

You should see that there is now a **row.names** column with the name of each university recorded. This means that R has given each row a name corresponding to the appropriate university. R will not try to perform calculations on the row names. However, we still need to elimnate the first colum in the data where the names are stored. Try

```
> college=college[,-1]
> fix(college)
```

Now you should see that the first data column is **Private**. Note that another column labeled **row.names** now apperas before the **Private** column. However, this is not a data column but rather teh name that R is giving to each row.

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

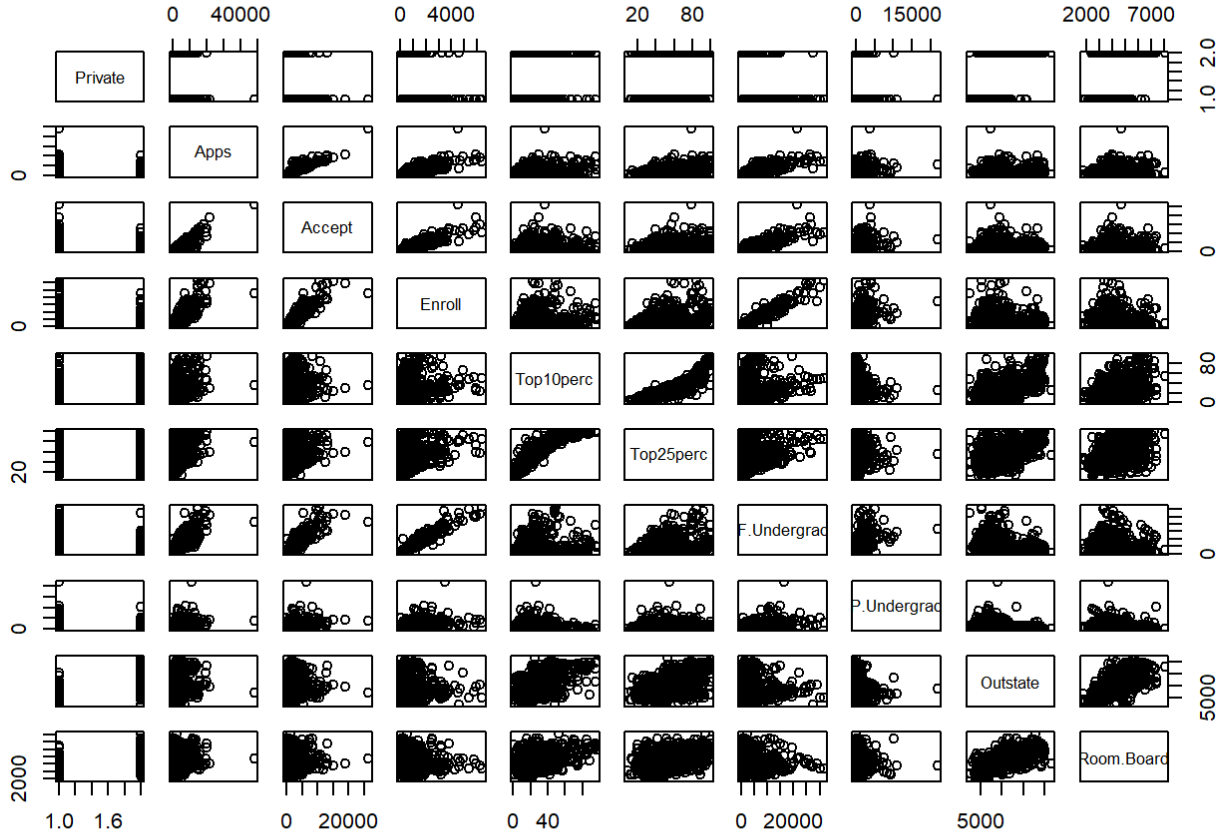
c. i. Use the **summary()** function to produce a numerical summary of the variables in the data set.

```
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  
```

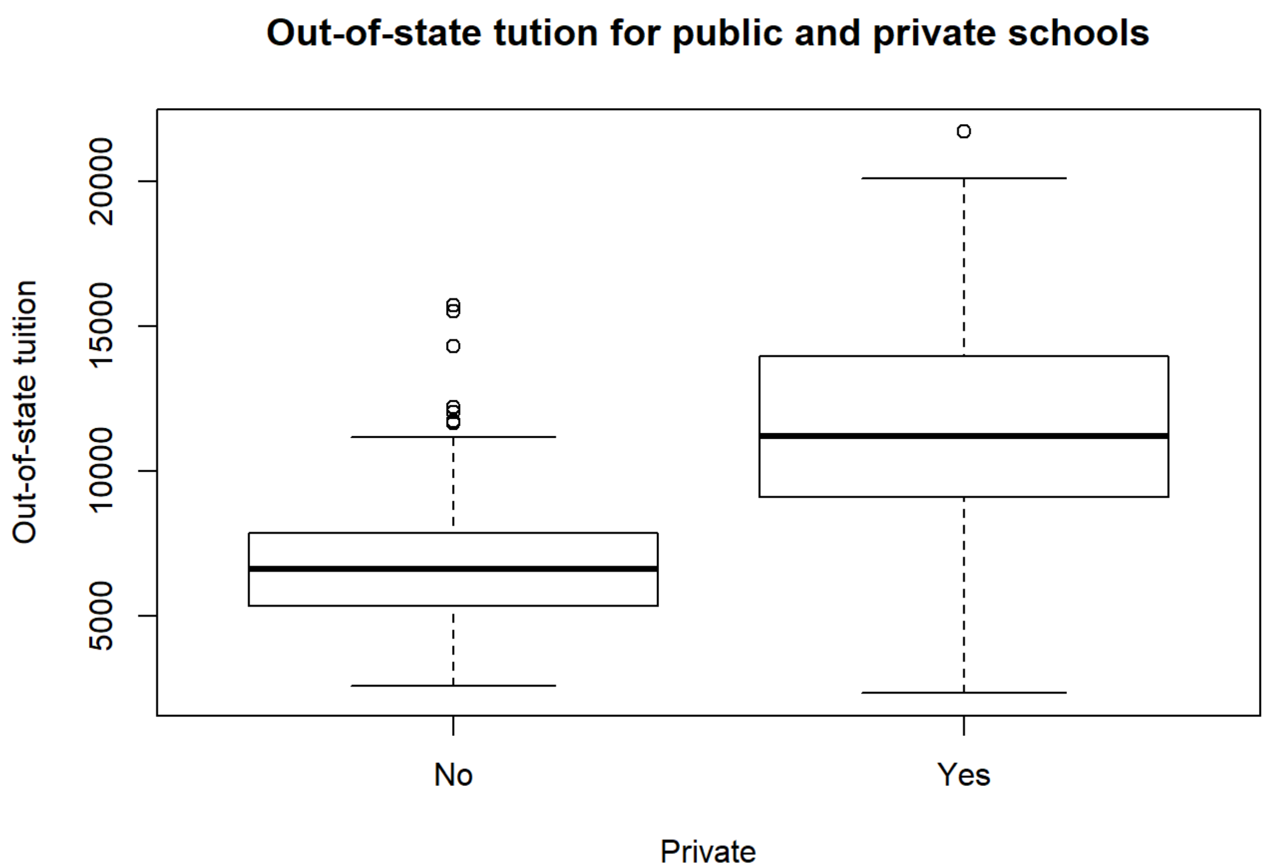
ii. Use the **pairs()** function to produce a scatterplot matrix of the first ten columns or variables of the data. Recall that you can reference the first ten columns of a matrix **A** using **A[,1:10]**.

```
pairs(college[,1:10])
```



iii. Use the **plot()** function to produce side-by-side boxplots of **Outstate** versus **Private**.

```
plot(x=college$Private, y=college$Outstate, xlab='Private', ylab='Out-of-state tui
tion', main='Out-of-state tution for public and private schools')
```



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%.

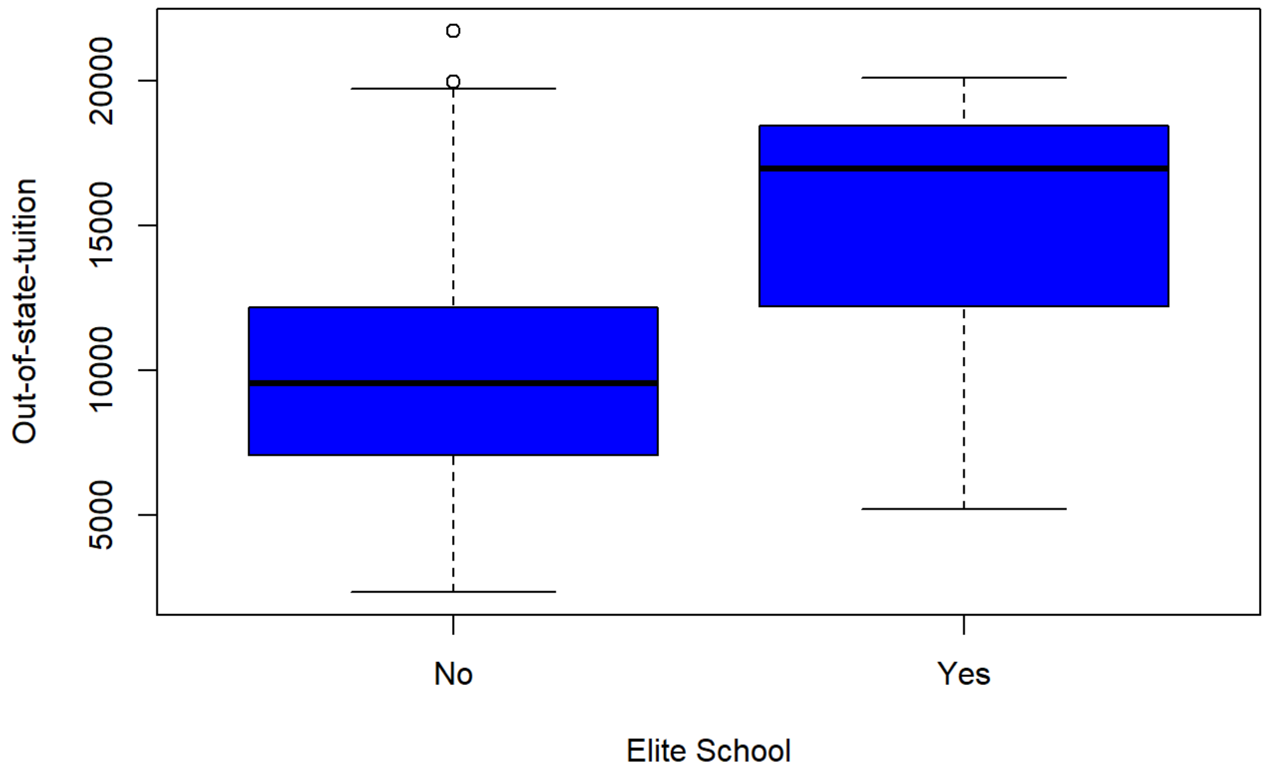
```
Elite=rep("No",nrow(college))
Elite[college$Top10perc>50]="Yes"
Elite=as.factor(Elite)
college=data.frame(college,Elite)
fix(college)
```

Use the summary() function to see how many elite universities there are. Now use the **plot()** function to produce side-by-side boxplots of **Outstate** versus **Elite**.

```
summary(Elite)
```

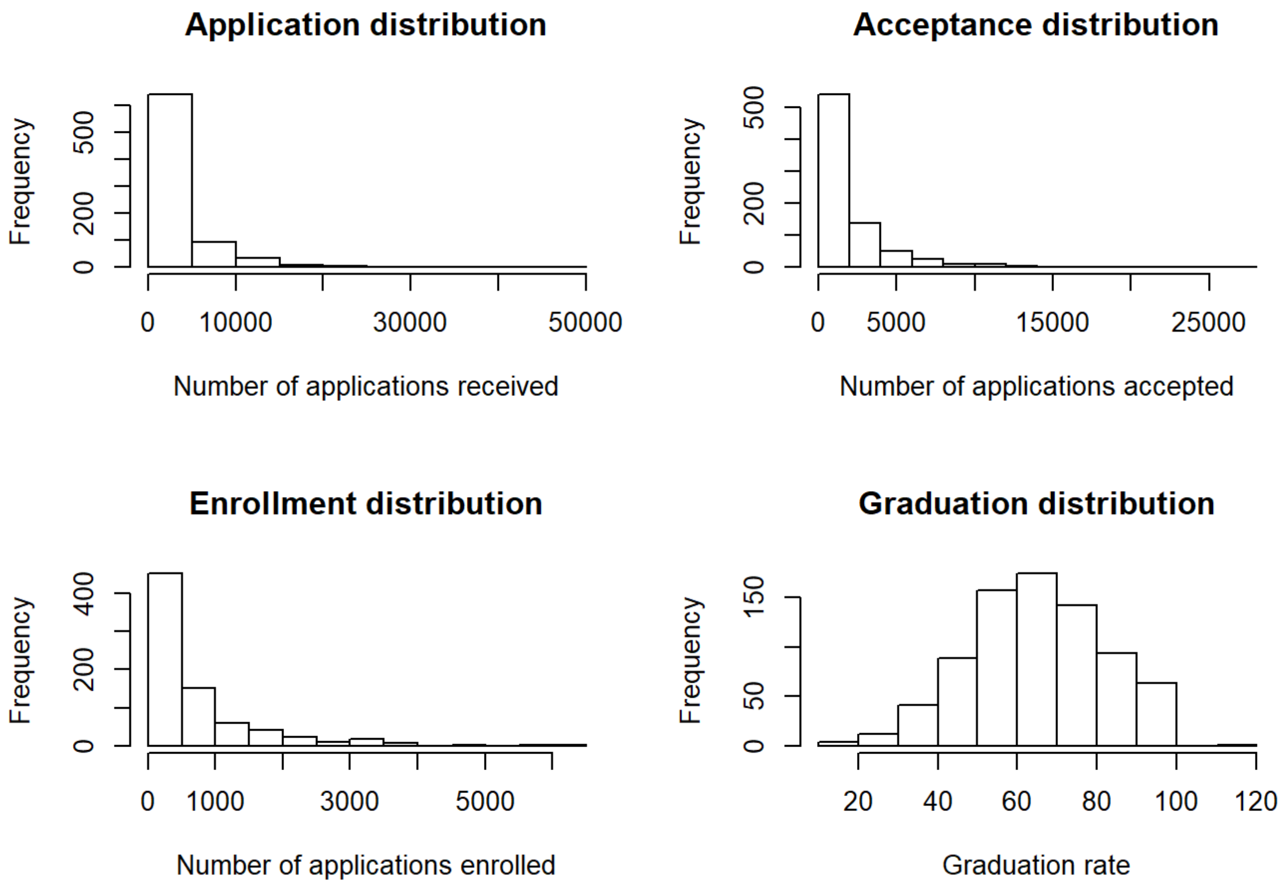
```
##  No Yes
## 699  78
```

```
plot(x=Elite, y=college$Outstate, xlab='Elite School', ylab='Out-of-state-tuition'
, col='blue')
```



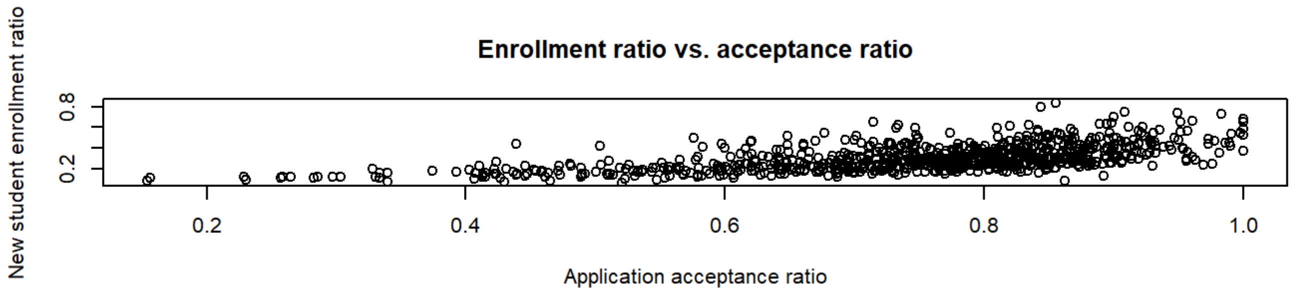
v. Use the **hist()** function to produce some histograms with diffeing numbers of bins for a few of the quantitative variables. You may find the command **par(mfrow=c(2,2))** useful: it will divide the print window into four regions so that four plots can be made simultaneously. Modifying the arguments to this function will divide the screen in other ways.

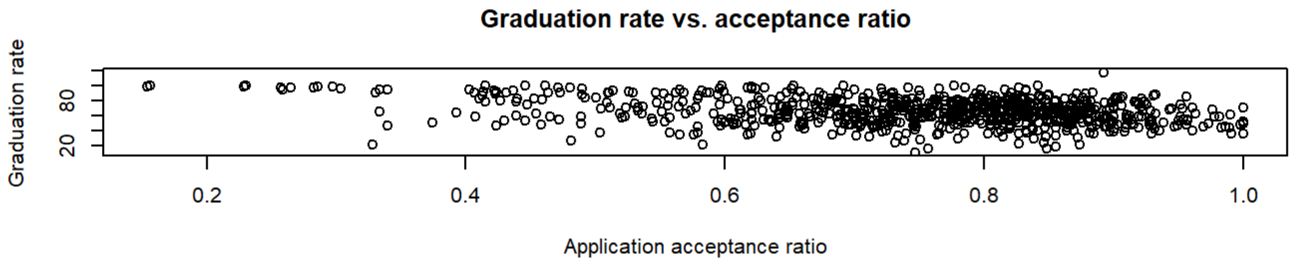
```
par(mfrow=c(2,2))
hist(x=college$Apps, xlab='Number of applications received', main='Application dis
tribution', breaks=15)
hist(x=college$Accept, xlab='Number of applications accepted', main='Acceptance di
stribution', breaks=15)
hist(x=college$Enroll, xlab='Number of applications enrolled', main='Enrollment di
stribution', breaks=15)
hist(x=college$Grad.Rate, xlab='Graduation rate', main='Graduation distribution',
 breaks=15)
```

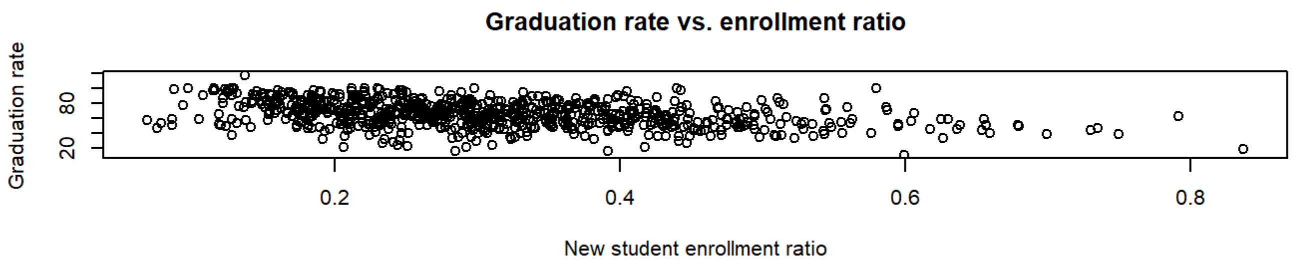


vi. Continue exploring the data, and provide a brief summary of what you discover.

```
acceptanceRatio = college$Accept / college$Apps
enrollmentRatio = college$Enroll / college$Apps
accLab = 'Application acceptance ratio'
enrLab = 'New student enrollment ratio'
graLab = 'Graduation rate'
par(mfrow=c(3,1))
plot(x=acceptanceRatio, y=enrollmentRatio, xlab=accLab, ylab=enrLab, main='Enrollm
ent ratio vs. acceptance ratio')
plot(x=acceptanceRatio, y=college$Grad.Rate, xlab=accLab, ylab=graLab, main='Gradu
ation rate vs. acceptance ratio')
plot(x=enrollmentRatio, y=college$Grad.Rate, xlab=enrLab, ylab=graLab, main='Gradu
ation rate vs. enrollment ratio')
```







These trends are surprising to me. I would have expected that enrollment ratio would be higher at more competitive schools (i.e. schools with a lower acceptance ratio), but the opposite seems to be true in this data. Also, I would have expected graduation rate to be lower at more competitive schools, but again the opposite seems to be true. There doesn't seem to be much of a trend with graduation rate vs. acceptance ratio.

Exercise 9

This exercise involves the **Auto** data set studied in the lab. Make sure that the missing values have been removed from the data.

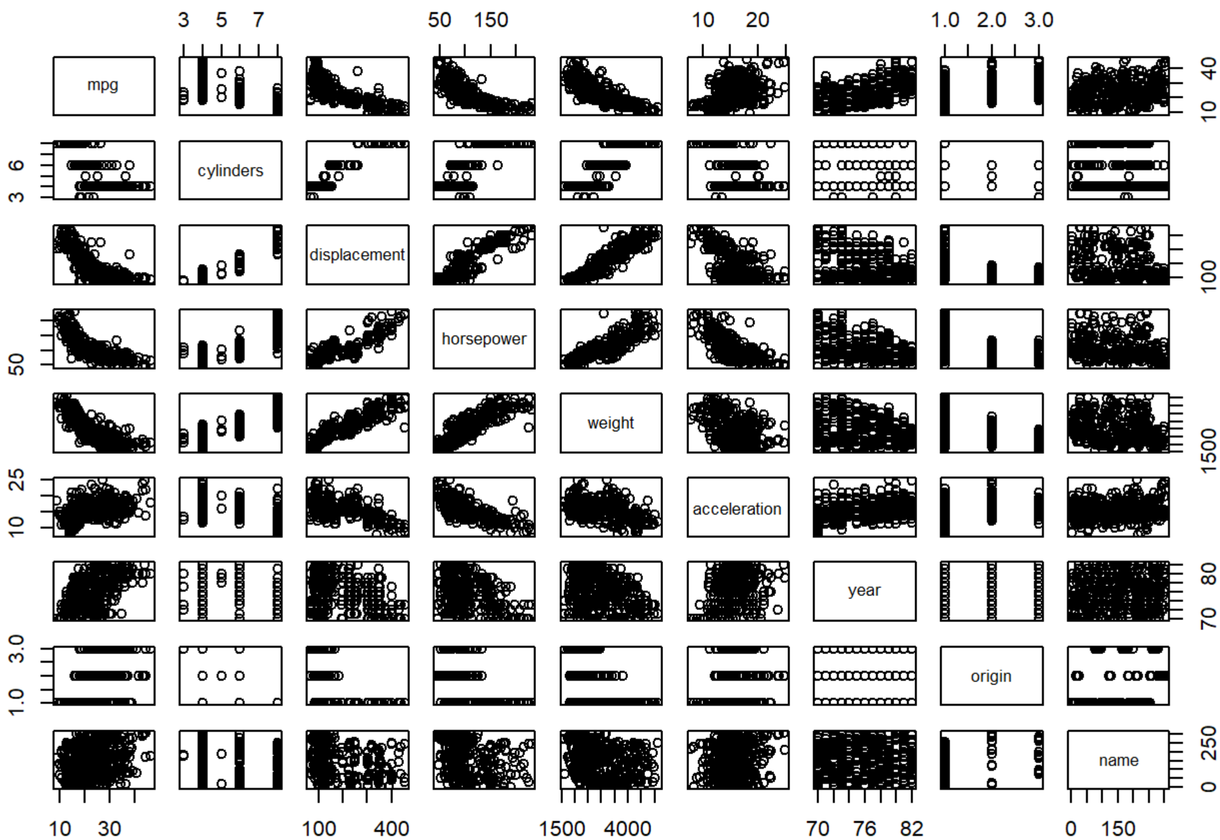
```
fh = 'D:/GoogleDrive/Introduction to Statistical Learning with Applications in R/data-se
ts/Auto.csv'
Auto = read.csv(file=fh, header=T, na.strings='?')
Auto = na.omit(Auto)
```

a. Which of the predictors are quantitative, and which are qualitative?

```
summary(Auto)
```

```
##       mpg          cylinders      displacement     horsepower   
##  Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0  
##  1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0  
##  Median :22.75   Median :4.000   Median :151.0   Median : 93.5  
##  Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5  
##  3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0  
##  Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0  
##                                                                 
##      weight      acceleration        year           origin     
##  Min.   :1613   Min.   : 8.00   Min.   :70.00   Min.   :1.000  
##  1st Qu.:2225   1st Qu.:13.78   1st Qu.:73.00   1st Qu.:1.000  
##  Median :2804   Median :15.50   Median :76.00   Median :1.000  
##  Mean   :2978   Mean   :15.54   Mean   :75.98   Mean   :1.577  
##  3rd Qu.:3615   3rd Qu.:17.02   3rd Qu.:79.00   3rd Qu.:2.000  
##  Max.   :5140   Max.   :24.80   Max.   :82.00   Max.   :3.000  
##                                                                 
##                  name    
##  amc matador       :  5  
##  ford pinto        :  5  
##  toyota corolla    :  5  
##  amc gremlin       :  4  
##  amc hornet        :  4  
##  chevrolet chevette:  4  
##  (Other)           :365
```

```
pairs(Auto)
```



Quantitative predictors: mpg, displacement, horsepower, weight, acceleration Qualitative predictors: cylinders, year, origin, name

b. What is the *range* of each quantitative predictor? You can answer this using the **range()** function.

```
quantFields = c('mpg', 'displacement', 'horsepower', 'weight', 'acceleration')
for (field in quantFields){
  x = range(Auto[[field]])
  print(sprintf('%s range: %.0f - %.0f', field, x[1], x[2]))
}
```

```
## [1] "mpg range: 9 - 47"
## [1] "displacement range: 68 - 455"
## [1] "horsepower range: 46 - 230"
## [1] "weight range: 1613 - 5140"
## [1] "acceleration range: 8 - 25"
```

c. What is the mean and standard deviation of each quantitative predictor?

```
# Note that we already have these values from the summary produced above, but I assume t
he author
# is looking for the mean and sd functions to be used here
for (field in quantFields){
  data = Auto[[field]]
  print(sprintf('%s %s: %.2f', field, c('mean', 'std'), c(mean(data), sd(data))))
}
```

```
## [1] "mpg mean: 23.45" "mpg std: 7.81"
## [1] "displacement mean: 194.41" "displacement std: 104.64"
## [1] "horsepower mean: 104.47" "horsepower std: 38.49"
## [1] "weight mean: 2977.58" "weight std: 849.40"
## [1] "acceleration mean: 15.54" "acceleration std: 2.76"
```

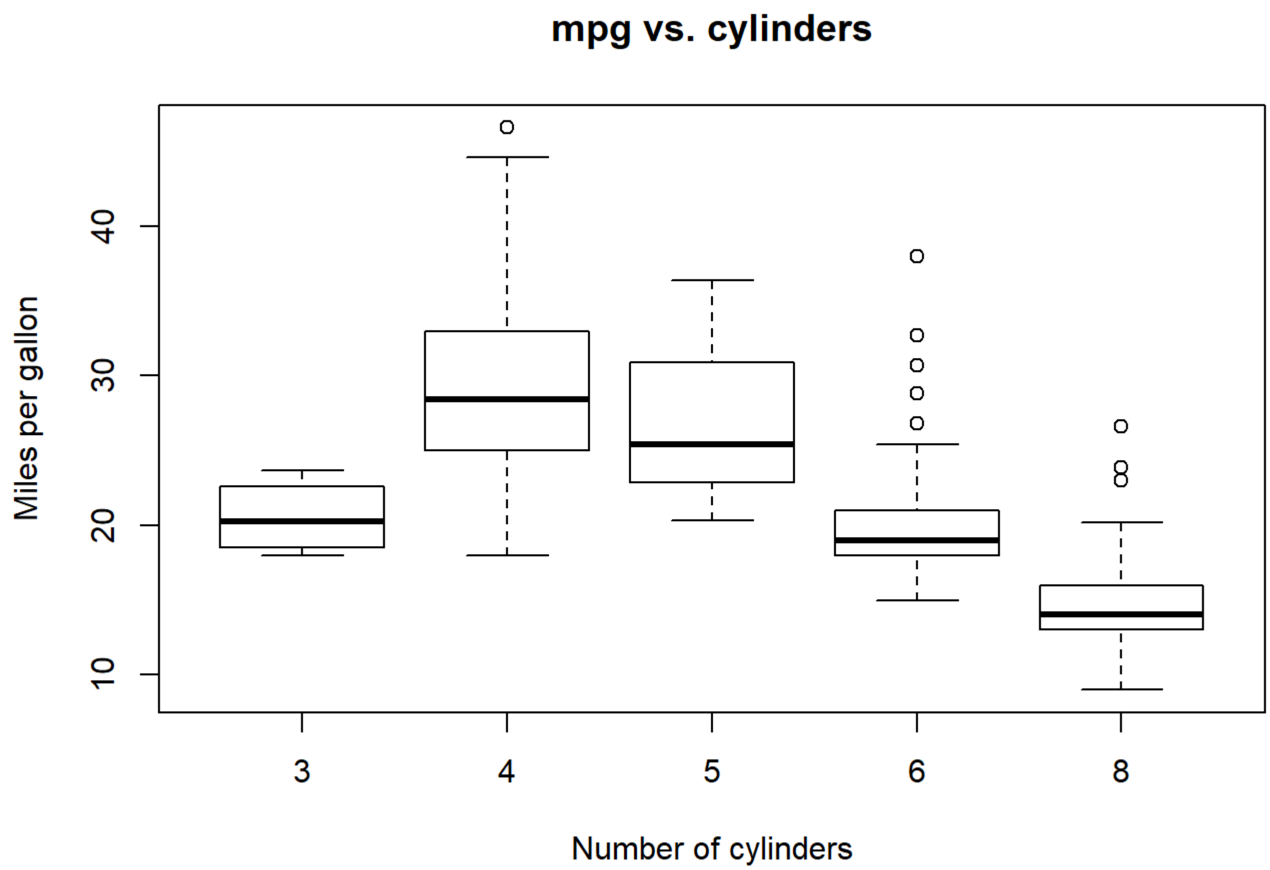
d. Now remove the 10th through 85th observations. What is the range, mean, and standard deviation of each predictor in the subset of the data that remains?

```
# Note: you can't use the 10:85 shorthand combined with negative indexing to drop rows;
 you have to write out the call to the seq() function
reducedAuto = Auto[-seq(10,85),]
for (field in quantFields){
  data = reducedAuto[[field]]
  print(sprintf('%s %s: %.2f', field, c('mean', 'std'), c(mean(data), sd(data))))
}
```

```
## [1] "mpg mean: 24.40" "mpg std: 7.87"
## [1] "displacement mean: 187.24" "displacement std: 99.68"
## [1] "horsepower mean: 100.72" "horsepower std: 35.71"
## [1] "weight mean: 2935.97" "weight std: 811.30"
## [1] "acceleration mean: 15.73" "acceleration std: 2.69"
```

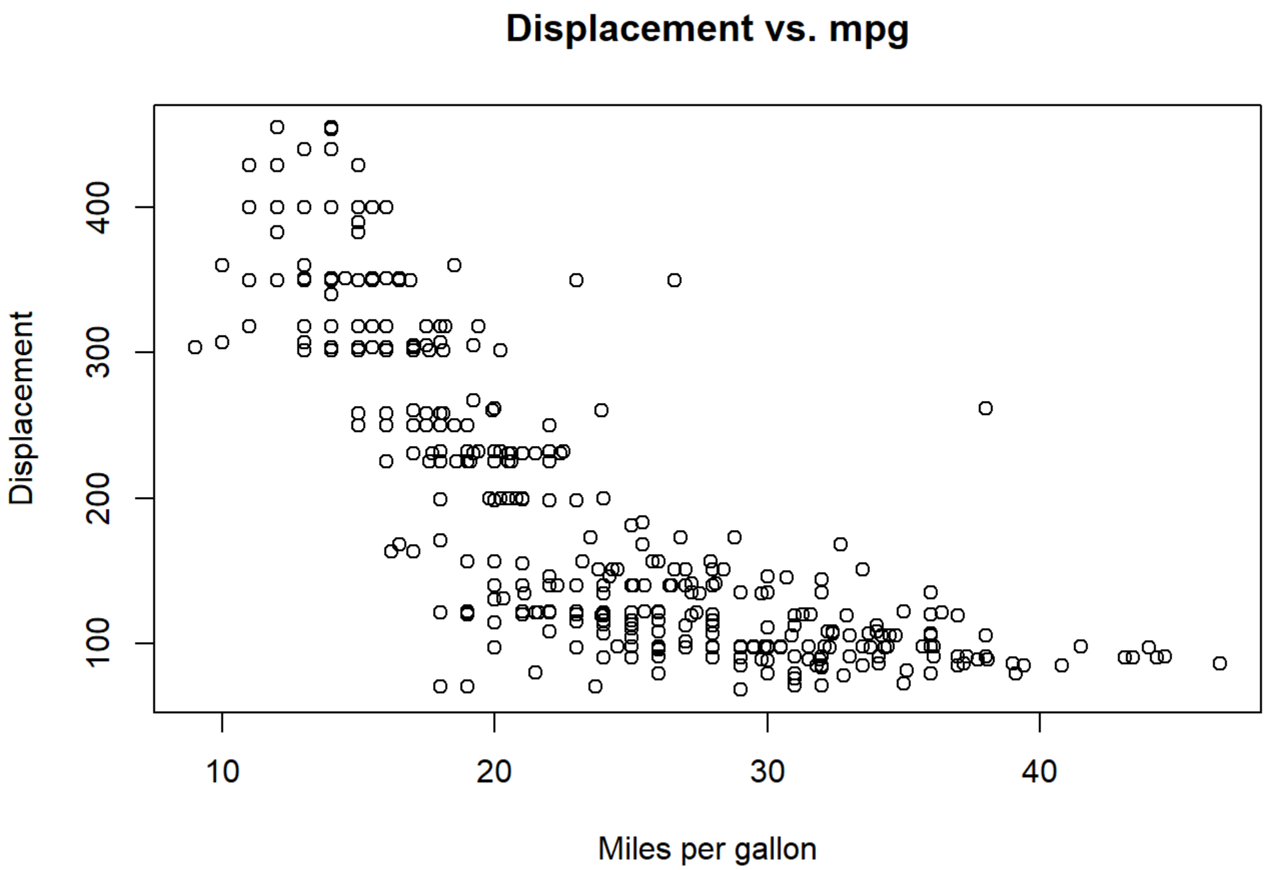
e. Using the full data set, investigate the predictors graphically, using scatterplots or other tools of your choice. Create some plots highlighting the relationships among the predictors. Comment on your findings.

```
cylinders = as.factor(Auto$cylinders)
plot(x=cylinders, y=Auto$mpg, xlab='Number of cylinders', ylab='Miles per gallon', main=
'mpg vs. cylinders')
```



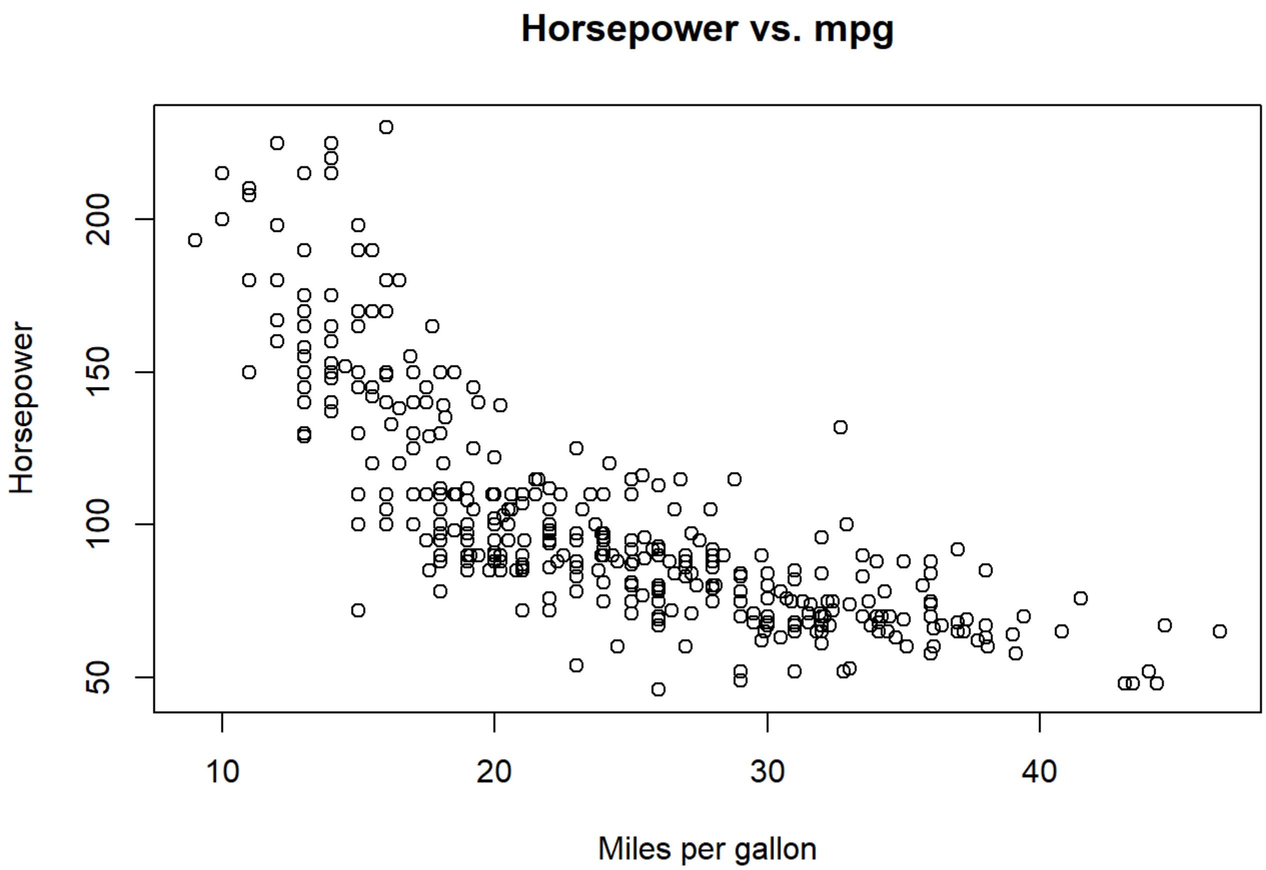
From this data, it appears that 4 cylinders is ideal for maximizing the gas mileage of a car. Additional cylinders beyond 4 begin to drastically decrease the fuel efficiency of a car.

```
plot(x=Auto$mpg, y=Auto$displacement, xlab='Miles per gallon', ylab='Displacement', main
='Displacement vs. mpg')
```



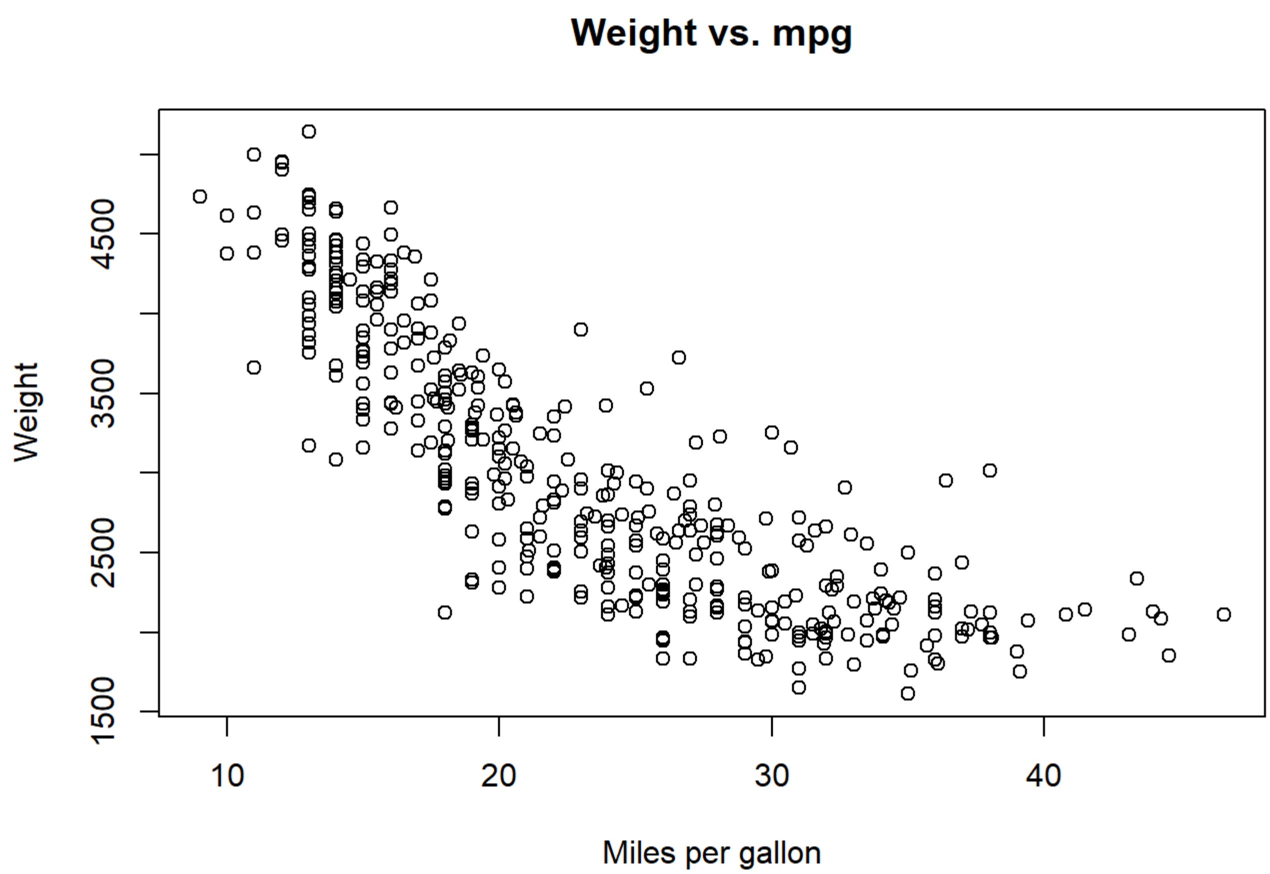
Miles per gallon seems to increase exponentially as displacement decreases.

```
plot(x=Auto$mpg, y=Auto$horsepower, xlab='Miles per gallon', ylab='Horsepower', main='Ho
rsepower vs. mpg')
```



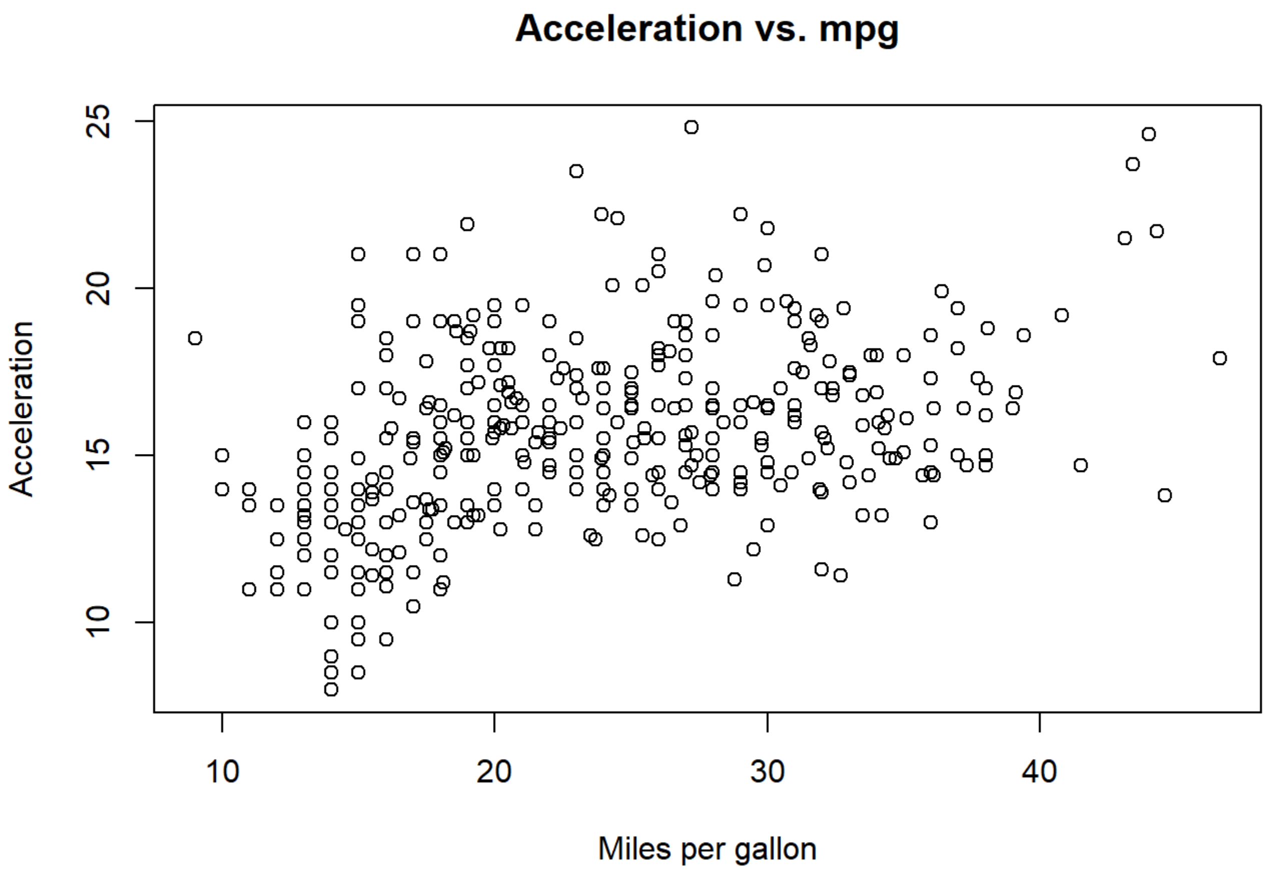
Horsepower seems to share a similar relationship to mpg as displacement does. (i.e. mpg tends to increase exponentially as horsepower decreases.)

```
plot(x=Auto$mpg, y=Auto$weight, xlab='Miles per gallon', ylab='Weight', main='Weight vs.
mpg')
```



Not surprisingly, fuel efficiency decreases exponentially as the weight of the car increases.

```
plot(x=Auto$mpg, y=Auto$acceleration, xlab='Miles per gallon', ylab='Acceleration', main
='Acceleration vs. mpg')
```

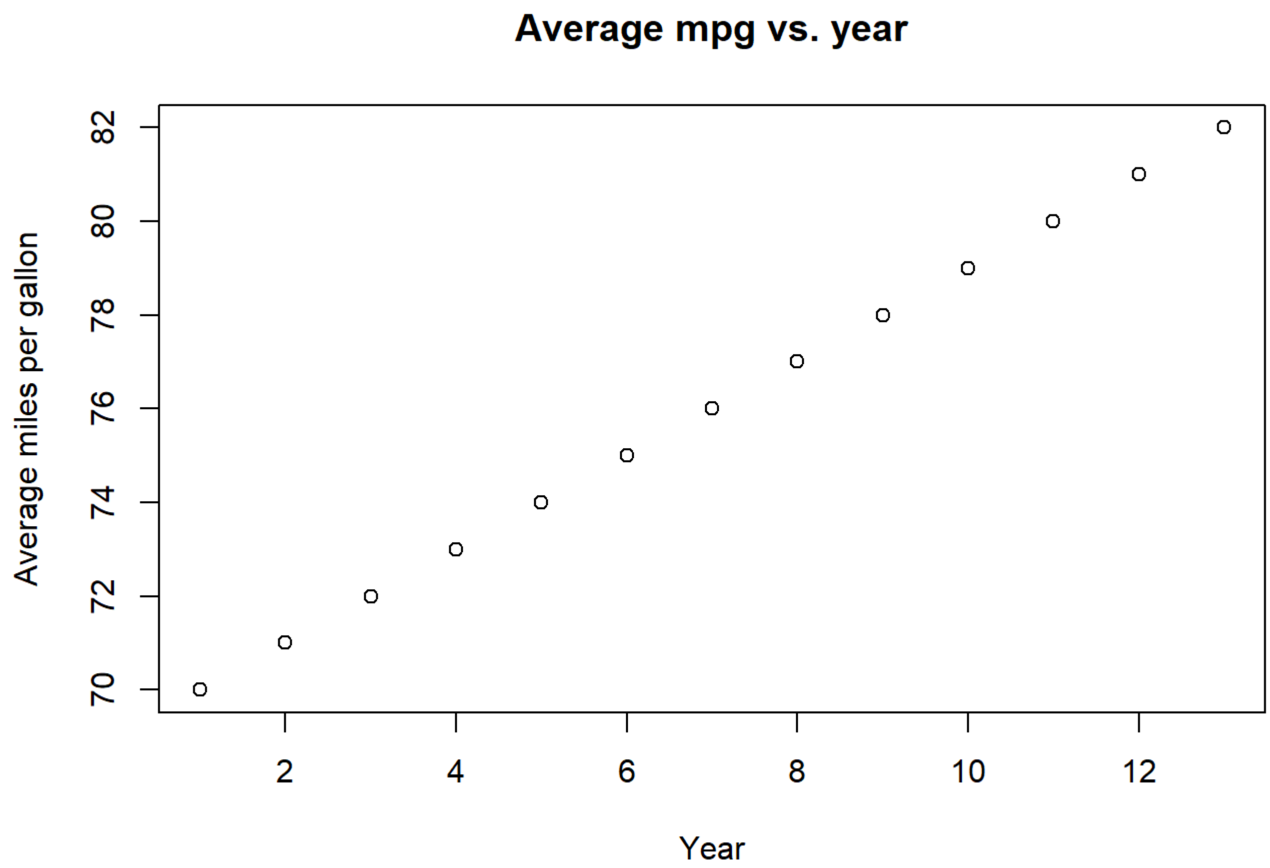



```
cor(x=Auto$mpg, y=Auto$acceleration)
```

```
## [1] 0.4233285
```

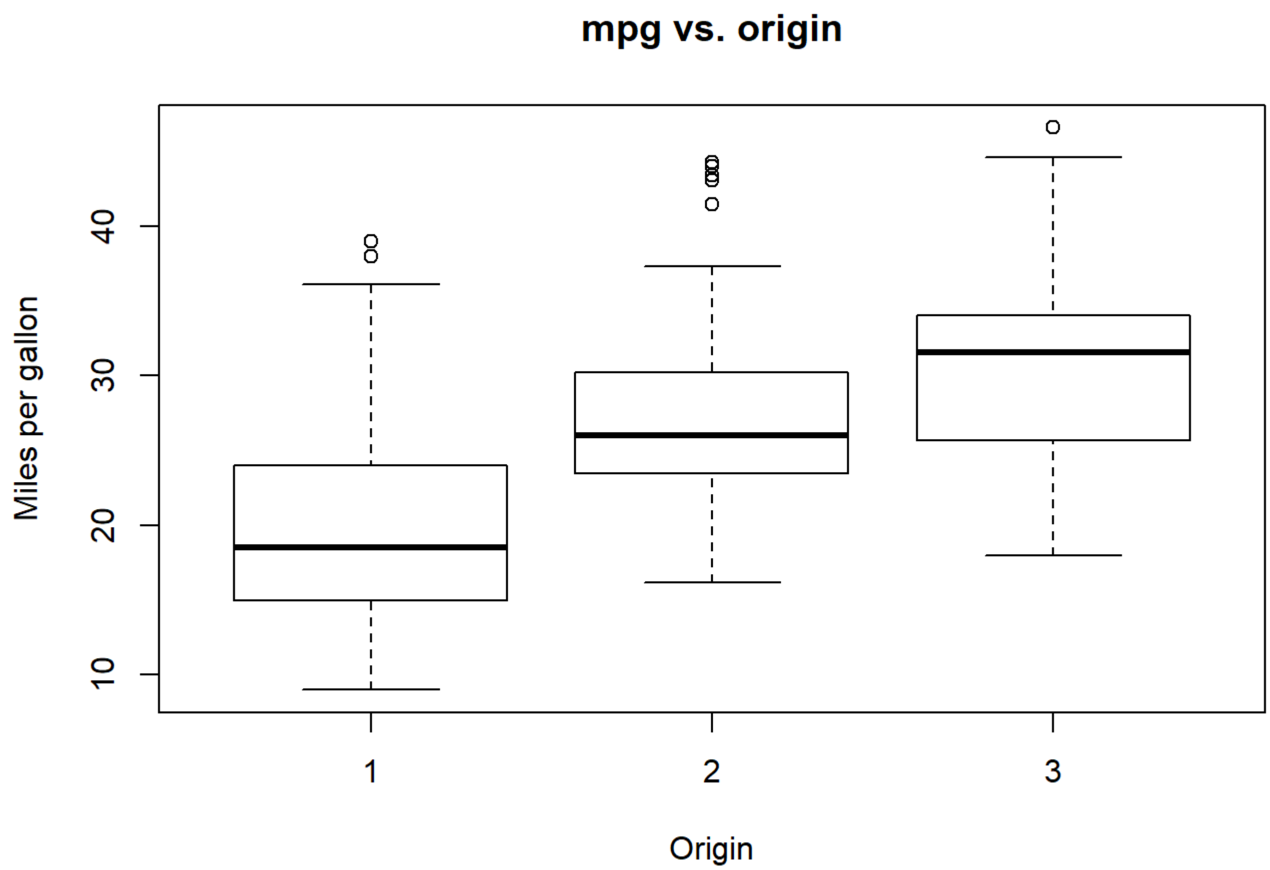
Acceleration is not strongly correlated to the gas mileage of a car. Although there is not a definitive relationship here, the mpg does tend to increase as acceleration also increases.

```
year_mpg = aggregate(Auto[,1], list(year=Auto$year), mean)
plot(x=year_mpg$year, y=year_mpg$mpg, xlab='Year', ylab='Average miles per gallon', main
='Average mpg vs. year')
```



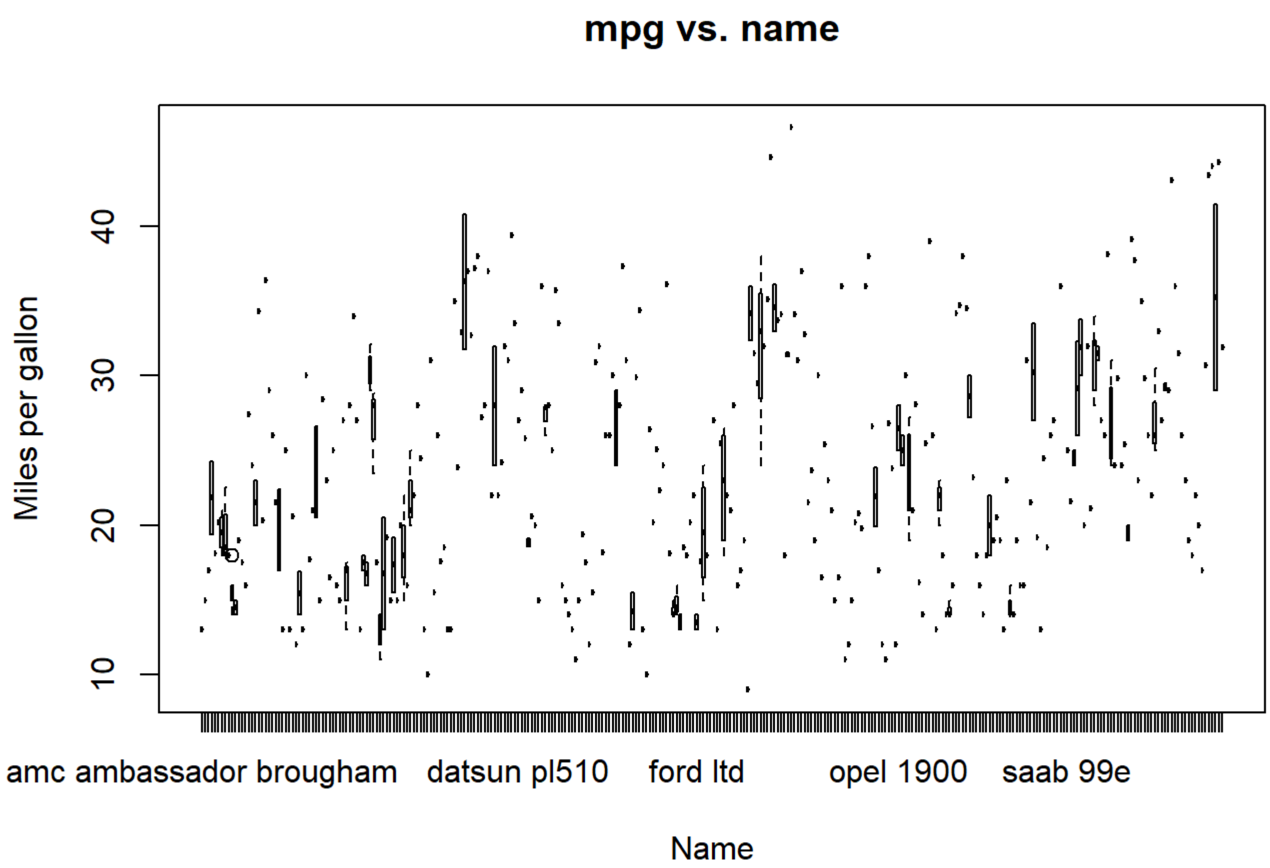
The average fuel effiency of a car increases linearly as time goes by.

```
origin = as.factor(Auto$origin)
plot(x=origin, y=Auto$mpg, xlab='Origin', ylab='Miles per gallon', main='mpg vs. origin'
)
```



I don't understand what origin is supposed to be, but fuel efficiency seems to increase as origin goes from 1 to 2 to 3.

```
plot(x=Auto$name, y=Auto$mpg, xlab='Name', ylab='Miles per gallon', main='mpg vs. name')
```



There are too many names to properly sort through them all, but some brands seem to be performing much better than others in terms of fuel efficiency. (f) Suppose that we wish to predict gas mileage (**mpg**) on the basis of the other variables. Do your plots suggest that any of the other variables might be useful in predicting **mpg**? Justify your answer. *From the above plots, you can see that the number of cylinders, displacement, horsepower, weight, year, and origin of a car are all useful predictors for determining that cars mpg.*

Exercise 10

This exercise involves the **Boston** housing data set.

a. To begin, load in the **Boston** data set. The **Boston** data set is part of the **MASS** *library* in R.

```
library(MASS)
```

Now the data set is contained in the object **Boston**.

```
?Boston
```

```
## starting httpd help server ... done
```

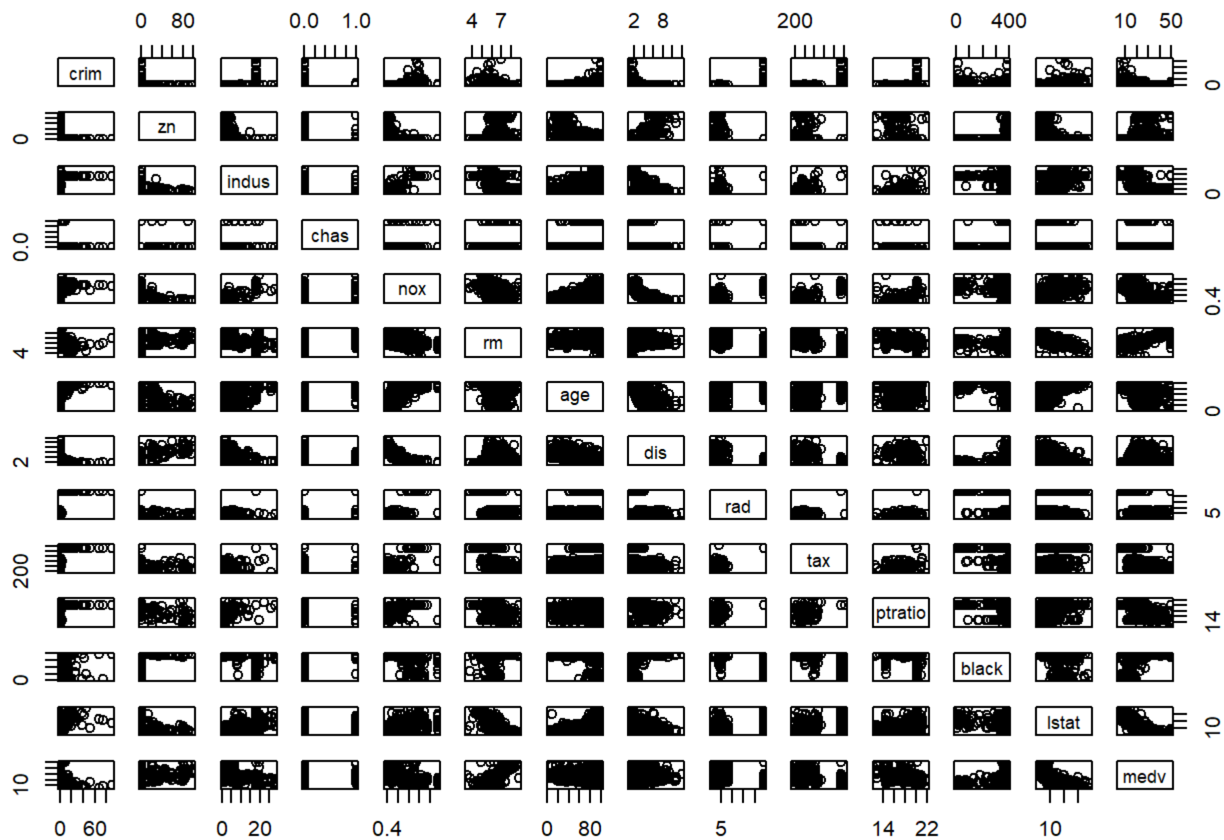
How many rows are in this data set? How many columns? What do the rows and columns represent?

```
dim(Boston)
```

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

There are 506 observations and 14 predictors in each observation. (b) Make some pairwise scatterplots of the predictors (columns) in this data set. Describe your findings.

```
pairs(Boston)
```



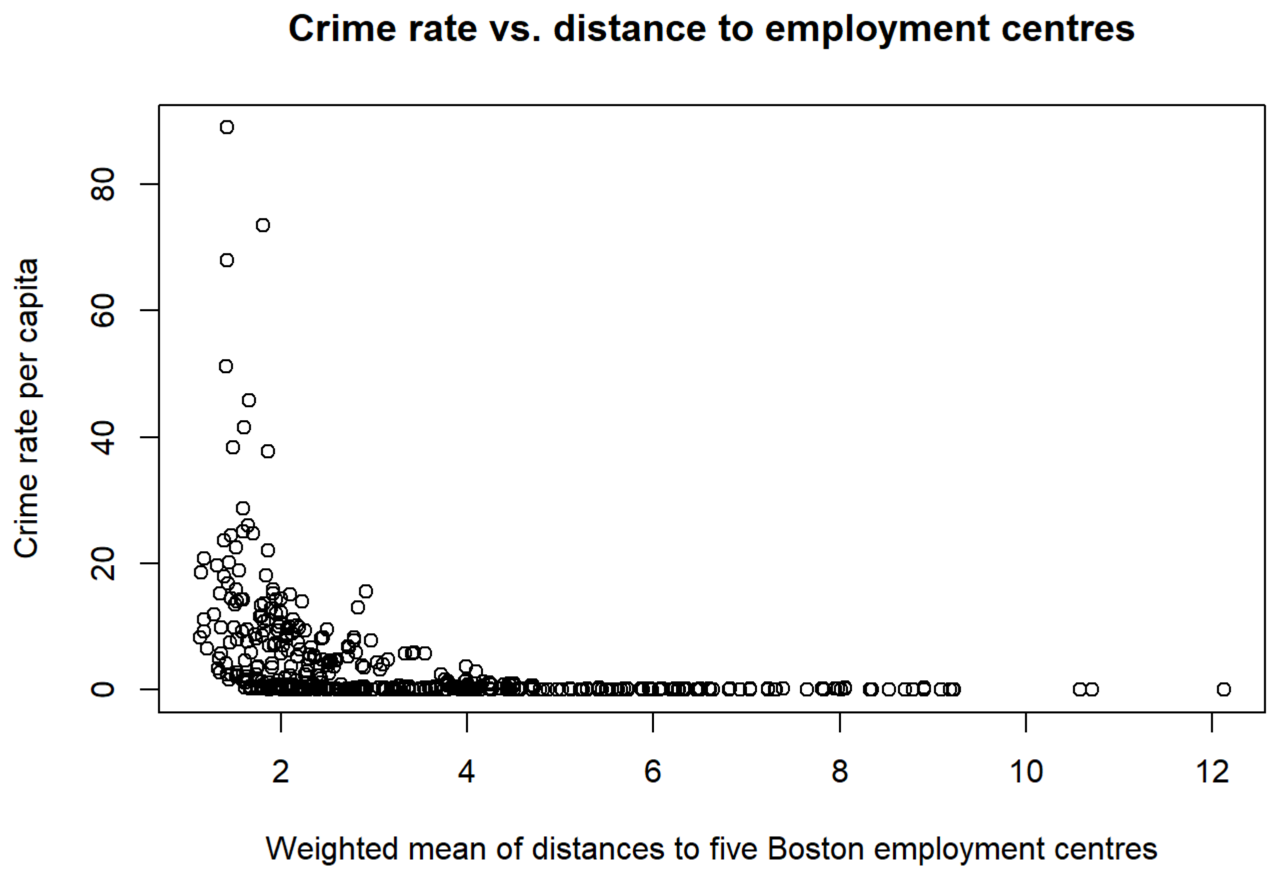
(c)

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

```
plot(y=Boston$crim, x=Boston$zn, ylab='per capita crime rate', xlab='proportion of resid
ential land zoned for lots over 25,000 sq.ft.', main='Crime rate vs. residential zoning'
)
```

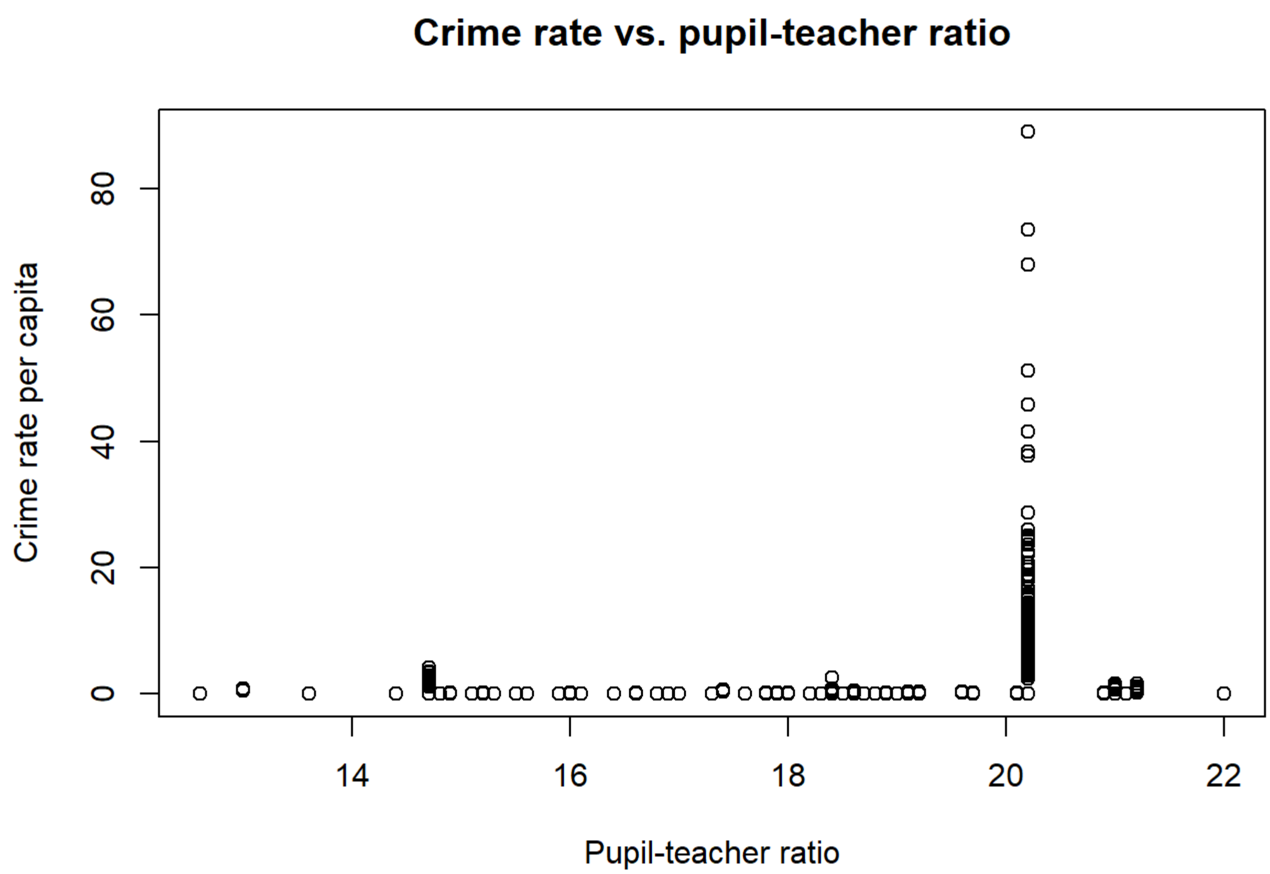
Almost of all the crime rate per capita is in towns with no proportion of residential land zoned for lots over 25,000 square feet.

```
plot(x=Boston$dis, y=Boston$crim, xlab='Weighted mean of distances to five Boston employment centres', ylab='Crime rate per capita', main='Crime rate vs. distance to employment centres')
```



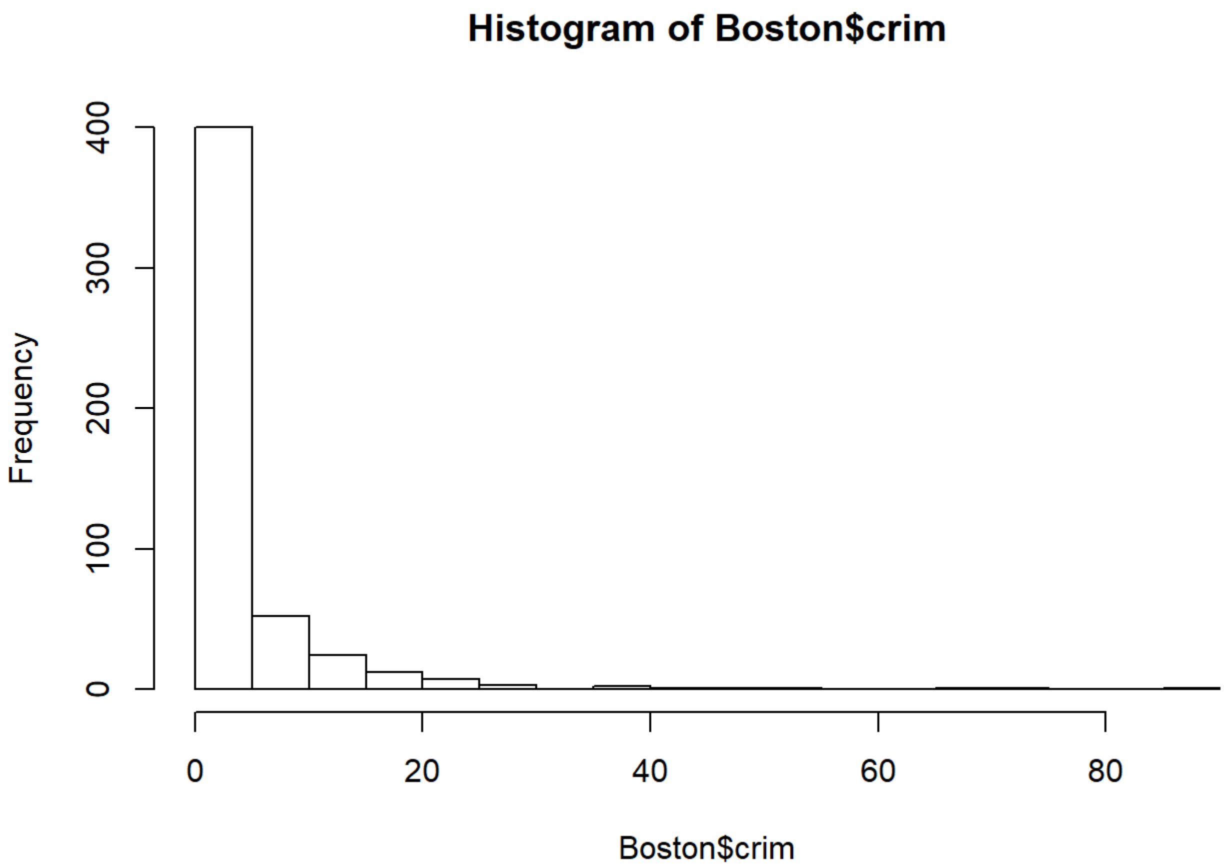
Much higher crime rate closer to the employment centres. The crime rate drastically decreases beyond 3-ish (miles maybe).

```
plot(x=Boston$ptratio, y=Boston$crim, xlab='Pupil-teacher ratio', ylab='Crime rate per c
apita', main='Crime rate vs. pupil-teacher ratio')
```



Almost all of the towns with a significant crime rate have a pupil-teacher ratio of just over 20 students to each teacher. Although, there is also a suprising fluctuation in crime rate at just under 15:1 pupil-teacher ratio. (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.

```
hist(Boston$crim, breaks=25)
```




```
cols = c('crim', 'tax', 'ptratio')
Boston[Boston$crim > 30, cols]
```

```
##        crim tax ptratio
## 381 88.9762 666    20.2
## 399 38.3518 666    20.2
## 405 41.5292 666    20.2
## 406 67.9208 666    20.2
## 411 51.1358 666    20.2
## 415 45.7461 666    20.2
## 419 73.5341 666    20.2
## 428 37.6619 666    20.2
```

All of the towns in this data set with crime rates of over 30 per capita a full-value property-tax rate of 666$10,000 and a pupil-teacher ratio of 20.2:1.

```
for (c in cols) {
  x = range(Boston[[c]])
  print(sprintf('%s range: %.2f - %.2f', c, x[1], x[2]))
}
```

```
## [1] "crim range: 0.01 - 88.98"
## [1] "tax range: 187.00 - 711.00"
## [1] "ptratio range: 12.60 - 22.00"
```

The values of crime rate, property-tax rate, and pupil-teacher ratio are all near their highest within the data set. (e) How many of the suburbs in this data set bound the Charles river?

```
summary(as.factor(Boston$chas))
```

```
##   0   1
## 471  35
```

35 are bound by the Charles river. 471 are not. (f) What is the median pupil-teacher ratio among the towns in this data set?

```
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.

```
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
```

```
summary(Boston)
```

```
##       crim                zn             indus            chas        
##  Min.   : 0.00632   Min.   :  0.00   Min.   : 0.46   Min.   :0.00000  
##  1st Qu.: 0.08204   1st Qu.:  0.00   1st Qu.: 5.19   1st Qu.:0.00000  
##  Median : 0.25651   Median :  0.00   Median : 9.69   Median :0.00000  
##  Mean   : 3.61352   Mean   : 11.36   Mean   :11.14   Mean   :0.06917  
##  3rd Qu.: 3.67708   3rd Qu.: 12.50   3rd Qu.:18.10   3rd Qu.:0.00000  
##  Max.   :88.97620   Max.   :100.00   Max.   :27.74   Max.   :1.00000  
##       nox               rm             age              dis        
##  Min.   :0.3850   Min.   :3.561   Min.   :  2.90   Min.   : 1.130  
##  1st Qu.:0.4490   1st Qu.:5.886   1st Qu.: 45.02   1st Qu.: 2.100  
##  Median :0.5380   Median :6.208   Median : 77.50   Median : 3.207  
##  Mean   :0.5547   Mean   :6.285   Mean   : 68.57   Mean   : 3.795  
##  3rd Qu.:0.6240   3rd Qu.:6.623   3rd Qu.: 94.08   3rd Qu.: 5.188  
##  Max.   :0.8710   Max.   :8.780   Max.   :100.00   Max.   :12.127  
##       rad              tax           ptratio          black       
##  Min.   : 1.000   Min.   :187.0   Min.   :12.60   Min.   :  0.32  
##  1st Qu.: 4.000   1st Qu.:279.0   1st Qu.:17.40   1st Qu.:375.38  
##  Median : 5.000   Median :330.0   Median :19.05   Median :391.44  
##  Mean   : 9.549   Mean   :408.2   Mean   :18.46   Mean   :356.67  
##  3rd Qu.:24.000   3rd Qu.:666.0   3rd Qu.:20.20   3rd Qu.:396.23  
##  Max.   :24.000   Max.   :711.0   Max.   :22.00   Max.   :396.90  
##      lstat            medv      
##  Min.   : 1.73   Min.   : 5.00  
##  1st Qu.: 6.95   1st Qu.:17.02  
##  Median :11.36   Median :21.20  
##  Mean   :12.65   Mean   :22.53  
##  3rd Qu.:16.95   3rd Qu.:25.00  
##  Max.   :37.97   Max.   :50.00
```

The town with the lowest median value of owner-occupied homes is one of the 8 towns that has a crime rate over 30. Also, all of the owner-occupied units in the town were built prior to 1940, and it has the highest index of accessibility to radial highways, one of the smallest weighted mean distances to five Boston employment centres, and the higheset proportion of blacks. It is definitely part of the old city and centrally located within Boston. (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.

```
sum(Boston$rm > 7)
```

```
## [1] 64
```

There are 64 towns with more than 7 rooms per dwelling on average.

```
manyRooms = Boston[Boston$rm > 8,]
fewRooms  = Boston[Boston$rm < 8,]
nrow(manyRooms)
```

```
## [1] 13
```

There are 13 towns with more than 8 rooms per dwelling on average.

```
summary(manyRooms)
```

```
##       crim               zn            indus             chas       
##  Min.   :0.02009   Min.   : 0.00   Min.   : 2.680   Min.   :0.0000  
##  1st Qu.:0.33147   1st Qu.: 0.00   1st Qu.: 3.970   1st Qu.:0.0000  
##  Median :0.52014   Median : 0.00   Median : 6.200   Median :0.0000  
##  Mean   :0.71879   Mean   :13.62   Mean   : 7.078   Mean   :0.1538  
##  3rd Qu.:0.57834   3rd Qu.:20.00   3rd Qu.: 6.200   3rd Qu.:0.0000  
##  Max.   :3.47428   Max.   :95.00   Max.   :19.580   Max.   :1.0000  
##       nox               rm             age              dis       
##  Min.   :0.4161   Min.   :8.034   Min.   : 8.40   Min.   :1.801  
##  1st Qu.:0.5040   1st Qu.:8.247   1st Qu.:70.40   1st Qu.:2.288  
##  Median :0.5070   Median :8.297   Median :78.30   Median :2.894  
##  Mean   :0.5392   Mean   :8.349   Mean   :71.54   Mean   :3.430  
##  3rd Qu.:0.6050   3rd Qu.:8.398   3rd Qu.:86.50   3rd Qu.:3.652  
##  Max.   :0.7180   Max.   :8.780   Max.   :93.90   Max.   :8.907  
##       rad              tax           ptratio          black      
##  Min.   : 2.000   Min.   :224.0   Min.   :13.00   Min.   :354.6  
##  1st Qu.: 5.000   1st Qu.:264.0   1st Qu.:14.70   1st Qu.:384.5  
##  Median : 7.000   Median :307.0   Median :17.40   Median :386.9  
##  Mean   : 7.462   Mean   :325.1   Mean   :16.36   Mean   :385.2  
##  3rd Qu.: 8.000   3rd Qu.:307.0   3rd Qu.:17.40   3rd Qu.:389.7  
##  Max.   :24.000   Max.   :666.0   Max.   :20.20   Max.   :396.9  
##      lstat           medv     
##  Min.   :2.47   Min.   :21.9  
##  1st Qu.:3.32   1st Qu.:41.7  
##  Median :4.14   Median :48.3  
##  Mean   :4.31   Mean   :44.2  
##  3rd Qu.:5.12   3rd Qu.:50.0  
##  Max.   :7.44   Max.   :50.0
```

```
summary(fewRooms)
```

```
##       crim                zn             indus            chas        
##  Min.   : 0.00632   Min.   :  0.0   Min.   : 0.46   Min.   :0.00000  
##  1st Qu.: 0.08014   1st Qu.:  0.0   1st Qu.: 5.19   1st Qu.:0.00000  
##  Median : 0.24522   Median :  0.0   Median : 9.69   Median :0.00000  
##  Mean   : 3.68986   Mean   : 11.3   Mean   :11.24   Mean   :0.06694  
##  3rd Qu.: 3.77498   3rd Qu.: 12.5   3rd Qu.:18.10   3rd Qu.:0.00000  
##  Max.   :88.97620   Max.   :100.0   Max.   :27.74   Max.   :1.00000  
##       nox               rm             age              dis        
##  Min.   :0.3850   Min.   :3.561   Min.   :  2.9   Min.   : 1.130  
##  1st Qu.:0.4490   1st Qu.:5.879   1st Qu.: 44.4   1st Qu.: 2.088  
##  Median :0.5380   Median :6.185   Median : 77.3   Median : 3.216  
##  Mean   :0.5551   Mean   :6.230   Mean   : 68.5   Mean   : 3.805  
##  3rd Qu.:0.6240   3rd Qu.:6.575   3rd Qu.: 94.3   3rd Qu.: 5.215  
##  Max.   :0.8710   Max.   :7.929   Max.   :100.0   Max.   :12.127  
##       rad              tax           ptratio          black       
##  Min.   : 1.000   Min.   :187.0   Min.   :12.60   Min.   :  0.32  
##  1st Qu.: 4.000   1st Qu.:280.0   1st Qu.:17.40   1st Qu.:374.71  
##  Median : 5.000   Median :334.0   Median :19.10   Median :391.83  
##  Mean   : 9.604   Mean   :410.4   Mean   :18.51   Mean   :355.92  
##  3rd Qu.:24.000   3rd Qu.:666.0   3rd Qu.:20.20   3rd Qu.:396.24  
##  Max.   :24.000   Max.   :711.0   Max.   :22.00   Max.   :396.90  
##      lstat            medv      
##  Min.   : 1.73   Min.   : 5.00  
##  1st Qu.: 7.34   1st Qu.:16.70  
##  Median :11.65   Median :21.00  
##  Mean   :12.87   Mean   :21.96  
##  3rd Qu.:17.11   3rd Qu.:24.80  
##  Max.   :37.97   Max.   :50.00  
```

In the towns with more than 8 average rooms per dwelling, there are smaller crime rates, less industrial zoning, a smaller lower status of the population, and a higher median value of owner-occupied homes.