Chapter 2 Problem 9

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

```
#Remove missing values from data
>auto= read.csv("Auto.csv",header=T,na.strings = "?")
>auto=na.omit(auto)

#Summary of auto data
>summary(auto)
```

```
      mpg          cylinders      displacement     horsepower        weight      acceleration
 Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0   Min.   :1613   Min.   : 8.00
 1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0   1st Qu.:2225   1st Qu.:13.78
 Median :22.75   Median :4.000   Median :151.0   Median : 93.5   Median :2804   Median :15.50
 Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5   Mean   :2978   Mean   :15.54
 3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0   3rd Qu.:3615   3rd Qu.:17.02
 Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0   Max.   :5140   Max.   :24.80

      year           origin                      name
 Min.   :70.00   Min.   :1.000   amc matador       :  5
 1st Qu.:73.00   1st Qu.:1.000   ford pinto        :  5
 Median :76.00   Median :1.000   toyota corolla    :  5
 Mean   :75.98   Mean   :1.577   amc gremlin       :  4
 3rd Qu.:79.00   3rd Qu.:2.000   amc hornet        :  4
 Max.   :82.00   Max.   :3.000   chevrolet chevette:  4
                                 (Other)           :365
```

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

All predictors except *name* and *origin* are quantitative.

b. What is the range of each quantitative predictor?

```
>range(auto$mpg)
[1]  9.0 46.6
```

Range mpg = 46.6 – 9.0 = 37.6

cylinders: 5
displacement: 387
horsepower: 184
weight: 3527
acceleration: 16.8
year: 12
origin: 2

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

Refer to the top table for mean values. We use `> sapply(auto[,1:7], sd)` to calculate standard deviations.

mpg: 7.805
cylinders: 1.706
displacement: 104.644
horsepower: 38.491
weight: 849.403
acceleration: 2.759
year: 3.684
origin: 0.806

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?

```
#Delete 10th to 85th observations
>auto <- auto[-(10:85),]
>summary(auto) #gives mean values; range of predictor X = Max_x - Min_x
```

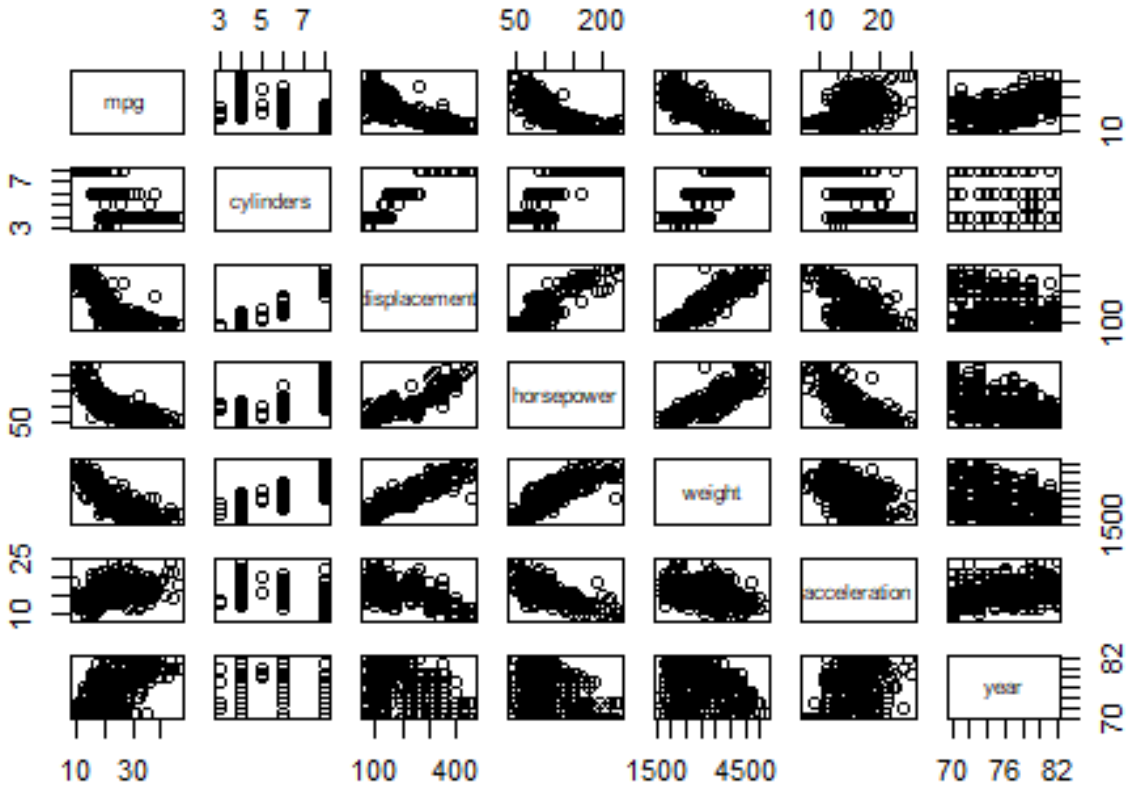
```
      mpg          cylinders      displacement     horsepower        weight      acceleration
 Min.   :11.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0   Min.   :1649   Min.   : 8.50
 1st Qu.:18.00   1st Qu.:4.000   1st Qu.:100.2   1st Qu.: 75.0   1st Qu.:2214   1st Qu.:14.00
 Median :23.95   Median :4.000   Median :145.5   Median : 90.0   Median :2792   Median :15.50
 Mean   :24.40   Mean   :5.373   Mean   :187.2   Mean   :100.7   Mean   :2936   Mean   :15.73
 3rd Qu.:30.55   3rd Qu.:6.000   3rd Qu.:250.0   3rd Qu.:115.0   3rd Qu.:3508   3rd Qu.:17.30
 Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0   Max.   :4997   Max.   :24.80

      year           origin                      name
 Min.   :70.00   Min.   :1.000   ford pinto           :  5
 1st Qu.:75.00   1st Qu.:1.000   toyota corolla       :  5
 Median :77.00   Median :1.000   amc matador          :  4
 Mean   :77.15   Mean   :1.601   chevrolet chevette   :  4
 3rd Qu.:80.00   3rd Qu.:2.000   amc hornet           :  3
 Max.   :82.00   Max.   :3.000   chevrolet caprice classic:  3
```

```
>sapply(auto[,1:7], sd) #standard deviations
```

```
       mpg   cylinders displacement  horsepower      weight acceleration        year
  7.867283    1.654179   99.678367   35.708853  811.300208    2.693721    3.106217
```

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

```
#Just getting the full data back :)
>auto= read.csv("Auto.csv",header=T,na.strings = "?")
>auto=na.omit(auto)
>pairs(auto[,1:7]) #scatterplot matrix
```



Positive correlations: mpg with years

Negative correlations: mpg with displacement, horsepower, weight

f. Suppose that we wish to predict the gas mileage based on other variables. Do your plots suggest that any of the other variables might be useful in predicting mpg? Justify your answer.

Yes, we were able to see relationships between mpg and other predictors (see above).