2.4 Exercises Problem 9

Section 23 - Group 6 Project Groups (Bosan Hsu, Fan Liu, Jimeng Yin, Michael Liu, Richard Wang, Zhuoqian Zhang

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

```
library(ISLR)
attach(Auto)
Auto = na.omit(Auto)
View(Auto)
summary(Auto)
```

```
##       mpg          cylinders      displacement     horsepower
 weight
##  Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0   Min.
   :1613
##  1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0   1st
 Qu.:2225
##  Median :22.75   Median :4.000   Median :151.0   Median : 93.5   Med
ian :2804
##  Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5   Mea
n   :2978
##  3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0   3rd
 Qu.:3615
##  Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0   Max.
   :5140
##
```

```
##   acceleration        year           origin                      nam
e
##  Min.   : 8.00   Min.   :70.00   Min.   :1.000   amc matador       :
  5
##  1st Qu.:13.78   1st Qu.:73.00   1st Qu.:1.000   ford pinto        :
  5
##  Median :15.50   Median :76.00   Median :1.000   toyota corolla    :
  5
##  Mean   :15.54   Mean   :75.98   Mean   :1.577   amc gremlin       :
  4
##  3rd Qu.:17.02   3rd Qu.:79.00   3rd Qu.:2.000   amc hornet        :
  4
##  Max.   :24.80   Max.   :82.00   Max.   :3.000   chevrolet chevette:
  4
##                                                  (Other)           :
365
```

```
names(Auto)
```

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

dim(Auto)

## [1] 392   9
```

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

```
sapply(Auto, class)

##          mpg    cylinders displacement   horsepower       weight acc
eleration
##    "numeric"    "numeric"    "numeric"    "numeric"    "numeric"
"numeric"
##         year       origin         name
##    "numeric"    "numeric"     "factor"
```

quantitative predictor: mpg, displacement, horsepower, weight, acceleration
qualitative predictor: cylinders, year, origin, name

```
quantitative_predictor = Auto[ ,c("mpg", "displacement", "horsepower",
"weight", "acceleration")]
qualitative_predictor = Auto[ ,c("cylinders", "year", "origin", "name")]
```

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

```
sapply(quantitative_predictor, range)

##       mpg displacement horsepower weight acceleration
## [1,]  9.0           68         46   1613          8.0
## [2,] 46.6          455        230   5140         24.8
```

The range of each quantitative predictor: mpg: [9.0, 46.6]; displacement: [68, 455]; horsepower: [46, 230]; weight: [1613, 5140]; acceleration: [8.0, 24.8]

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

```
sapply(quantitative_predictor, mean)

##          mpg displacement   horsepower       weight acceleration
##     23.44592    194.41199    104.46939   2977.58418     15.54133

sapply(quantitative_predictor, sd)

##          mpg displacement   horsepower       weight acceleration
##     7.805007   104.644004    38.491160   849.402560     2.758864
```

The mean of each quantitative predictor: mpg: 23.44592; displacement: 194.41199; horsepower: 104.46939; weight: 2977.58418; acceleration: 15.54133 The standard deviation of each quantitative predictor: mpg: 7.805007; displacement: 104.644004; horsepower: 38.491160; weight: 849.402560; acceleration: 2.758864

(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?

```
remove_Auto = quantitative_predictor[-c(10:85), ]
sapply(remove_Auto, range)
```

```
##       mpg displacement horsepower weight acceleration
## [1,] 11.0           68         46   1649          8.5
## [2,] 46.6          455        230   4997         24.8
```

```
sapply(remove_Auto, mean)
```

```
##          mpg displacement   horsepower       weight acceleration
##     24.40443    187.24051    100.72152   2935.97152     15.72690
```

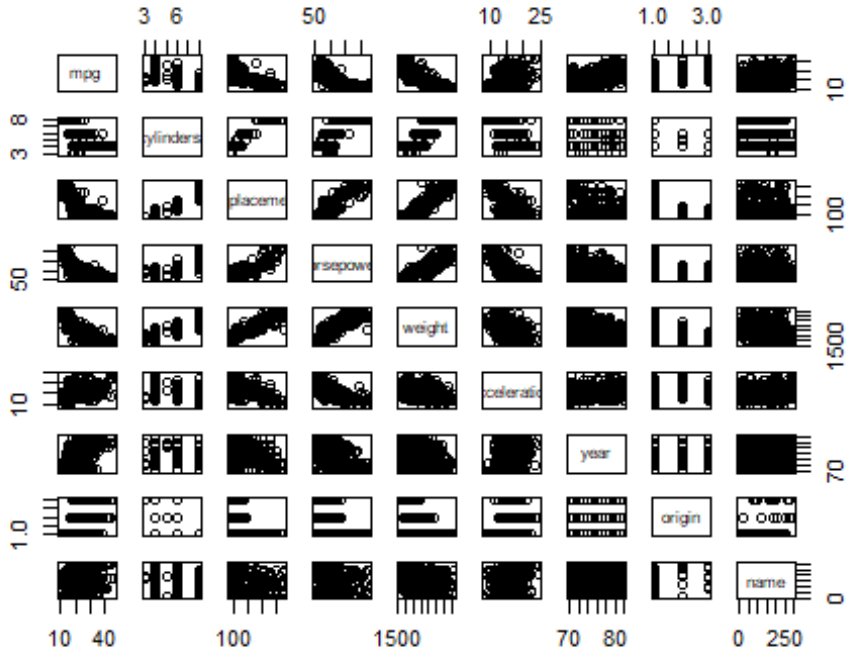
```
sapply(remove_Auto, sd)
```

```
##          mpg displacement   horsepower       weight acceleration
##     7.867283    99.678367    35.708853   811.300208     2.693721
```

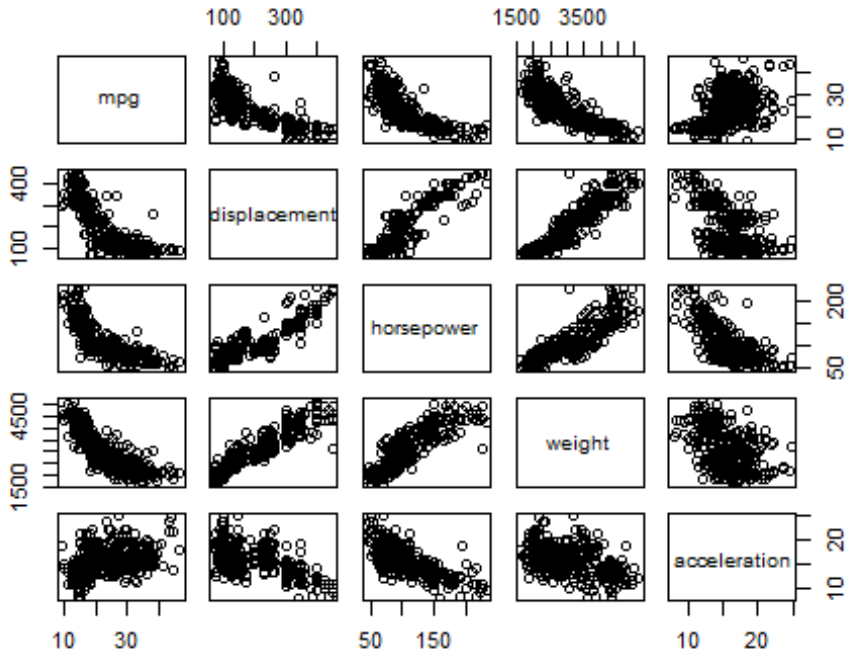
The range of each quantitative predictor: mpg: [11.0, 46.6]; displacement: [68, 455]; horsepower: [46, 230]; weight: [1649, 4997]; acceleration: [8.5, 24.8] The mean of each quantitative predictor: mpg: 24.40443; displacement: 187.24051; horsepower: 100.72152; weight: 2935.97152; acceleration: 15.72690 The standard deviation of each quantitative predictor: mpg: 7.867283; displacement: 99.678367; horsepower: 35.708853; weight: 811.300208; acceleration: 2.693721

(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.

```
pairs(Auto)
```




```
pairs(~mpg + displacement + horsepower + weight + acceleration, Auto)
```



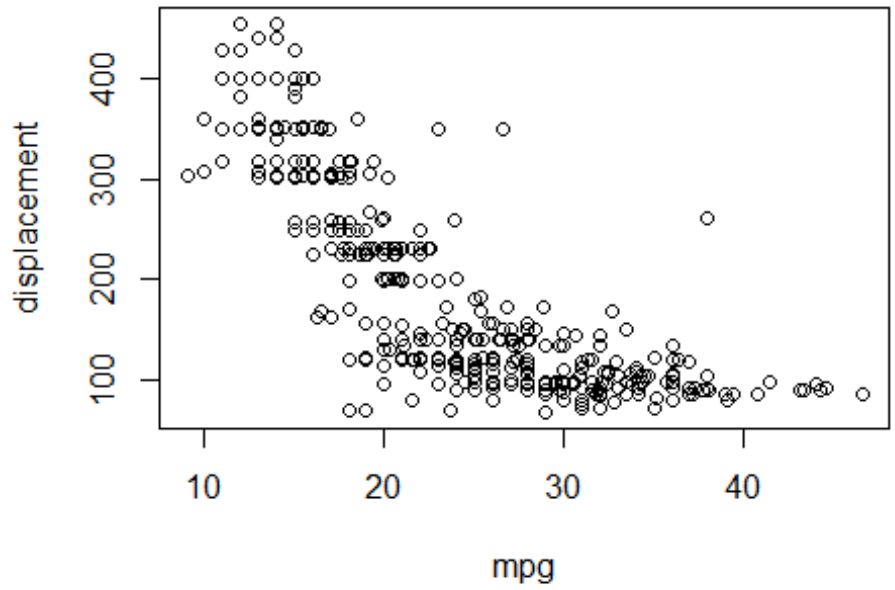
findings:

mpg, displacement, horsepower, weight, and acceleration have transparent linear

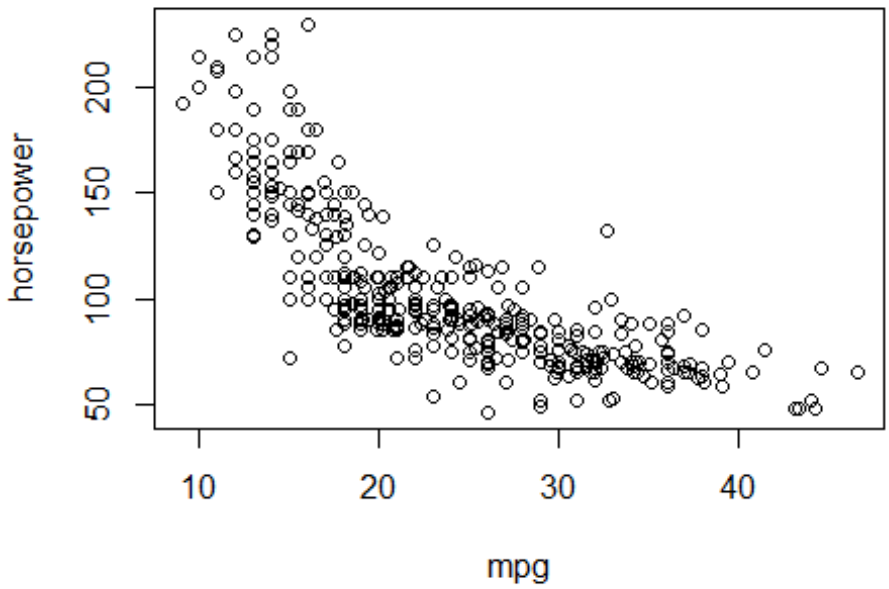
relationships. mpg has negatively correlated with displacement, horsepower, and weight; while positively correlated with acceleration.

(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.

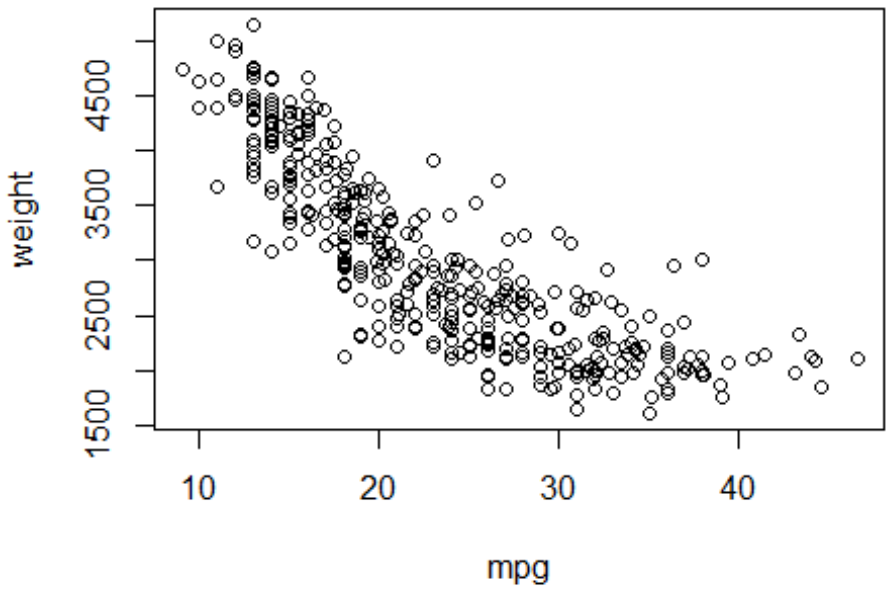
```
plot(mpg, displacement)
```



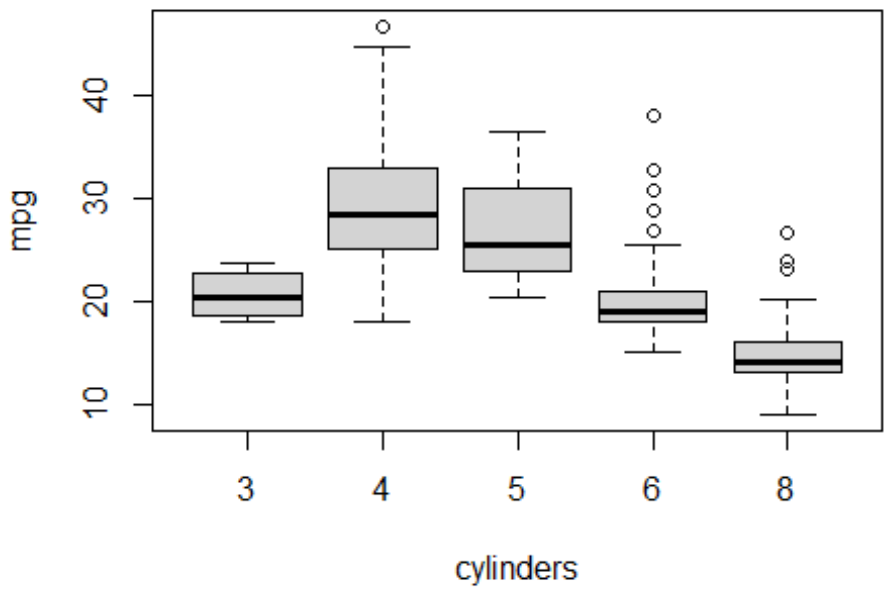

```
plot(mpg, horsepower)
```



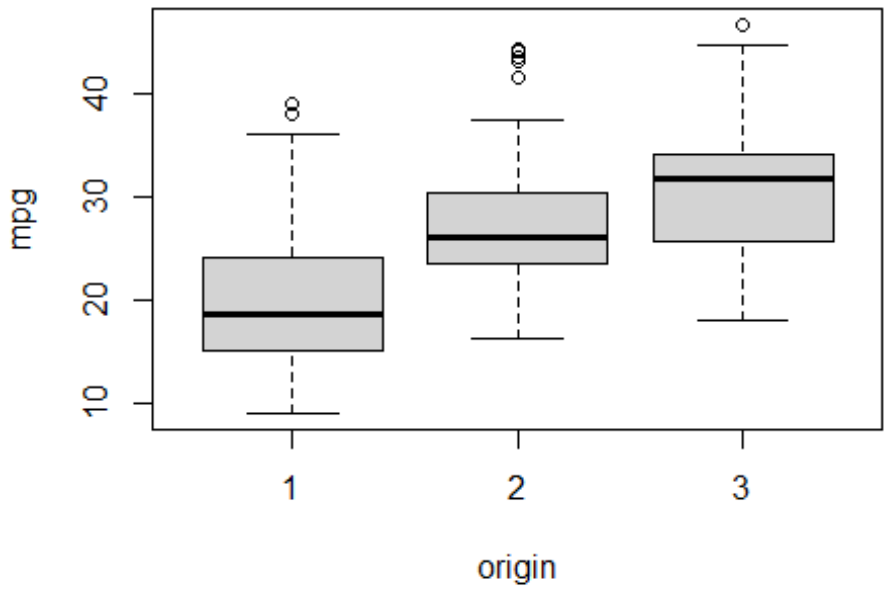

```r
plot(mpg, weight)
```




```r
boxplot(mpg~cylinders)
```




```
boxplot(mpg~origin)
```



findings:

mpg has negatively correlated with displacement, horsepower, and weight. origin 3 has the best mpg, while origin 1 has the worst the mpg; we think origin 1 is the US,

origin 2 is Europe, origin 3 is Japan. cars with four cylinders have the best mpg, cars with eight cylinders have the worst mpg.