Homework 1

Samikshya

2023-01-09

Question 2

a. It is a regression problem because the CEO's salary which is response variable i.e the y in this problem is a numeric value. It is also a inference problem because we are interested in understanding the factors that affect the CEO's salary. For this, we need to understand the functional form and how each input variables affect the CEO's salary.

b. It is a classification problem because we are interested in categorical output (success or failure). It is a prediction problem, because we are more interested in whether the model yields a accurate prediction of Y rather than understanding the exact form the model and how each input is related to the output variable.

c. It is a regression problem because the output/response variable is a numerical value as it is a % change. It is also a prediction problem because we are more interested in getting the accurate value for y than on understanding how each of the input variable (% change in US market, British market etc) affect the % change in exchange rate.

Question 4

a. Classification/ inference or prediction

   1. Diabetes diagnosis: It is a classification problem because the response variable is whether one has a diabetes diagnosis or not. Some of the potential predictors could be: genetic disposition, diet, overall health etc. It is also a prediction problem because we are more interested in the accuracy of the response rather than on how each of the predictors could affect the output.
   2. Academic grades' impact on job search. The output is whether a person gets a job offer or not. The predictors could be:grades, previous internship experience, relevant coursework, public speaking etc. It is a inference problem because we are interested in understanding how each of these predictors specifically academic grades affects one's chance of getting a job.
   3. Has grocery bills increased by more than 10% compared to last year. It is a classification problem with response variable as : increased or not increased. Some predictors could be: inflation, supply chain disruption, yield level etc. It is a prediction problem because we are more interested in the output than understanding how each of these input variables could have affected the output. We are not focused on understanding the functional form of the model.

   b. Regression/inference or prediction

   1. Impact of recession on consumer habits. Changes in consumer spending because of threat of recession. It is a regression because we are quantifying the consumer spending. Some potential predictors could be: inflation rate, increase in salary, percieved job security etc. It is a prediction model because the question is framed such as we are more interested in how has the consumer spending changed than on understanding how each factor affect the consumer spending.

2. Percentage growth in COVID cases after holidays. The y variable is % increase in COVID cases. The input variables could be travel rate, vaccination rate etc. It is a inference problem because we want to understand the impact of holiday on COVID cases so we are interested in understanding the functional form and how each factor affect y.
3. How has graduation rate changed over time. It is a regression problem because our output variable (graduation rate) is measured quantitatively. Some predictors could be information access, parents were graduates, income level etc. It is a prediction model because we are more interested in how has the graduation rate changed over time rather than on factors affecting the change.

c. Cluster

1. Grouping eating habits of certain birds together.
2. Items brought together by an average consumer in a supermarket.
3. TV shows watched by different age groups

Question 7

a) Calculating the Euclidean distance between each observation and the test point, obs 1: sqrt ((0-0)^2 + (0-3)^2 + (0- 0)^2) = 3 obs 2: sqrt ((2-0)^2 + (0-)^2 + (0- 0)^2) = 2 obs 3: sqrt ((0-0)^2 + (1-0)^2 + (3- 0)^2) = sqrt 10 = 3.16 obs 4: sqrt ((0-0)^2 + (0-1)^2 + (2- 0)^2) = sqrt 5 =2.24 obs 5: sqrt ((-1-0)^2 + (0-0)^2 + (1- 0)^2) = sqrt 2 = 1.414 obs 6: sqrt ((1-0)^2 + (1-0)^2 + (1- 0)^2) = sqrt 3 = 1.73

b) With k = 1, our prediction is that it will be observation 5 i.e our classification would be that the test point would fall in class Green; smallest distance between test point and each observation is for obs green.

c) Based on sqaure distance, the 3 smallest distance are for observation are obs 2, obs 5 and obs 6. There class are Red, Green and Red respectively. Therefore, with k = 3, it would fall in Red class.

d) If the problem is highly non-linear than the K value should be small so that the model can adopt to a more flexible model.

Question 8

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(dplyr)
library(ISLR2)
```

a)

```r
#Loading the dataset
#college <- read.csv('/Users/samik/Downloads/College.csv')
#View
#View(college)
```

```r
#Loading the dataset
college <- College
view(college)
```

b)

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

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

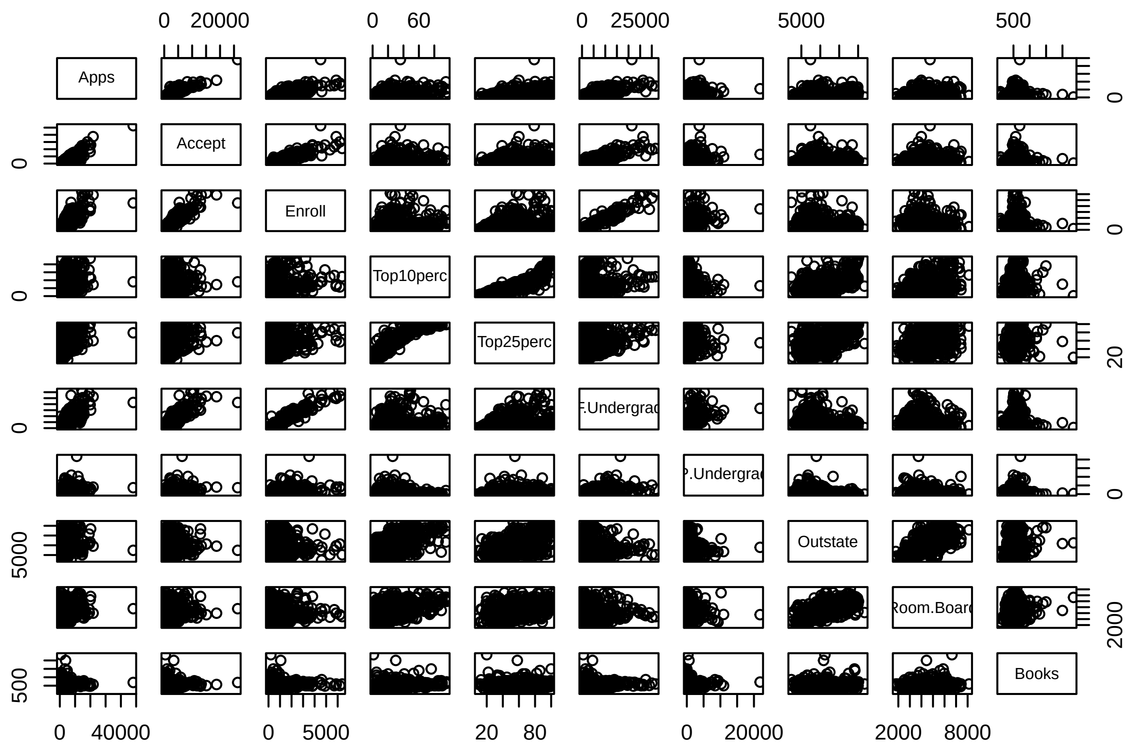
c)

```r
summary(college)
```

```
##       Apps           Accept          Enroll       Top10perc       Top25perc    
##  Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00   Min.   :  9.0  
##  1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00   1st Qu.: 41.0  
##  Median : 1558   Median : 1110   Median : 434   Median :23.00   Median : 54.0  
##  Mean   : 3002   Mean   : 2019   Mean   : 780   Mean   :27.56   Mean   : 55.8  
##  3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.00   3rd Qu.: 69.0  
##  Max.   :48094   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
```

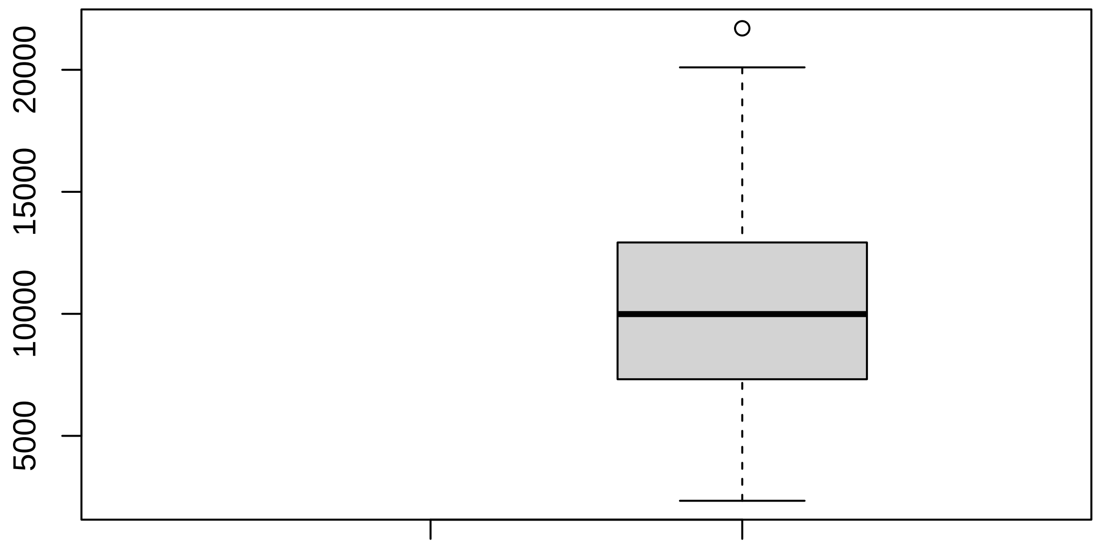
The numerical summary for each of the varialbes are listed above.

```r
col_A10 <- college[,1:10]   #choosing the first 10 columns
#col_A10
pairs(col_A10)
```




```r
#ggplot(data = college, aes(x = Outstate , y = Private)) +
  #geom_boxplot()
```

```r
boxplot(college$Private, college$Outstate, xlim = c(0,3))
```



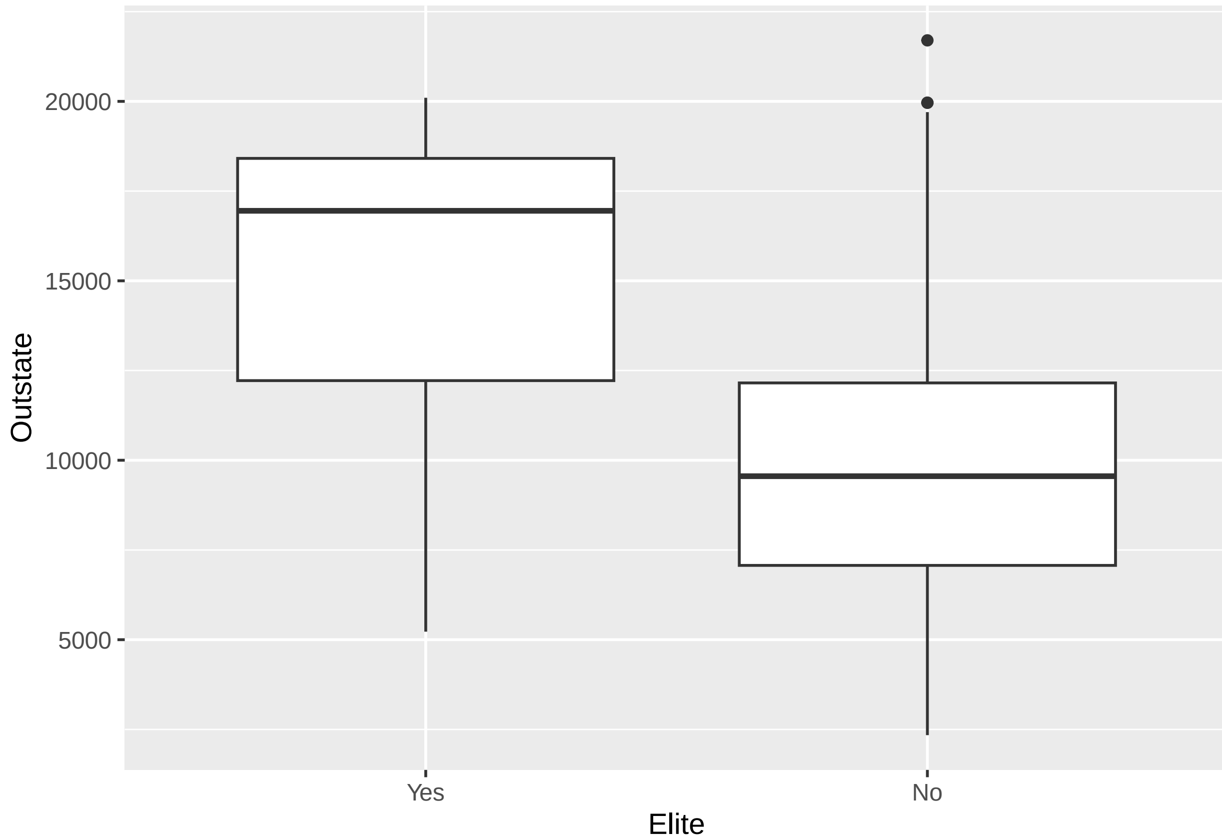

```r
Elite <- rep ("No", nrow (college))
Elite[college$Top10perc > 50] <- " Yes "
Elite <- as.factor(Elite)
college <- data.frame (college , Elite)
```

```r
summary(college$Elite)
```

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

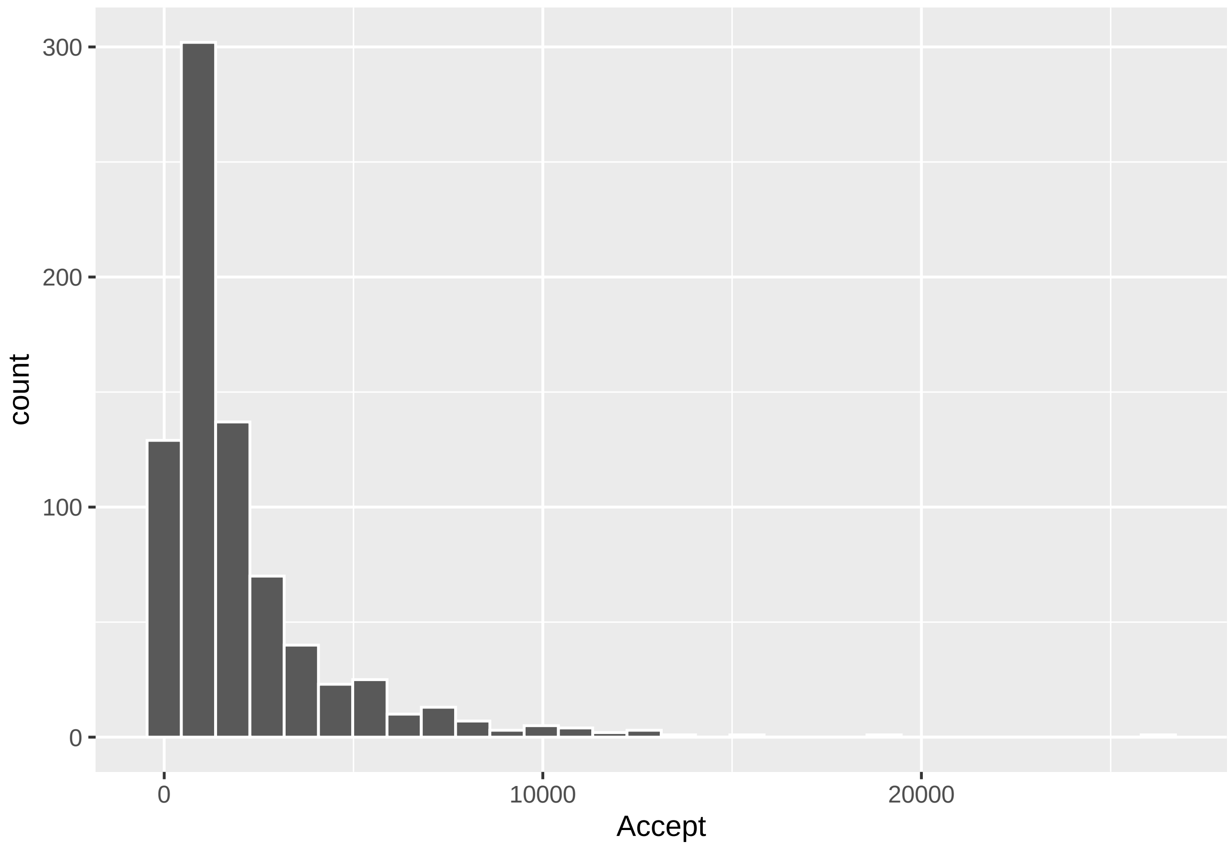
We can see that there are 78 elite colleges and 699 not elite universities.

```r
#boxplot(college$Outstate, college$Elite)

ggplot(data = college, aes(x = Elite , y = Outstate)) +
  geom_boxplot()
```




```
ggplot(data = college, mapping = aes(x = Accept)) +
  geom_histogram(color="white")
```

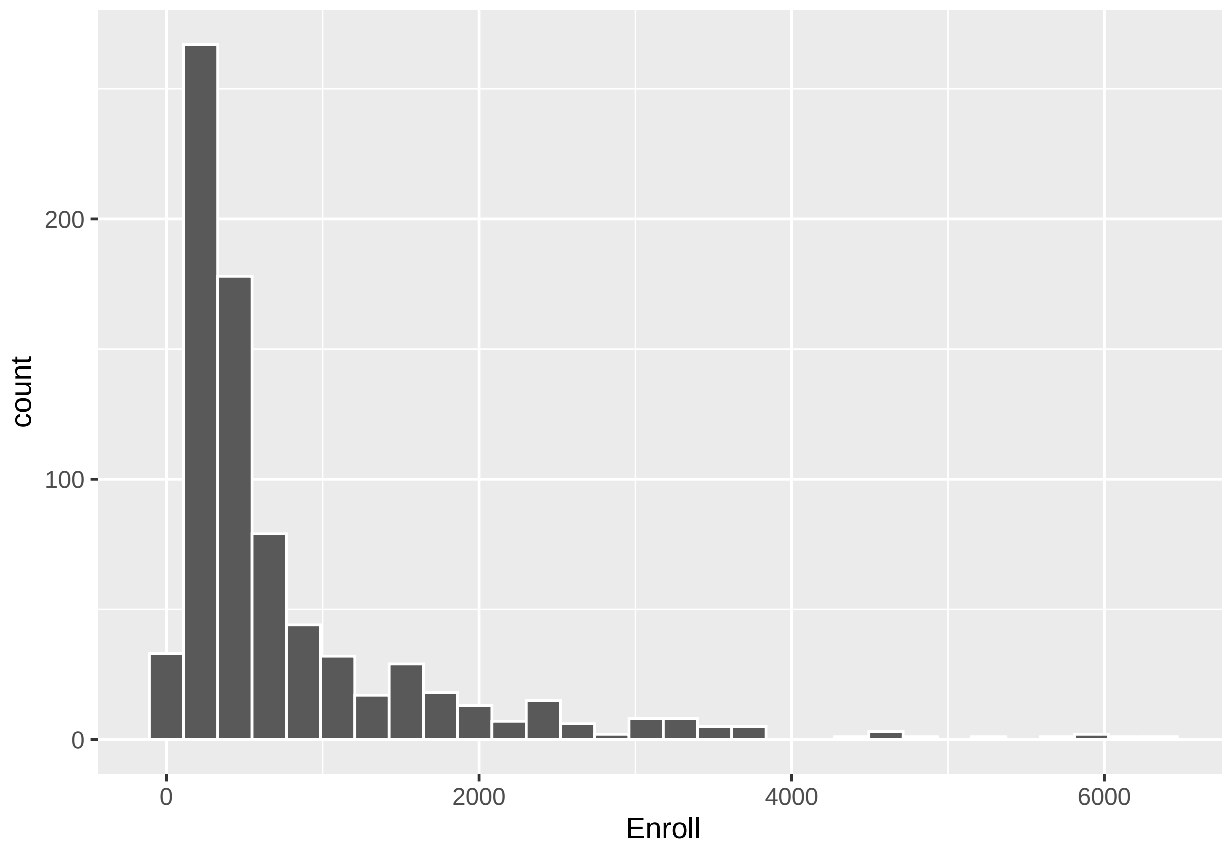
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The accept tends to follow a normal distribution as well.

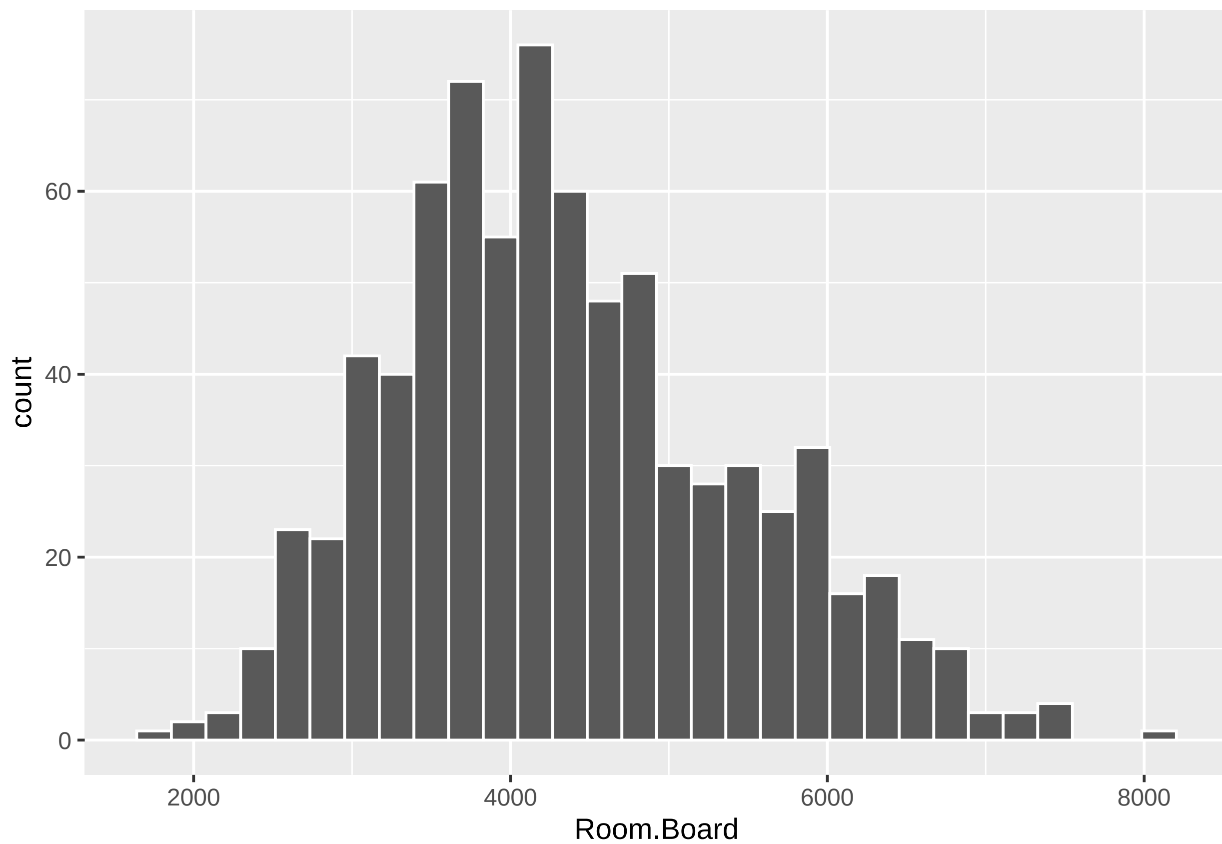
```
ggplot(data = college, mapping = aes(x = Enroll)) +
  geom_histogram(color="white")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(data = college, mapping = aes(x = Room.Board)) +
  geom_histogram(color="white")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The room and board costs tends to center around 3800-5000.

Question 9

```r
auto <- Auto
auto <- auto %>%
  drop_na() #drop na values
```

```r
view(auto)
```

All the column mpg, displacement, horsepower, weighty, acceleration are quantitative. Origin and name, year are qualitative in my opinion.

b)

```r
head(auto, 10)
```

```
##    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
## 7   14         8          454        220   4354          9.0   70       1
## 8   14         8          440        215   4312          8.5   70       1
## 9   14         8          455        225   4425         10.0   70       1
## 10  15         8          390        190   3850          8.5   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
## 7           chevrolet impala
## 8          plymouth fury iii
## 9           pontiac catalina
## 10        amc ambassador dpl
```

```
quan <- auto[, 1:6]
#quan
print(apply(quan, 2, range))
```

```
##       mpg cylinders displacement horsepower weight acceleration
## [1,]  9.0         3           68         46   1613          8.0
## [2,] 46.6         8          455        230   5140         24.8
```

c) Mean, SD

```
#mean
print(apply(quan, 2, mean))
```

```
##          mpg    cylinders displacement   horsepower       weight acceleration
##    23.445918     5.471939   194.411990   104.469388  2977.584184    15.541327
```

```
#SD
print(apply(quan, 2, sd))
```

```
##          mpg    cylinders displacement   horsepower       weight acceleration
##     7.805007     1.705783   104.644004    38.491160   849.402560     2.758864
```

d) Remove 10-85th observation:

```
dim(auto)
```

```
## [1] 392   9
```

```
auto_quan <- quan[-c(10:85), ]
dim(auto_quan)
```

```
## [1] 316   6
```

We can see from the dimension that the rows from 10:85 has been removed.

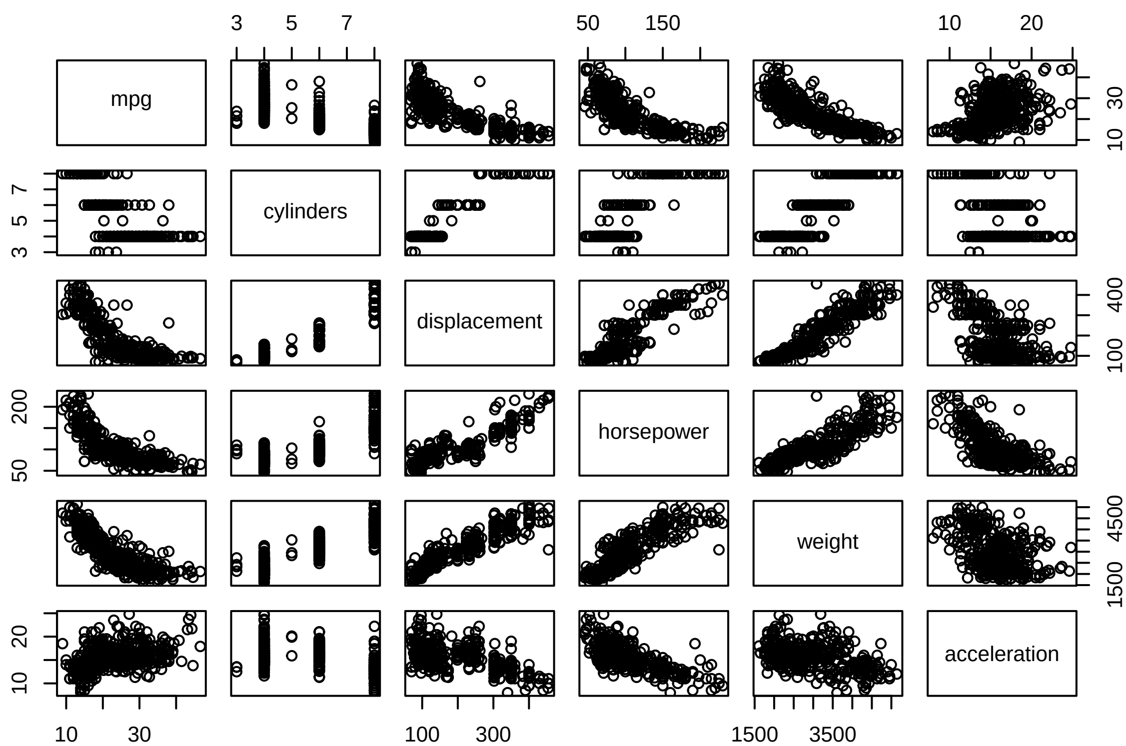
```r
print(apply(auto_quan, 2, range))
```

```
##       mpg cylinders displacement horsepower weight acceleration
## [1,] 11.0         3           68         46   1649          8.5
## [2,] 46.6         8          455        230   4997         24.8
```

The range of the predictors of the value has remained fairly same so we can say that specific for range values, dropping the rows didnot lead to drastic change.

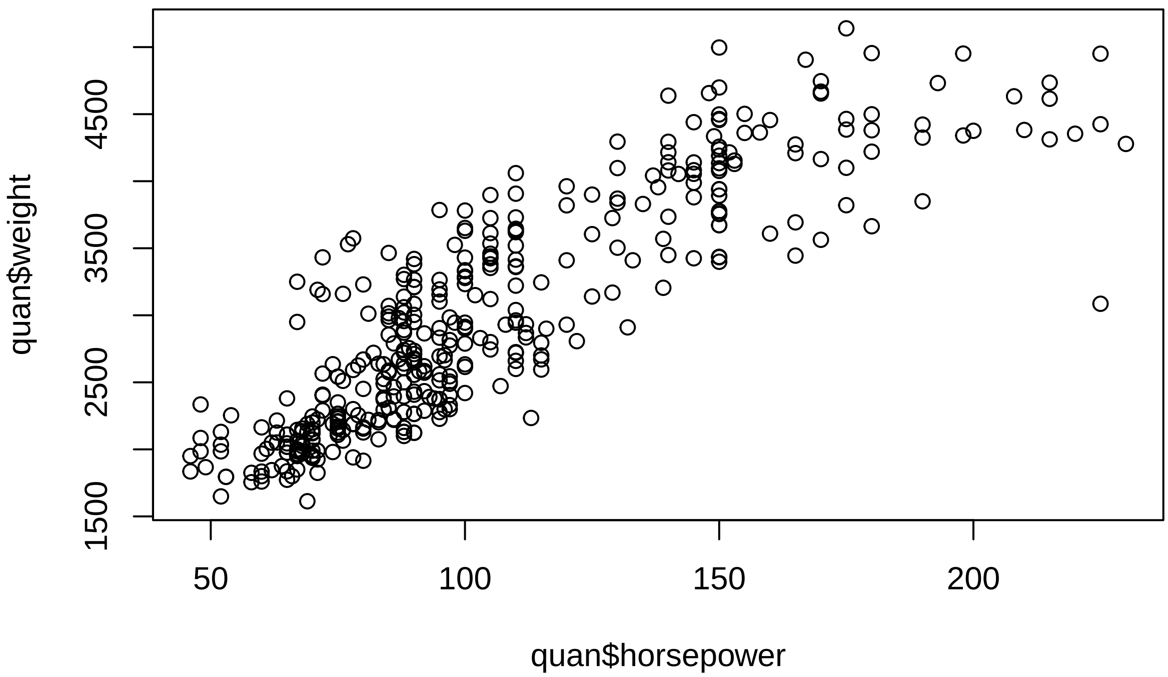
d)

```r
pairs(quan)  #Using the original dataset,
```



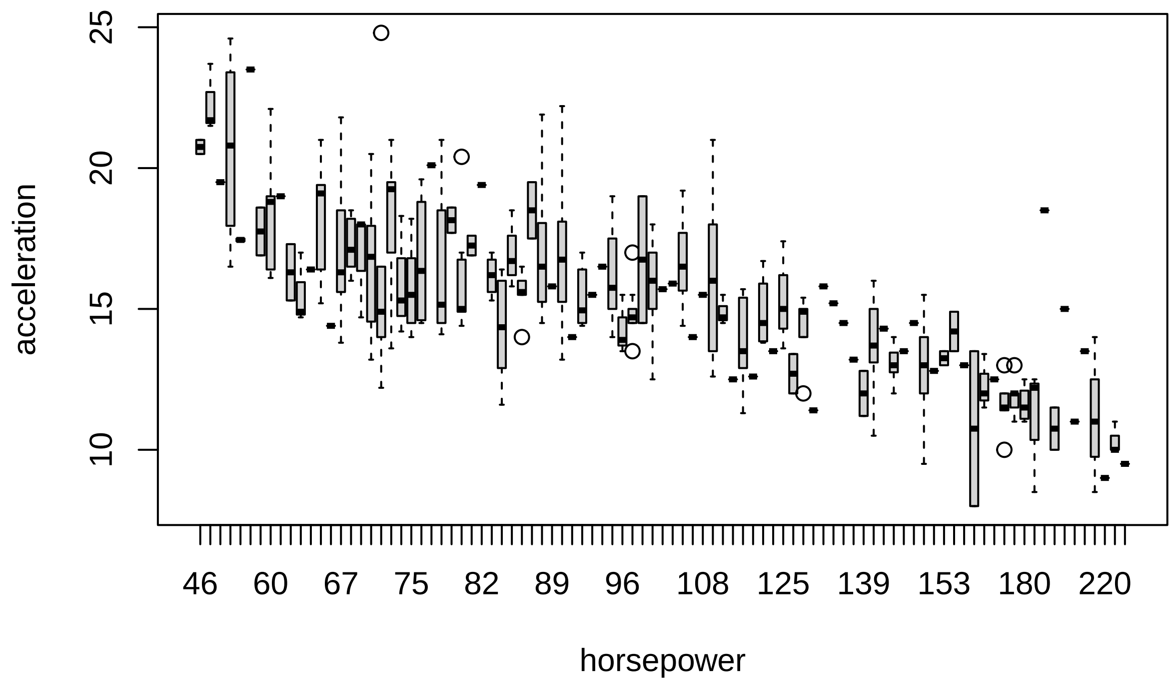
From the pair plot above, we can see intersting relationship between mpg and displacement, horsepwoer and weight. mpg seems to share negative relation will all three of these variables, Similarly, displacement seems to share positive relationship between horsepower and weight. There seems to be lack of clear relationship between acceleration and other variables, however, we could explore if acceleration shares negative relationship with displacement.

```r
#Exploring relationship between horsepower and weight.

plot(quan$horsepower, quan$weight)
```

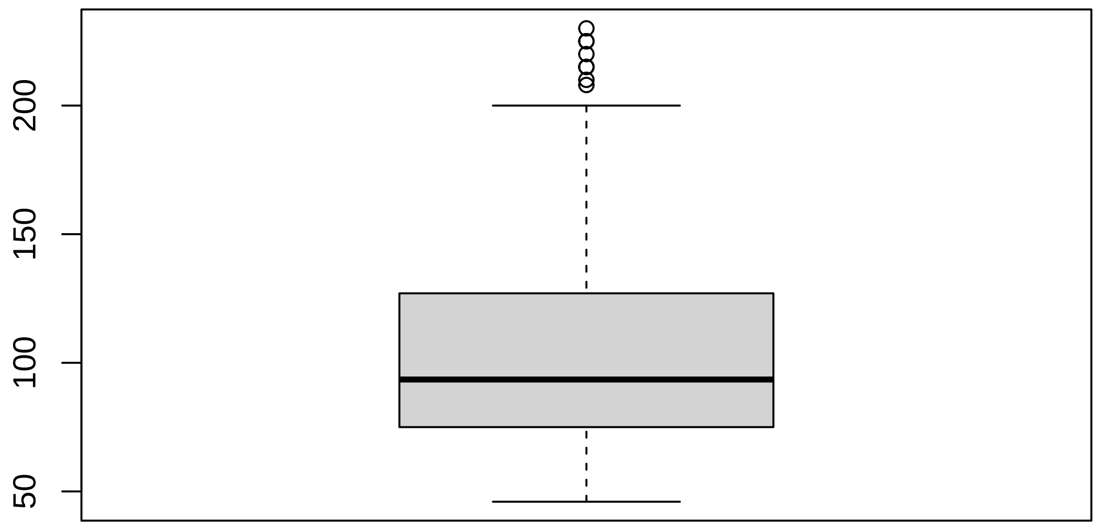


There seems to be positive relationship.

```
# doing boxplot between acceleration and horsepower,

boxplot(acceleration~horsepower, data = quan)
```




```
boxplot(quan$horsepower)
```



f) mpg: horsepower, displacement and weight seem to share fairly negative relationship with mpg so I would include these values when conducting test to predict gas mileage (mpg)

Question 10

a)

```
glimpse(Boston)
```

```
## Rows: 506
## Columns: 13
## $ crim    <dbl> 0.00632, 0.02731, 0.02729, 0.03237, 0.06905, 0.02985, 0.08829,~
## $ zn      <dbl> 18.0, 0.0, 0.0, 0.0, 0.0, 0.0, 12.5, 12.5, 12.5, 12.5, 12.5, 1~
## $ indus   <dbl> 2.31, 7.07, 7.07, 2.18, 2.18, 2.18, 7.87, 7.87, 7.87, 7.87, 7.~
## $ chas    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ nox     <dbl> 0.538, 0.469, 0.469, 0.458, 0.458, 0.458, 0.524, 0.524, 0.524,~
## $ rm      <dbl> 6.575, 6.421, 7.185, 6.998, 7.147, 6.430, 6.012, 6.172, 5.631,~
## $ age     <dbl> 65.2, 78.9, 61.1, 45.8, 54.2, 58.7, 66.6, 96.1, 100.0, 85.9, 9~
## $ dis     <dbl> 4.0900, 4.9671, 4.9671, 6.0622, 6.0622, 6.0622, 5.5605, 5.9505~
## $ rad     <int> 1, 2, 2, 3, 3, 3, 5, 5, 5, 5, 5, 5, 5, 4, 4, 4, 4, 4, 4, 4,~
## $ tax     <dbl> 296, 242, 242, 222, 222, 222, 311, 311, 311, 311, 311, 311, 31~
## $ ptratio <dbl> 15.3, 17.8, 17.8, 18.7, 18.7, 18.7, 15.2, 15.2, 15.2, 15.2, 15~
## $ lstat   <dbl> 4.98, 9.14, 4.03, 2.94, 5.33, 5.21, 12.43, 19.15, 29.93, 17.10~
## $ medv    <dbl> 24.0, 21.6, 34.7, 33.4, 36.2, 28.7, 22.9, 27.1, 16.5, 18.9, 15~
```

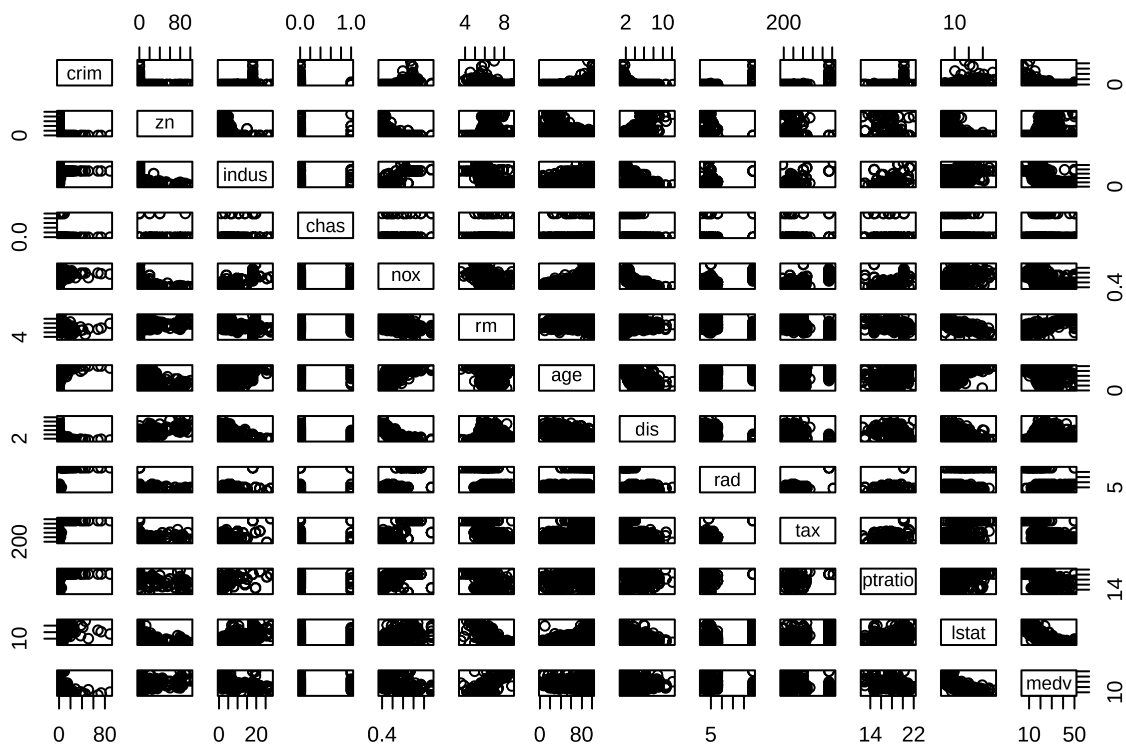
```r
dim(Boston)
```

```
## [1] 506  13
```

There are 506 rows and 13 columns in the dataset. Rows represent a set of observations and columns represents set of predictor values for a neighbourhood in Boston region.

b)

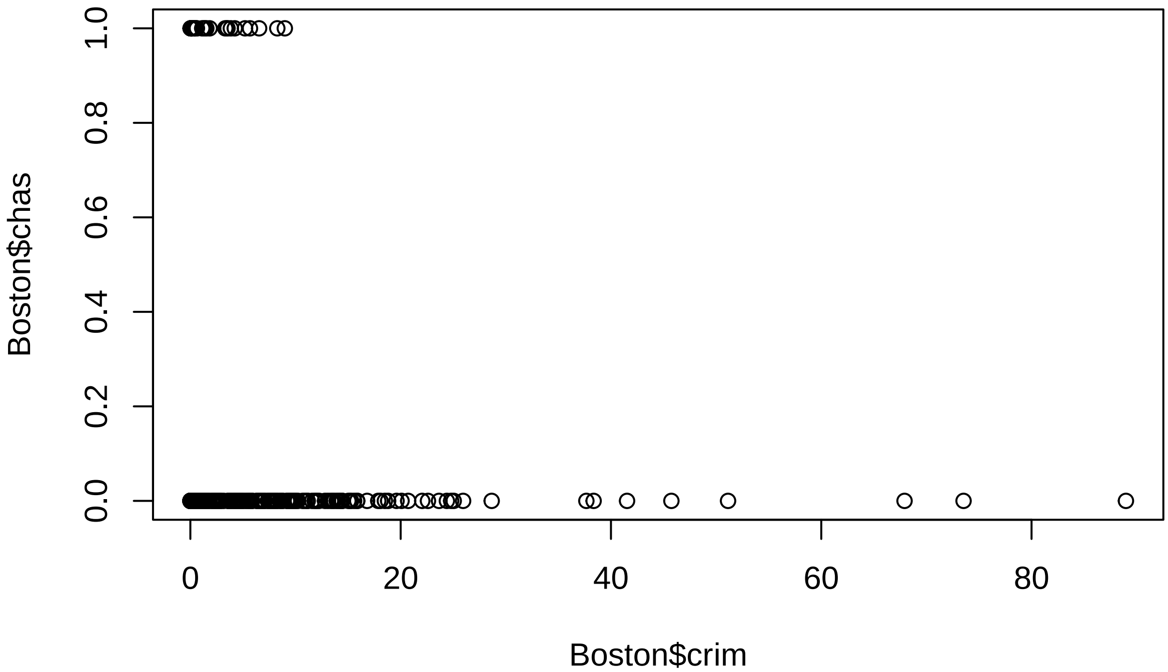
```r
pairs(Boston)
```



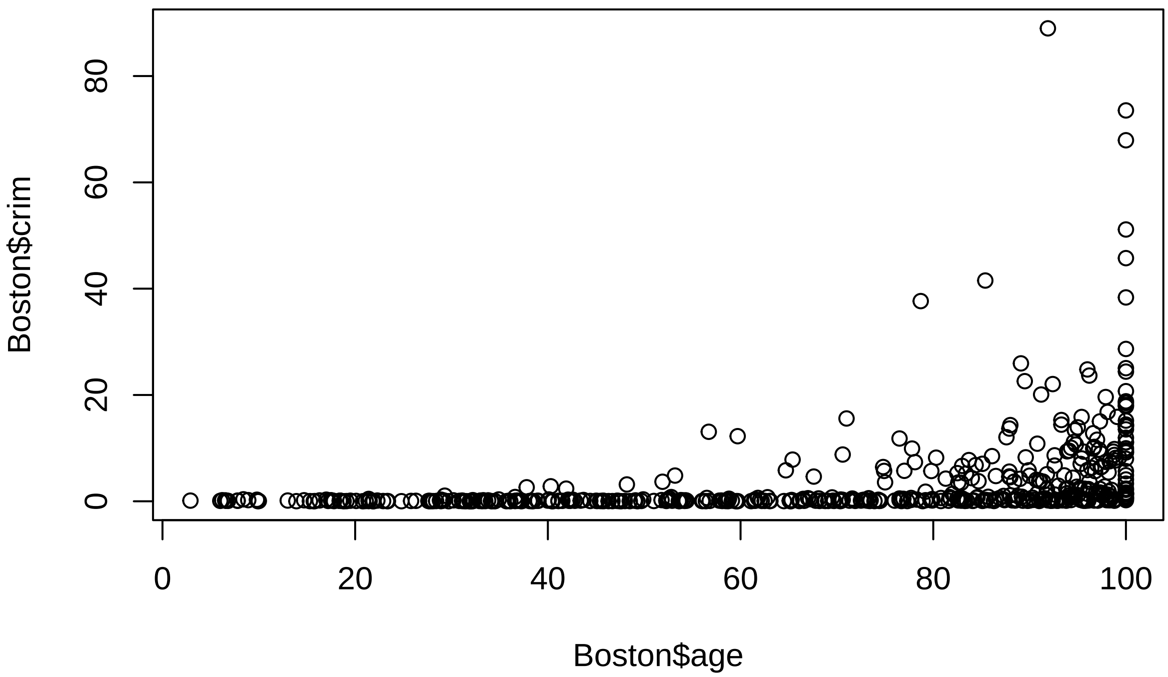
It is hard to make conclusive observations from the pair plot. All the variables seems to distributed randomly, only relationship I could observe were: dis and nax some negative relationship and lstat and medv negative relationship.

c)

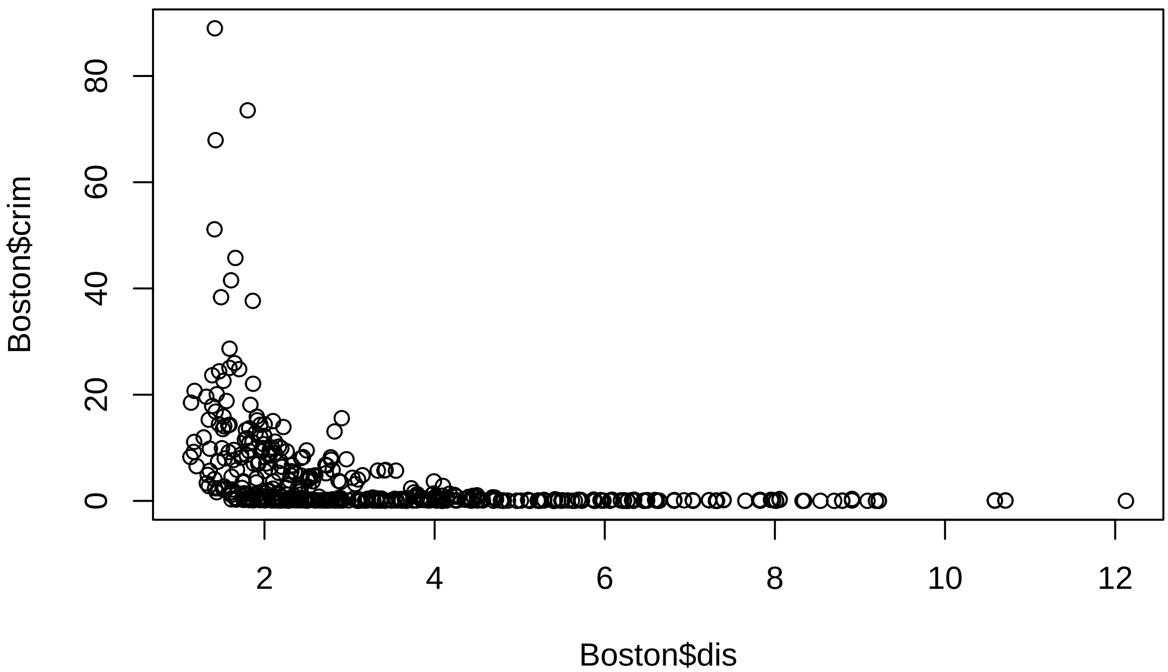
```r
plot(Boston$crim, Boston$chas)
```



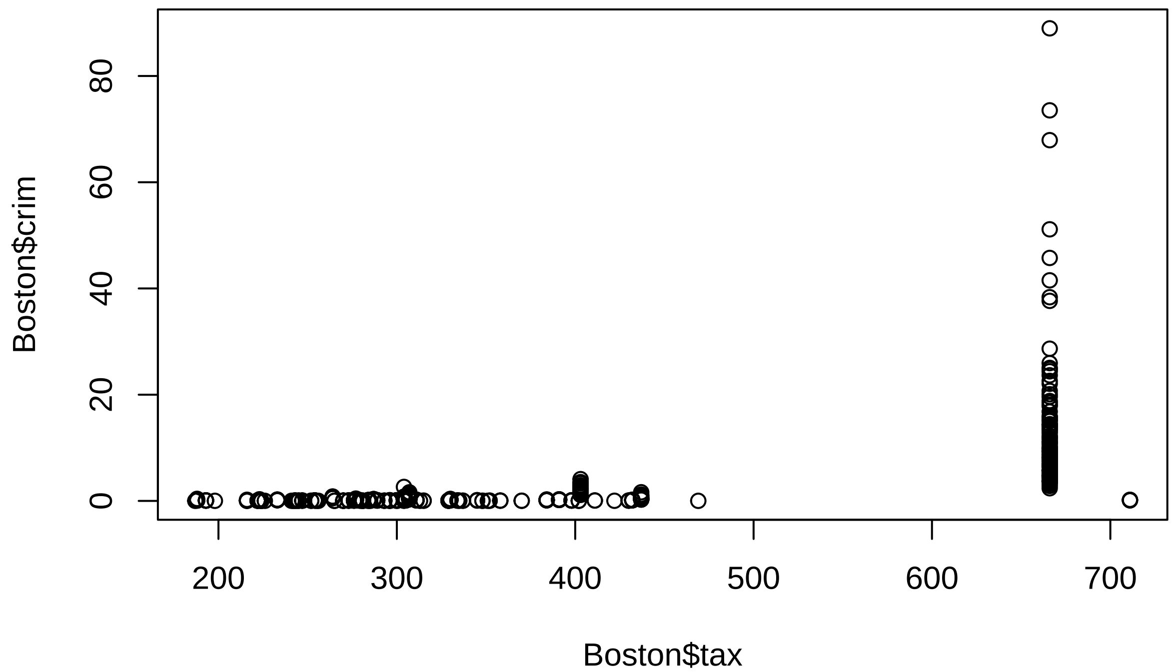

```
plot(Boston$age, Boston$crim)
```




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




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



Just from graphs above we can observe that crime is related to age, dis, tax etc

d)

```
#glimpse(Boston)
crim_rel <- Boston[, c(1, 10,11)]
crim_rel
```

```
##          crim tax ptratio
## 1     0.00632 296    15.3
## 2     0.02731 242    17.8
## 3     0.02729 242    17.8
## 4     0.03237 222    18.7
## 5     0.06905 222    18.7
## 6     0.02985 222    18.7
## 7     0.08829 311    15.2
## 8     0.14455 311    15.2
## 9     0.21124 311    15.2
## 10    0.17004 311    15.2
## 11    0.22489 311    15.2
## 12    0.11747 311    15.2
## 13    0.09378 311    15.2
## 14    0.62976 307    21.0
## 15    0.63796 307    21.0
## 16    0.62739 307    21.0
## 17    1.05393 307    21.0
```

```
## 18   0.78420 307    21.0
## 19   0.80271 307    21.0
## 20   0.72580 307    21.0
## 21   1.25179 307    21.0
## 22   0.85204 307    21.0
## 23   1.23247 307    21.0
## 24   0.98843 307    21.0
## 25   0.75026 307    21.0
## 26   0.84054 307    21.0
## 27   0.67191 307    21.0
## 28   0.95577 307    21.0
## 29   0.77299 307    21.0
## 30   1.00245 307    21.0
## 31   1.13081 307    21.0
## 32   1.35472 307    21.0
## 33   1.38799 307    21.0
## 34   1.15172 307    21.0
## 35   1.61282 307    21.0
## 36   0.06417 279    19.2
## 37   0.09744 279    19.2
## 38   0.08014 279    19.2
## 39   0.17505 279    19.2
## 40   0.02763 252    18.3
## 41   0.03359 252    18.3
## 42   0.12744 233    17.9
## 43   0.14150 233    17.9
## 44   0.15936 233    17.9
## 45   0.12269 233    17.9
## 46   0.17142 233    17.9
## 47   0.18836 233    17.9
## 48   0.22927 233    17.9
## 49   0.25387 233    17.9
## 50   0.21977 233    17.9
## 51   0.08873 243    16.8
## 52   0.04337 243    16.8
## 53   0.05360 243    16.8
## 54   0.04981 243    16.8
## 55   0.01360 469    21.1
## 56   0.01311 226    17.9
## 57   0.02055 313    17.3
## 58   0.01432 256    15.1
## 59   0.15445 284    19.7
## 60   0.10328 284    19.7
## 61   0.14932 284    19.7
## 62   0.17171 284    19.7
## 63   0.11027 284    19.7
## 64   0.12650 284    19.7
## 65   0.01951 216    18.6
## 66   0.03584 337    16.1
## 67   0.04379 337    16.1
## 68   0.05789 345    18.9
## 69   0.13554 345    18.9
## 70   0.12816 345    18.9
## 71   0.08826 305    19.2
```

```
## 72   0.15876 305    19.2
## 73   0.09164 305    19.2
## 74   0.19539 305    19.2
## 75   0.07896 398    18.7
## 76   0.09512 398    18.7
## 77   0.10153 398    18.7
## 78   0.08707 398    18.7
## 79   0.05646 398    18.7
## 80   0.08387 398    18.7
## 81   0.04113 281    19.0
## 82   0.04462 281    19.0
## 83   0.03659 281    19.0
## 84   0.03551 281    19.0
## 85   0.05059 247    18.5
## 86   0.05735 247    18.5
## 87   0.05188 247    18.5
## 88   0.07151 247    18.5
## 89   0.05660 270    17.8
## 90   0.05302 270    17.8
## 91   0.04684 270    17.8
## 92   0.03932 270    17.8
## 93   0.04203 270    18.2
## 94   0.02875 270    18.2
## 95   0.04294 270    18.2
## 96   0.12204 276    18.0
## 97   0.11504 276    18.0
## 98   0.12083 276    18.0
## 99   0.08187 276    18.0
## 100  0.06860 276    18.0
## 101  0.14866 384    20.9
## 102  0.11432 384    20.9
## 103  0.22876 384    20.9
## 104  0.21161 384    20.9
## 105  0.13960 384    20.9
## 106  0.13262 384    20.9
## 107  0.17120 384    20.9
## 108  0.13117 384    20.9
## 109  0.12802 384    20.9
## 110  0.26363 384    20.9
## 111  0.10793 384    20.9
## 112  0.10084 432    17.8
## 113  0.12329 432    17.8
## 114  0.22212 432    17.8
## 115  0.14231 432    17.8
## 116  0.17134 432    17.8
## 117  0.13158 432    17.8
## 118  0.15098 432    17.8
## 119  0.13058 432    17.8
## 120  0.14476 432    17.8
## 121  0.06899 188    19.1
## 122  0.07165 188    19.1
## 123  0.09299 188    19.1
## 124  0.15038 188    19.1
## 125  0.09849 188    19.1
```

```
## 126  0.16902 188    19.1
## 127  0.38735 188    19.1
## 128  0.25915 437    21.2
## 129  0.32543 437    21.2
## 130  0.88125 437    21.2
## 131  0.34006 437    21.2
## 132  1.19294 437    21.2
## 133  0.59005 437    21.2
## 134  0.32982 437    21.2
## 135  0.97617 437    21.2
## 136  0.55778 437    21.2
## 137  0.32264 437    21.2
## 138  0.35233 437    21.2
## 139  0.24980 437    21.2
## 140  0.54452 437    21.2
## 141  0.29090 437    21.2
## 142  1.62864 437    21.2
## 143  3.32105 403    14.7
## 144  4.09740 403    14.7
## 145  2.77974 403    14.7
## 146  2.37934 403    14.7
## 147  2.15505 403    14.7
## 148  2.36862 403    14.7
## 149  2.33099 403    14.7
## 150  2.73397 403    14.7
## 151  1.65660 403    14.7
## 152  1.49632 403    14.7
## 153  1.12658 403    14.7
## 154  2.14918 403    14.7
## 155  1.41385 403    14.7
## 156  3.53501 403    14.7
## 157  2.44668 403    14.7
## 158  1.22358 403    14.7
## 159  1.34284 403    14.7
## 160  1.42502 403    14.7
## 161  1.27346 403    14.7
## 162  1.46336 403    14.7
## 163  1.83377 403    14.7
## 164  1.51902 403    14.7
## 165  2.24236 403    14.7
## 166  2.92400 403    14.7
## 167  2.01019 403    14.7
## 168  1.80028 403    14.7
## 169  2.30040 403    14.7
## 170  2.44953 403    14.7
## 171  1.20742 403    14.7
## 172  2.31390 403    14.7
## 173  0.13914 296    16.6
## 174  0.09178 296    16.6
## 175  0.08447 296    16.6
## 176  0.06664 296    16.6
## 177  0.07022 296    16.6
## 178  0.05425 296    16.6
## 179  0.06642 296    16.6
```

```
## 180  0.05780 193    17.8
## 181  0.06588 193    17.8
## 182  0.06888 193    17.8
## 183  0.09103 193    17.8
## 184  0.10008 193    17.8
## 185  0.08308 193    17.8
## 186  0.06047 193    17.8
## 187  0.05602 193    17.8
## 188  0.07875 398    15.2
## 189  0.12579 398    15.2
## 190  0.08370 398    15.2
## 191  0.09068 398    15.2
## 192  0.06911 398    15.2
## 193  0.08664 398    15.2
## 194  0.02187 265    15.6
## 195  0.01439 265    15.6
## 196  0.01381 255    14.4
## 197  0.04011 329    12.6
## 198  0.04666 329    12.6
## 199  0.03768 329    12.6
## 200  0.03150 402    17.0
## 201  0.01778 402    17.0
## 202  0.03445 348    14.7
## 203  0.02177 348    14.7
## 204  0.03510 224    14.7
## 205  0.02009 224    14.7
## 206  0.13642 277    18.6
## 207  0.22969 277    18.6
## 208  0.25199 277    18.6
## 209  0.13587 277    18.6
## 210  0.43571 277    18.6
## 211  0.17446 277    18.6
## 212  0.37578 277    18.6
## 213  0.21719 277    18.6
## 214  0.14052 277    18.6
## 215  0.28955 277    18.6
## 216  0.19802 277    18.6
## 217  0.04560 276    16.4
## 218  0.07013 276    16.4
## 219  0.11069 276    16.4
## 220  0.11425 276    16.4
## 221  0.35809 307    17.4
## 222  0.40771 307    17.4
## 223  0.62356 307    17.4
## 224  0.61470 307    17.4
## 225  0.31533 307    17.4
## 226  0.52693 307    17.4
## 227  0.38214 307    17.4
## 228  0.41238 307    17.4
## 229  0.29819 307    17.4
## 230  0.44178 307    17.4
## 231  0.53700 307    17.4
## 232  0.46296 307    17.4
## 233  0.57529 307    17.4
```

```
## 234  0.33147 307    17.4
## 235  0.44791 307    17.4
## 236  0.33045 307    17.4
## 237  0.52058 307    17.4
## 238  0.51183 307    17.4
## 239  0.08244 300    16.6
## 240  0.09252 300    16.6
## 241  0.11329 300    16.6
## 242  0.10612 300    16.6
## 243  0.10290 300    16.6
## 244  0.12757 300    16.6
## 245  0.20608 330    19.1
## 246  0.19133 330    19.1
## 247  0.33983 330    19.1
## 248  0.19657 330    19.1
## 249  0.16439 330    19.1
## 250  0.19073 330    19.1
## 251  0.14030 330    19.1
## 252  0.21409 330    19.1
## 253  0.08221 330    19.1
## 254  0.36894 330    19.1
## 255  0.04819 315    16.4
## 256  0.03548 315    16.4
## 257  0.01538 244    15.9
## 258  0.61154 264    13.0
## 259  0.66351 264    13.0
## 260  0.65665 264    13.0
## 261  0.54011 264    13.0
## 262  0.53412 264    13.0
## 263  0.52014 264    13.0
## 264  0.82526 264    13.0
## 265  0.55007 264    13.0
## 266  0.76162 264    13.0
## 267  0.78570 264    13.0
## 268  0.57834 264    13.0
## 269  0.54050 264    13.0
## 270  0.09065 223    18.6
## 271  0.29916 223    18.6
## 272  0.16211 223    18.6
## 273  0.11460 223    18.6
## 274  0.22188 223    18.6
## 275  0.05644 254    17.6
## 276  0.09604 254    17.6
## 277  0.10469 254    17.6
## 278  0.06127 254    17.6
## 279  0.07978 254    17.6
## 280  0.21038 216    14.9
## 281  0.03578 216    14.9
## 282  0.03705 216    14.9
## 283  0.06129 216    14.9
## 284  0.01501 198    13.6
## 285  0.00906 285    15.3
## 286  0.01096 300    15.3
## 287  0.01965 241    18.2
```

```
## 288  0.03871 293    16.6
## 289  0.04590 293    16.6
## 290  0.04297 293    16.6
## 291  0.03502 245    19.2
## 292  0.07886 245    19.2
## 293  0.03615 245    19.2
## 294  0.08265 289    16.0
## 295  0.08199 289    16.0
## 296  0.12932 289    16.0
## 297  0.05372 289    16.0
## 298  0.14103 289    16.0
## 299  0.06466 358    14.8
## 300  0.05561 358    14.8
## 301  0.04417 358    14.8
## 302  0.03537 329    16.1
## 303  0.09266 329    16.1
## 304  0.10000 329    16.1
## 305  0.05515 222    18.4
## 306  0.05479 222    18.4
## 307  0.07503 222    18.4
## 308  0.04932 222    18.4
## 309  0.49298 304    18.4
## 310  0.34940 304    18.4
## 311  2.63548 304    18.4
## 312  0.79041 304    18.4
## 313  0.26169 304    18.4
## 314  0.26938 304    18.4
## 315  0.36920 304    18.4
## 316  0.25356 304    18.4
## 317  0.31827 304    18.4
## 318  0.24522 304    18.4
## 319  0.40202 304    18.4
## 320  0.47547 304    18.4
## 321  0.16760 287    19.6
## 322  0.18159 287    19.6
## 323  0.35114 287    19.6
## 324  0.28392 287    19.6
## 325  0.34109 287    19.6
## 326  0.19186 287    19.6
## 327  0.30347 287    19.6
## 328  0.24103 287    19.6
## 329  0.06617 430    16.9
## 330  0.06724 430    16.9
## 331  0.04544 430    16.9
## 332  0.05023 304    16.9
## 333  0.03466 304    16.9
## 334  0.05083 224    20.2
## 335  0.03738 224    20.2
## 336  0.03961 224    20.2
## 337  0.03427 224    20.2
## 338  0.03041 224    20.2
## 339  0.03306 224    20.2
## 340  0.05497 224    20.2
## 341  0.06151 224    20.2
```

```
## 342  0.01301 284    15.5
## 343  0.02498 422    15.9
## 344  0.02543 370    17.6
## 345  0.03049 370    17.6
## 346  0.03113 352    18.8
## 347  0.06162 352    18.8
## 348  0.01870 351    17.9
## 349  0.01501 280    17.0
## 350  0.02899 335    19.7
## 351  0.06211 335    19.7
## 352  0.07950 411    18.3
## 353  0.07244 411    18.3
## 354  0.01709 187    17.0
## 355  0.04301 334    22.0
## 356  0.10659 334    22.0
## 357  8.98296 666    20.2
## 358  3.84970 666    20.2
## 359  5.20177 666    20.2
## 360  4.26131 666    20.2
## 361  4.54192 666    20.2
## 362  3.83684 666    20.2
## 363  3.67822 666    20.2
## 364  4.22239 666    20.2
## 365  3.47428 666    20.2
## 366  4.55587 666    20.2
## 367  3.69695 666    20.2
## 368 13.52220 666    20.2
## 369  4.89822 666    20.2
## 370  5.66998 666    20.2
## 371  6.53876 666    20.2
## 372  9.23230 666    20.2
## 373  8.26725 666    20.2
## 374 11.10810 666    20.2
## 375 18.49820 666    20.2
## 376 19.60910 666    20.2
## 377 15.28800 666    20.2
## 378  9.82349 666    20.2
## 379 23.64820 666    20.2
## 380 17.86670 666    20.2
## 381 88.97620 666    20.2
## 382 15.87440 666    20.2
## 383  9.18702 666    20.2
## 384  7.99248 666    20.2
## 385 20.08490 666    20.2
## 386 16.81180 666    20.2
## 387 24.39380 666    20.2
## 388 22.59710 666    20.2
## 389 14.33370 666    20.2
## 390  8.15174 666    20.2
## 391  6.96215 666    20.2
## 392  5.29305 666    20.2
## 393 11.57790 666    20.2
## 394  8.64476 666    20.2
## 395 13.35980 666    20.2
```

```
## 396  8.71675 666    20.2
## 397  5.87205 666    20.2
## 398  7.67202 666    20.2
## 399 38.35180 666    20.2
## 400  9.91655 666    20.2
## 401 25.04610 666    20.2
## 402 14.23620 666    20.2
## 403  9.59571 666    20.2
## 404 24.80170 666    20.2
## 405 41.52920 666    20.2
## 406 67.92080 666    20.2
## 407 20.71620 666    20.2
## 408 11.95110 666    20.2
## 409  7.40389 666    20.2
## 410 14.43830 666    20.2
## 411 51.13580 666    20.2
## 412 14.05070 666    20.2
## 413 18.81100 666    20.2
## 414 28.65580 666    20.2
## 415 45.74610 666    20.2
## 416 18.08460 666    20.2
## 417 10.83420 666    20.2
## 418 25.94060 666    20.2
## 419 73.53410 666    20.2
## 420 11.81230 666    20.2
## 421 11.08740 666    20.2
## 422  7.02259 666    20.2
## 423 12.04820 666    20.2
## 424  7.05042 666    20.2
## 425  8.79212 666    20.2
## 426 15.86030 666    20.2
## 427 12.24720 666    20.2
## 428 37.66190 666    20.2
## 429  7.36711 666    20.2
## 430  9.33889 666    20.2
## 431  8.49213 666    20.2
## 432 10.06230 666    20.2
## 433  6.44405 666    20.2
## 434  5.58107 666    20.2
## 435 13.91340 666    20.2
## 436 11.16040 666    20.2
## 437 14.42080 666    20.2
## 438 15.17720 666    20.2
## 439 13.67810 666    20.2
## 440  9.39063 666    20.2
## 441 22.05110 666    20.2
## 442  9.72418 666    20.2
## 443  5.66637 666    20.2
## 444  9.96654 666    20.2
## 445 12.80230 666    20.2
## 446 10.67180 666    20.2
## 447  6.28807 666    20.2
## 448  9.92485 666    20.2
## 449  9.32909 666    20.2
```

```
## 450  7.52601 666    20.2
## 451  6.71772 666    20.2
## 452  5.44114 666    20.2
## 453  5.09017 666    20.2
## 454  8.24809 666    20.2
## 455  9.51363 666    20.2
## 456  4.75237 666    20.2
## 457  4.66883 666    20.2
## 458  8.20058 666    20.2
## 459  7.75223 666    20.2
## 460  6.80117 666    20.2
## 461  4.81213 666    20.2
## 462  3.69311 666    20.2
## 463  6.65492 666    20.2
## 464  5.82115 666    20.2
## 465  7.83932 666    20.2
## 466  3.16360 666    20.2
## 467  3.77498 666    20.2
## 468  4.42228 666    20.2
## 469 15.57570 666    20.2
## 470 13.07510 666    20.2
## 471  4.34879 666    20.2
## 472  4.03841 666    20.2
## 473  3.56868 666    20.2
## 474  4.64689 666    20.2
## 475  8.05579 666    20.2
## 476  6.39312 666    20.2
## 477  4.87141 666    20.2
## 478 15.02340 666    20.2
## 479 10.23300 666    20.2
## 480 14.33370 666    20.2
## 481  5.82401 666    20.2
## 482  5.70818 666    20.2
## 483  5.73116 666    20.2
## 484  2.81838 666    20.2
## 485  2.37857 666    20.2
## 486  3.67367 666    20.2
## 487  5.69175 666    20.2
## 488  4.83567 666    20.2
## 489  0.15086 711    20.1
## 490  0.18337 711    20.1
## 491  0.20746 711    20.1
## 492  0.10574 711    20.1
## 493  0.11132 711    20.1
## 494  0.17331 391    19.2
## 495  0.27957 391    19.2
## 496  0.17899 391    19.2
## 497  0.28960 391    19.2
## 498  0.26838 391    19.2
## 499  0.23912 391    19.2
## 500  0.17783 391    19.2
## 501  0.22438 391    19.2
## 502  0.06263 273    21.0
## 503  0.04527 273    21.0
```

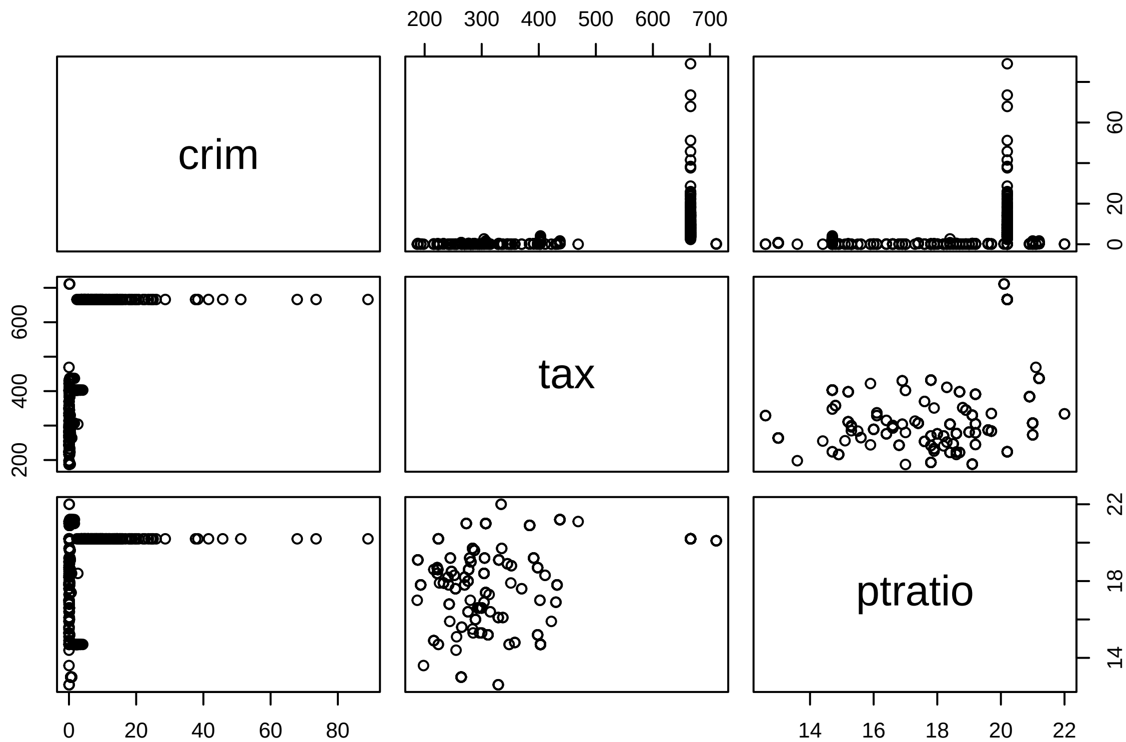
```
## 504  0.06076 273    21.0
## 505  0.10959 273    21.0
## 506  0.04741 273    21.0
```

```
apply(crim_rel, 2, range)
```

```
##             crim tax ptratio
## [1,]  0.00632 187    12.6
## [2,] 88.97620 711    22.0
```

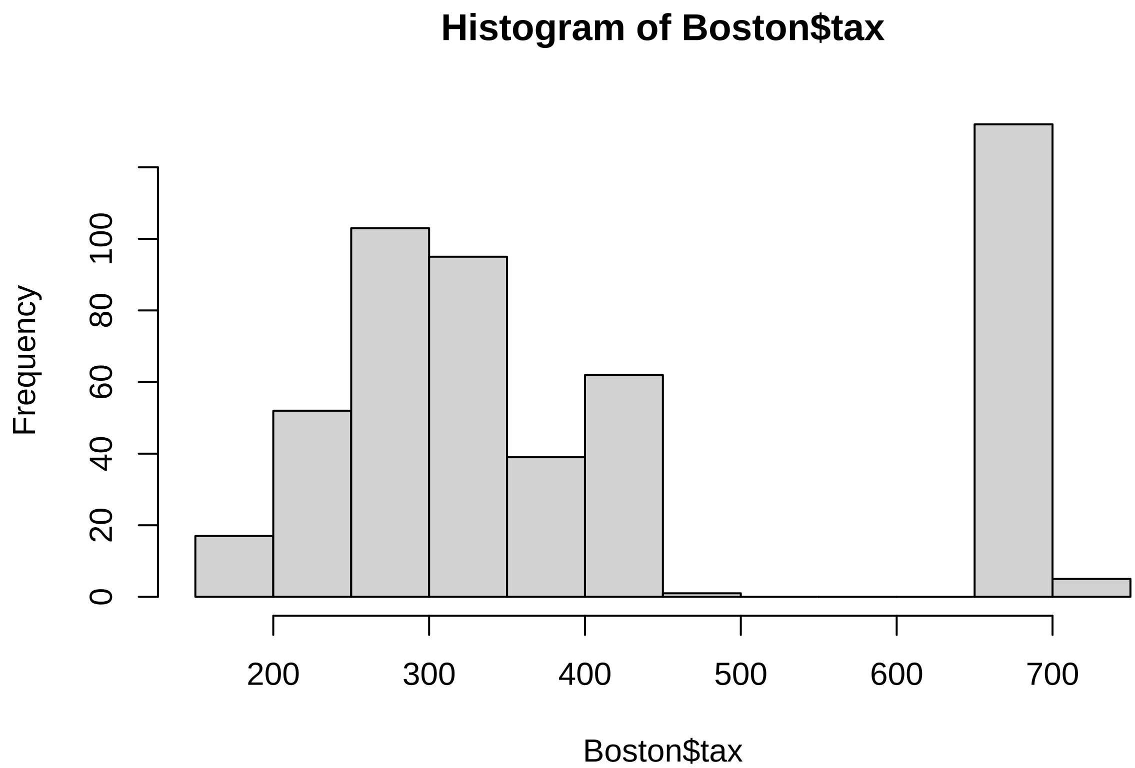
From above value we can see the range of crime level, tax and the pupil-teacher ratio.

```
pairs(crim_rel)
```



We can observe relationship between crime rate and tax and crime rate and pupil teacher ratio from the figure above.

```
hist(Boston$tax)
```



e)

```
nrow(subset(Boston, chas == 1))
```

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

There are 35 suburbs that are bound by the Charles river.

f)

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

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

On average, the pupil to tracher ratio is 19.05.

g)

```
lowest <- subset(Boston, medv == min(medv))
lowest
```

```
##        crim zn indus chas   nox    rm age    dis rad tax ptratio lstat medv
## 399 38.3518  0  18.1    0 0.693 5.453 100 1.4896  24 666    20.2 30.59    5
## 406 67.9208  0  18.1    0 0.693 5.683 100 1.4254  24 666    20.2 22.98    5
```

With the value of 5 in medv there are two census tract that give up this value, row 399 and row 406.

h)

```
rm_7 <-subset(Boston, rm >7)
nrow(rm_7)
```

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

There are 64 suburbs with more than 7 rooms peer dwelling.

```
rm_8 <-subset(Boston, rm >8)
nrow(rm_8)
```

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

There are 13 suburbs with more than 8 rooms peer dwelling.