DSE6011 - Module 1 - CH 2 HW

Nathan Monges

2024-07-07

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- 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 errors
```

Exercise 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("College.csv")
```

b)

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

```
##                                                         X Private Apps Accept
## Abilene Christian University Abilene Christian University     Yes 1660   1232
## Adelphi University                     Adelphi University     Yes 2186   1924
## Adrian College                             Adrian College     Yes 1428   1097
## Agnes Scott College                   Agnes Scott College     Yes  417    349
## Alaska Pacific University       Alaska Pacific University     Yes  193    146
## Albertson College                       Albertson College     Yes  587    479
##                              Enroll Top10perc Top25perc F.Undergrad P.Undergrad
## Abilene Christian University    721        23        52        2885         537
## Adelphi University              512        16        29        2683        1227
## Adrian College                  336        22        50        1036          99
## Agnes Scott College             137        60        89         510          63
## Alaska Pacific University        55        16        44         249         869
## Albertson College               158        38        62         678          41
##                              Outstate Room.Board Books Personal PhD Terminal
## Abilene Christian University     7440       3300   450     2200  70       78
## Adelphi University              12280       6450   750     1500  29       30
## Adrian College                  11250       3750   400     1165  53       66
## Agnes Scott College             12960       5450   450      875  92       97
```

```
## Alaska Pacific University        7560       4120   800     1500  76       72
## Albertson College               13500       3335   500      675  67       73
##                              S.F.Ratio perc.alumni Expend Grad.Rate
## Abilene Christian University      18.1          12   7041        60
## Adelphi University                12.2          16  10527        56
## Adrian College                    12.9          30   8735        54
## Agnes Scott College                7.7          37  19016        59
## Alaska Pacific University         11.9           2  10922        15
## Albertson College                  9.4          11   9727        55
```

```
college <- college[ , -1]

head(college)
```

```
##                              Private Apps Accept Enroll Top10perc Top25perc
## Abilene Christian University     Yes 1660   1232    721        23        52
## Adelphi University               Yes 2186   1924    512        16        29
## Adrian College                   Yes 1428   1097    336        22        50
## Agnes Scott College              Yes  417    349    137        60        89
## Alaska Pacific University        Yes  193    146     55        16        44
## Albertson College                Yes  587    479    158        38        62
##                              F.Undergrad P.Undergrad Outstate Room.Board Books
## Abilene Christian University        2885         537     7440       3300   450
## Adelphi University                  2683        1227    12280       6450   750
## Adrian College                      1036          99    11250       3750   400
## Agnes Scott College                  510          63    12960       5450   450
## Alaska Pacific University            249         869     7560       4120   800
## Albertson College                    678          41    13500       3335   500
##                              Personal PhD Terminal S.F.Ratio perc.alumni Expend
## Abilene Christian University     2200  70       78      18.1          12   7041
## Adelphi University               1500  29       30      12.2          16  10527
## Adrian College                   1165  53       66      12.9          30   8735
## Agnes Scott College               875  92       97       7.7          37  19016
## Alaska Pacific University        1500  76       72      11.9           2  10922
## Albertson College                 675  67       73       9.4          11   9727
##                              Grad.Rate
## Abilene Christian University        60
## Adelphi University                  56
## Adrian College                      54
## Agnes Scott College                 59
## Alaska Pacific University           15
## Albertson College                   55
```

c)

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

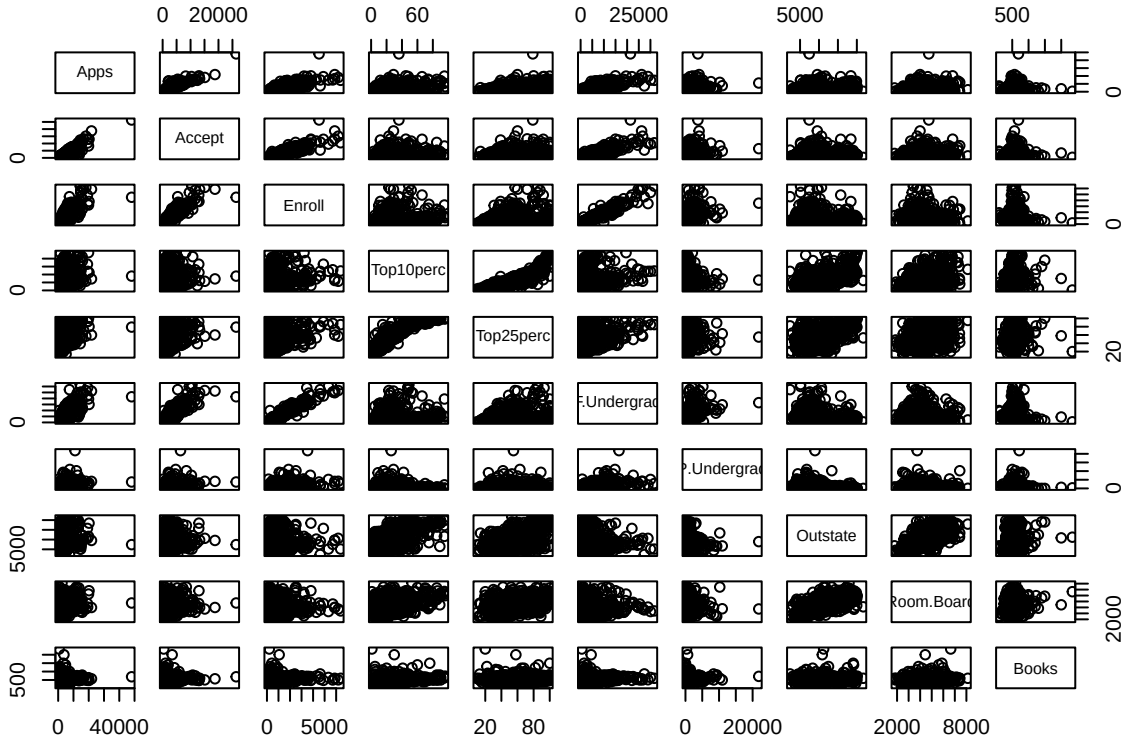
```
summary(college)
```

```
##    Private               Apps           Accept          Enroll
##  Length:777         Min.   :   81   Min.   :   72   Min.   :  35
##  Class :character   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242
##  Mode  :character   Median : 1558   Median : 1110   Median : 434
##                     Mean   : 3002   Mean   : 2019   Mean   : 780
##                     3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902
##                     Max.   :48094   Max.   :26330   Max.   :6392
```

```
##    Top10perc       Top25perc      F.Undergrad     P.Undergrad
##  Min.   : 1.00   Min.   :  9.0   Min.   :  139   Min.   :    1.0
##  1st Qu.:15.00   1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0
##  Median :23.00   Median : 54.0   Median : 1707   Median :  353.0
##  Mean   :27.56   Mean   : 55.8   Mean   : 3700   Mean   :  855.3
##  3rd Qu.:35.00   3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0
##  Max.   :96.00   Max.   :100.0   Max.   :31643   Max.   :21836.0
##     Outstate       Room.Board       Books           Personal
##  Min.   : 2340   Min.   :1780   Min.   :  96.0   Min.   : 250
##  1st Qu.: 7320   1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850
##  Median : 9990   Median :4200   Median : 500.0   Median :1200
##  Mean   :10441   Mean   :4358   Mean   : 549.4   Mean   :1341
##  3rd Qu.:12925   3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700
##  Max.   :21700   Max.   :8124   Max.   :2340.0   Max.   :6800
##       PhD            Terminal       S.F.Ratio      perc.alumni
##  Min.   :  8.00   Min.   : 24.0   Min.   : 2.50   Min.   : 0.00
##  1st Qu.: 62.00   1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00
##  Median : 75.00   Median : 82.0   Median :13.60   Median :21.00
##  Mean   : 72.66   Mean   : 79.7   Mean   :14.09   Mean   :22.74
##  3rd Qu.: 85.00   3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00
##  Max.   :103.00   Max.   :100.0   Max.   :39.80   Max.   :64.00
##      Expend        Grad.Rate
##  Min.   : 3186   Min.   : 10.00
##  1st Qu.: 6751   1st Qu.: 53.00
##  Median : 8377   Median : 65.00
##  Mean   : 9660   Mean   : 65.46
##  3rd Qu.:10830   3rd Qu.: 78.00
##  Max.   :56233   Max.   :118.00
```

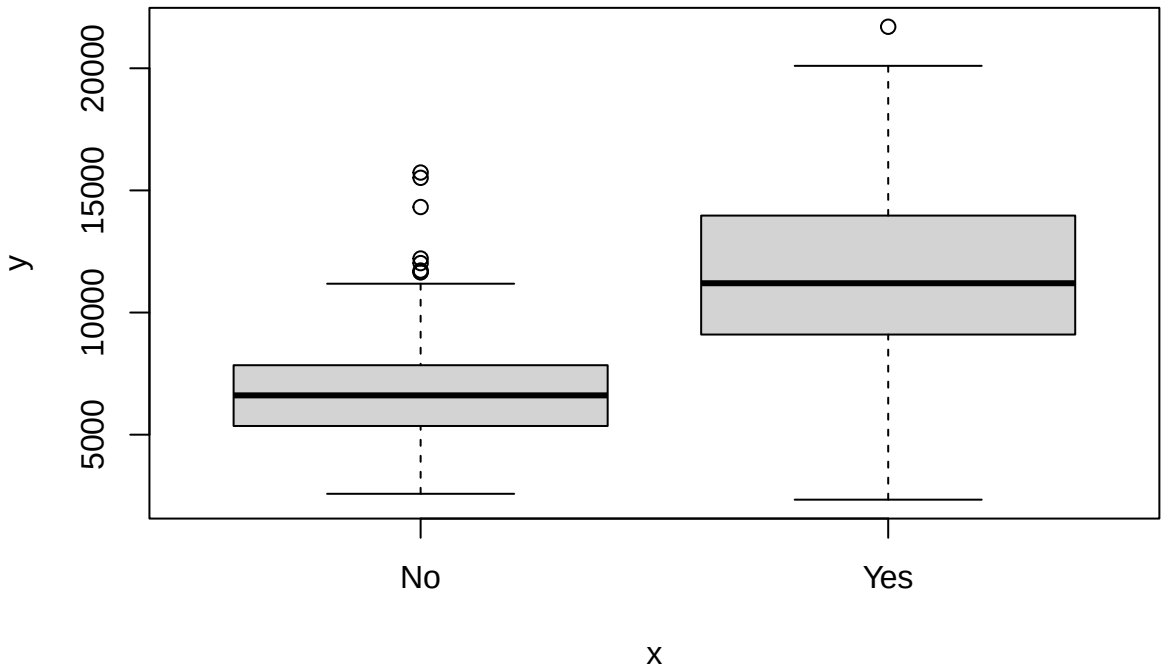
ii. Use the pairs() function to produce a scatterplot matrix of the first ten columns or variables of the data. Recall that you can reference the first ten columns of a matrix A using A[,1:10].

```
pairs(college[2:11])
```



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

```
college$Private <- as.factor(college$Private)

plot(college$Private, college$Outstate)
```



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 %.

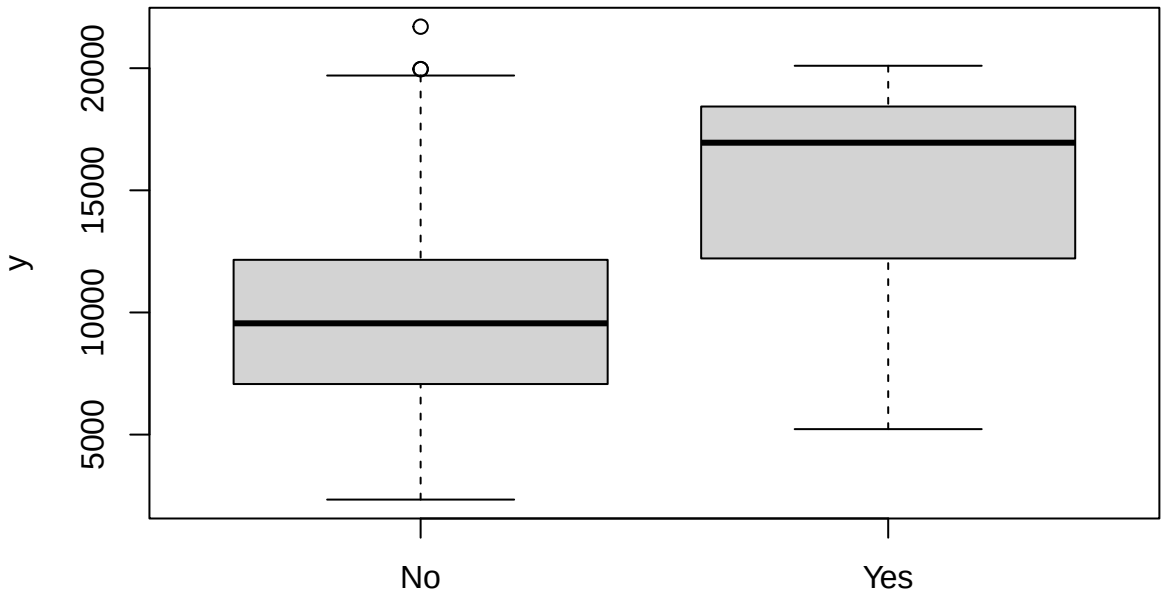
Use the summary() function to see how many elite univer- sities there are. Now use the plot() function to

produce side-by-side boxplots of Outstate versus Elite.

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

summary(college)
```

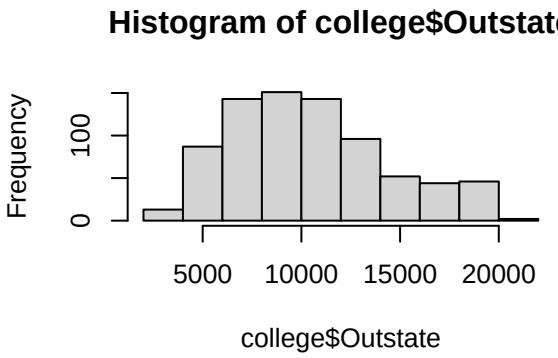
```
##  Private        Apps           Accept          Enroll       Top10perc
##  No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00
##  Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00
##            Median : 1558   Median : 1110   Median : 434   Median :23.00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

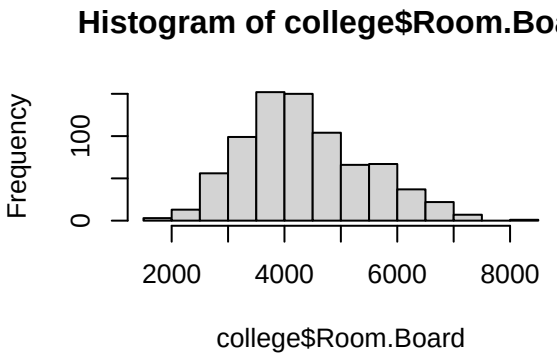
```
plot(college$Elite, college$Outstate)
```

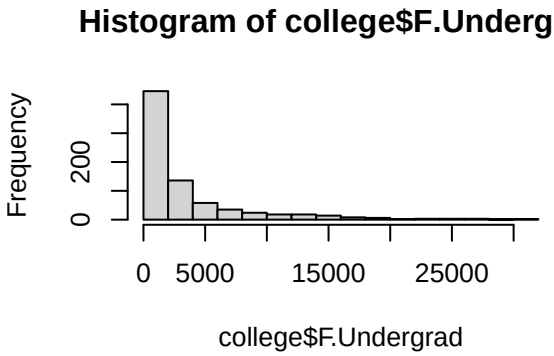


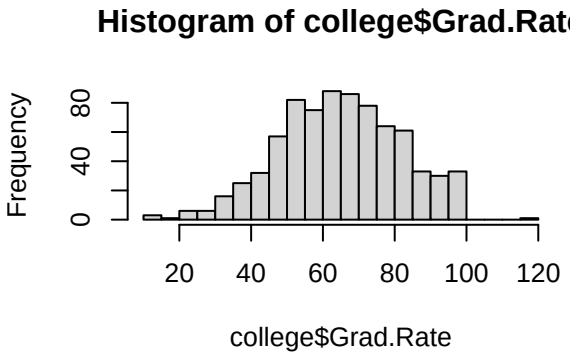
v. Use the hist() function to produce some histograms with differing numbers of bins for a few of the quantitative variables. You may find the command par(mfrow = c(2, 2)) useful: it will divide the print window into four regions so that four plots can be made simultaneously. Modifying the arguments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

```
par(mfrow = c(2,2))

hist(college$Outstate, breaks = 10)

hist(college$Room.Board, breaks = 20)

hist(college$F.Undergrad, breaks = 15)

hist(college$Grad.Rate, breaks = 25)
```









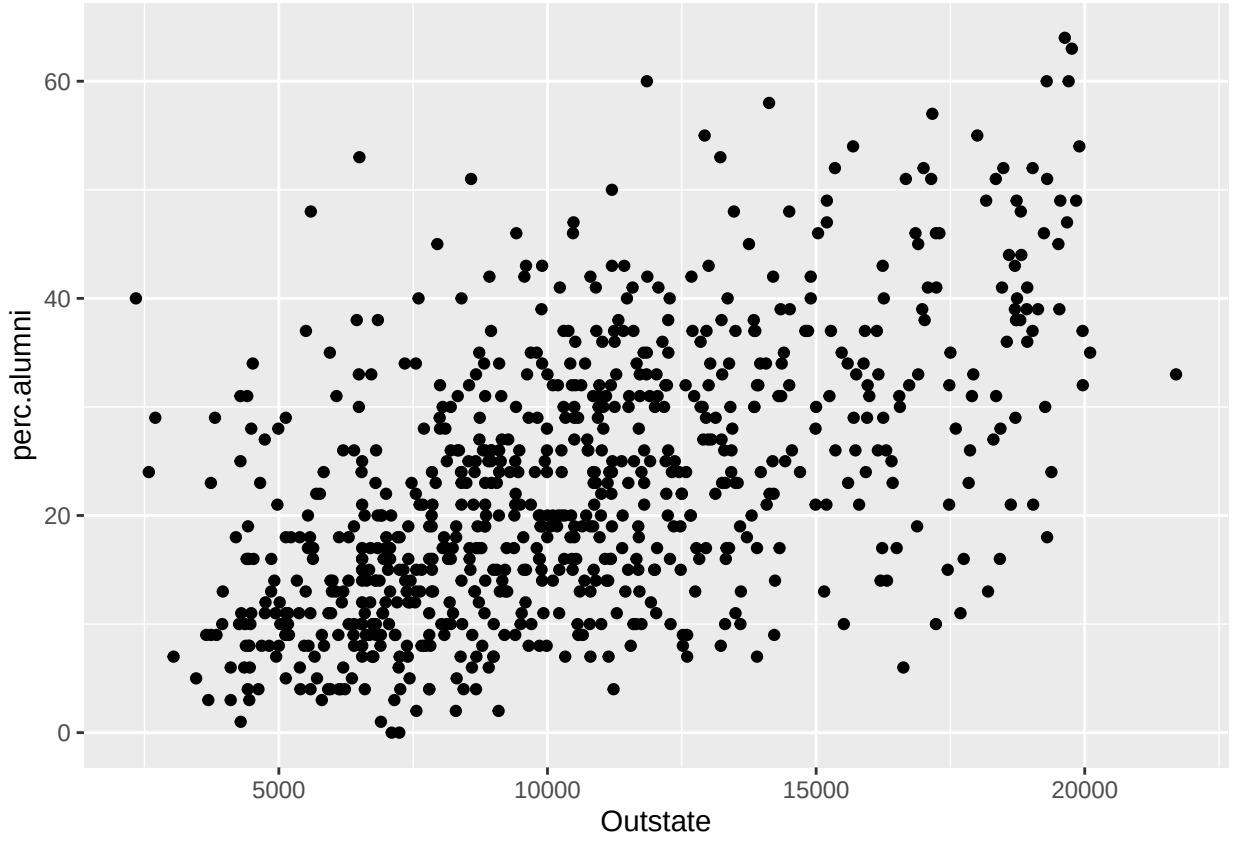
vi. Continue exploring the data, and provide a brief summary of what you discover.

```
summary(college)
```

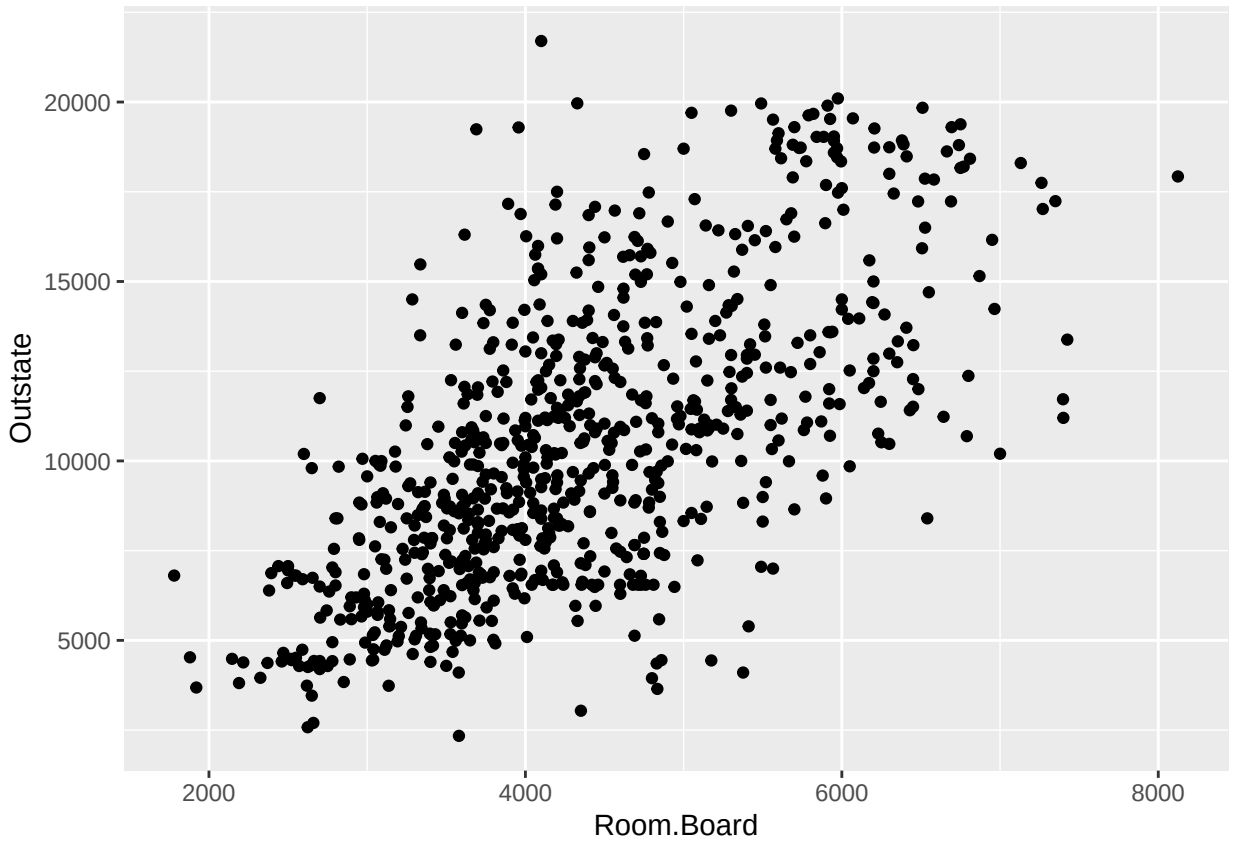
```
##  Private        Apps           Accept          Enroll       Top10perc
##  No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00
##  Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00
##            Median : 1558   Median : 1110   Median : 434   Median :23.00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

```
college %>%
  ggplot(aes(x = Outstate, y = perc.alumni)) +
  geom_point()
```




```
college %>%
  ggplot(aes(x = Room.Board, y = Outstate)) +
  geom_point()
```



After further analysis of the colleges dataset, I have noticed a couple of interesting observations. One beging that it seems like the percent of alumni and cost of outstate schools have no direct correlation as the data can be seen to be scattered around the entire plot, signifying that cost of outstae college may have no explanation as to percentag of alumni of a college. Another interesting observation I found was in the relationship between room & board cost and outstate cost. From the data, a positive relationship is evident as when the room & board cost increases, so does the cost to attend the college.

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.

(a) Which of the predictors are quantitative, and which are quali-tative?

```
auto_data <- read.table("Auto.data", header = TRUE, sep = "", na.strings = "?")
auto_data <- na.omit(auto_data)

str(auto_data)
```

```
## 'data.frame':    392 obs. of  9 variables:
##  $ mpg         : num  18 15 18 16 17 15 14 14 14 15 ...
##  $ cylinders   : int  8 8 8 8 8 8 8 8 8 8 ...
##  $ displacement: num  307 350 318 304 302 429 454 440 455 390 ...
##  $ horsepower  : num  130 165 150 150 140 198 220 215 225 190 ...
##  $ weight      : num  3504 3693 3436 3433 3449 ...
##  $ acceleration: num  12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...
##  $ year        : int  70 70 70 70 70 70 70 70 70 70 ...
##  $ origin      : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ name        : chr  "chevrolet chevelle malibu" "buick skylark 320" "plymouth satellite" "amc rebe
```

```
##  - attr(*, "na.action")= 'omit' Named int [1:5] 33 127 331 337 355
##   ..- attr(*, "names")= chr [1:5] "33" "127" "331" "337" ...
```

The quantitative variables are: mpg, displacement, horsepower, and weight. The qualitative variables are: name, origin, cylinders and year.

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

```
range(auto_data$mpg)
```

```
## [1]  9.0 46.6
```

```
range(auto_data$displacement)
```

```
## [1]  68 455
```

```
range(auto_data$horsepower)
```

```
## [1]  46 230
```

```
range(auto_data$weight)
```

```
## [1] 1613 5140
```

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

```
auto_data %>%
  summarise(
    mgp_mean = mean(mpg),
    displacement_mean = mean(displacement),
    horsepower_mean = mean(horsepower),
    weight_mean = mean(weight),

    mgp_sd = sd(mpg),
    displacement_sd = sd(displacement),
    horsepower_sd = sd(horsepower),
    weight_sd = sd(weight)
  )
```

```
##   mgp_mean displacement_mean horsepower_mean weight_mean   mgp_sd
## 1 23.44592           194.412        104.4694    2977.584 7.805007
##   displacement_sd horsepower_sd weight_sd
## 1         104.644      38.49116  849.4026
```

(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?

```
auto_data_subset <- auto_data[-(10:85), ]

auto_data_subset %>%
  summarise(
    mgp_mean = mean(mpg),
    displacement_mean = mean(displacement),
    horsepower_mean = mean(horsepower),
    weight_mean = mean(weight),

    mgp_sd = sd(mpg),
    displacement_sd = sd(displacement),
    horsepower_sd = sd(horsepower),
```

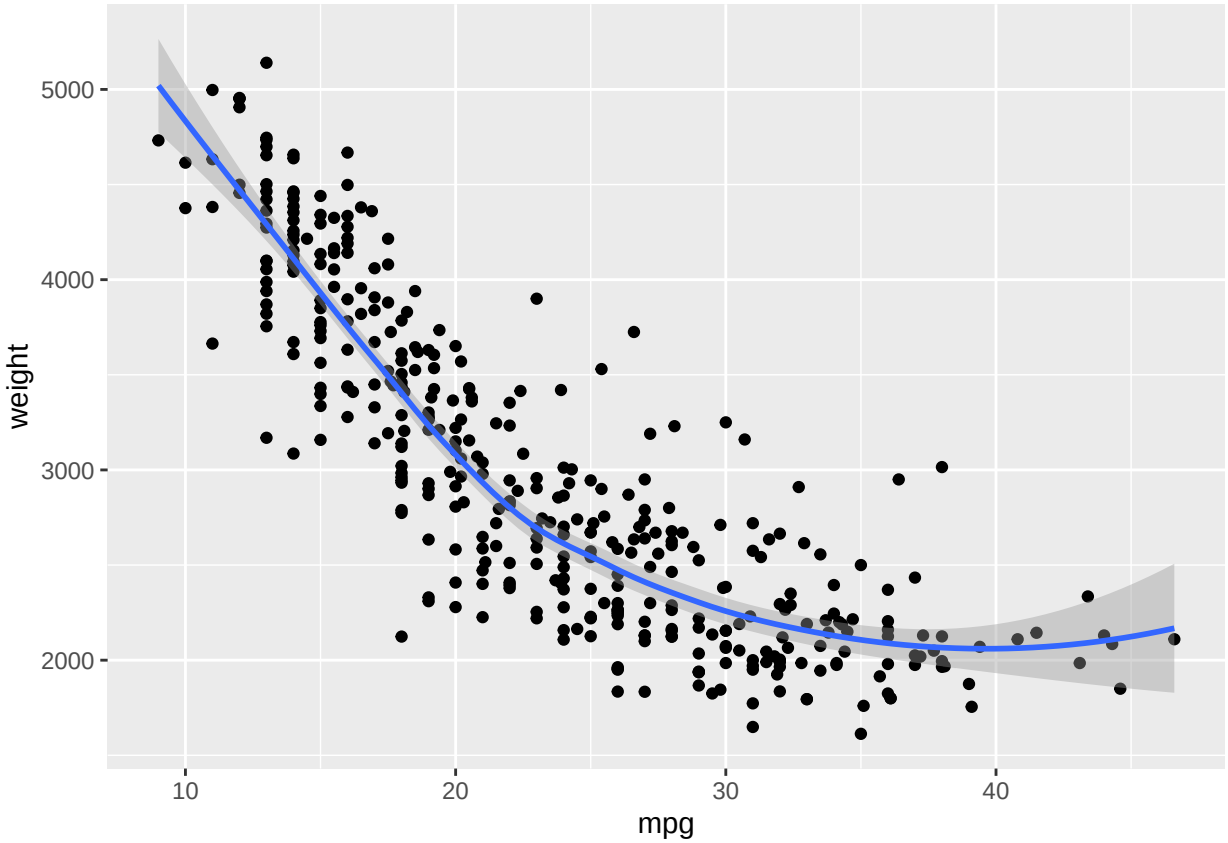
```
    weight_sd = sd(weight)
  )
```

```
##   mgp_mean displacement_mean horsepower_mean weight_mean   mgp_sd
## 1 24.40443          187.2405        100.7215    2935.972 7.867283
##   displacement_sd horsepower_sd weight_sd
## 1        99.67837      35.70885  811.3002
```

(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.

```
auto_data %>%
  ggplot(aes(x = mpg, y = weight)) +
  geom_point() +
  geom_smooth()
```

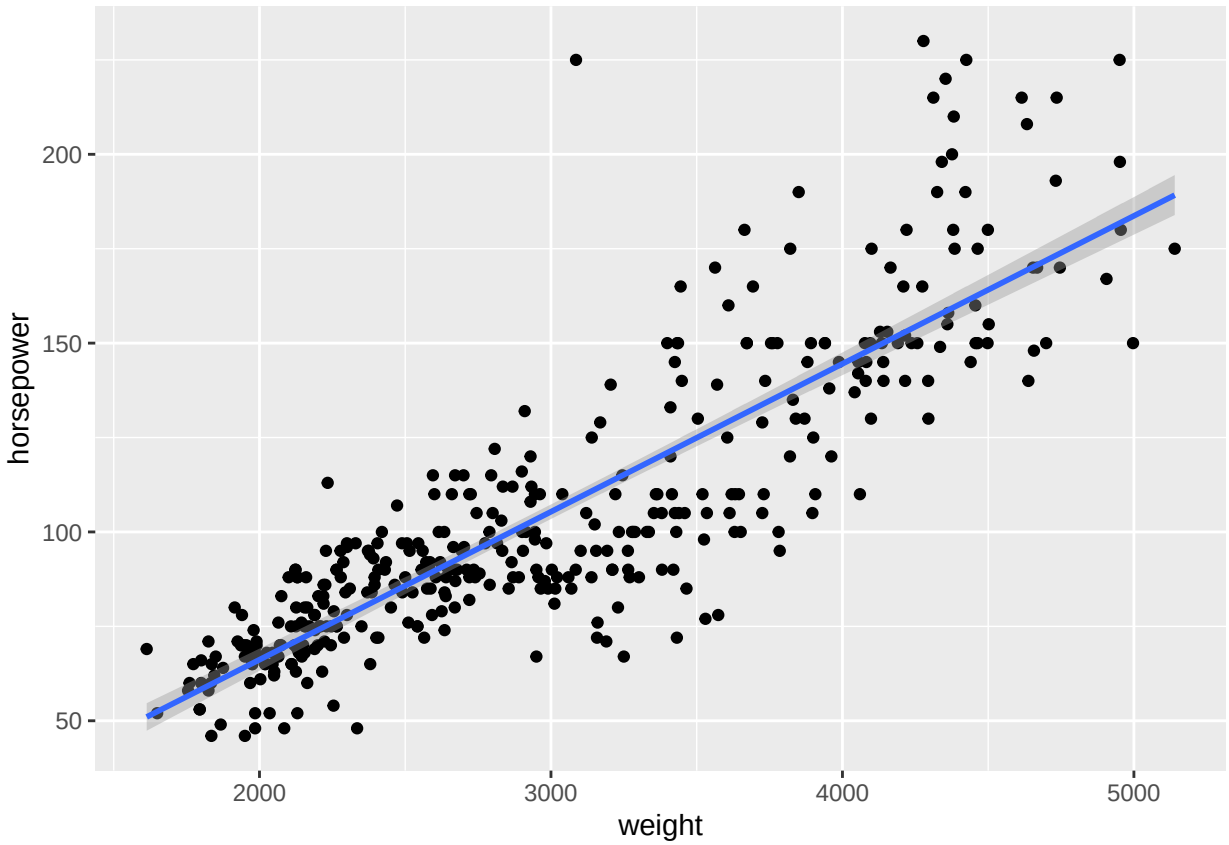
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



From the relationship between miles per gallon (mpg) and the weight of the vehicle, we can see the two variables are negatively related as when mpg increases, the weight of the vehcile tends to decrease. The downward curvature of the plot aids the eye in viewing this negative quadratic relationship. From the fitted line in the plot, the numerous amount of residuals signify the vairance in the relationship.

```
auto_data %>%
  ggplot(aes(x = weight, y = horsepower)) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



From the relationship of weight of vehicle and amount of horsepower it is evident from the plot that as when weight of vehicle increases, the horsepower also increases showing a positve linear relationship. This makes sense as for vehicles that tend to weigh more, more force is required form the engine of the vehicle.

Exercise 10

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

```
library(ISLR2)
Boston <- Boston

nrow(Boston) #row number
```

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

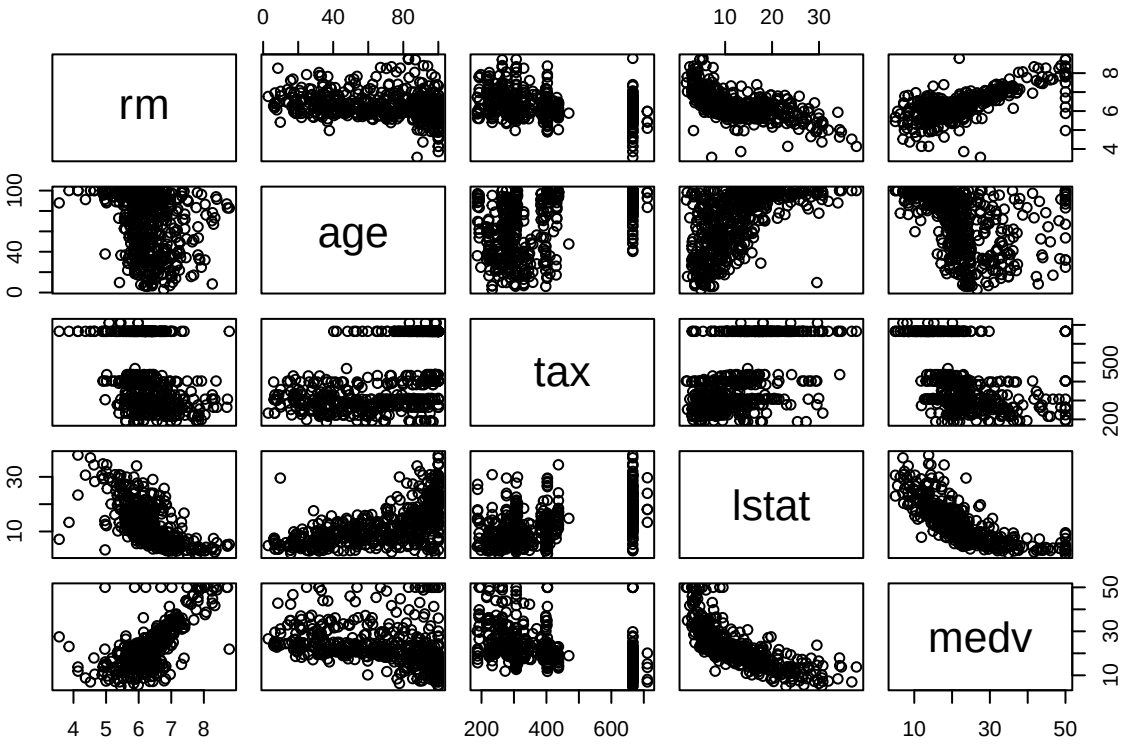
```
ncol(Boston) #col number
```

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

The columns in this dataset represent the varaibles associated with the housing values in Boston. Some of these variables being "rm" which represent the average numeber of rooms dwelling or "age" representing the proportion of owner-occupied units built prior to 1940. The rows in this datasets the observaion of the variables associated with the housing values in Boston. With there being 506 rows, this means that there are 508 observations for each 13 variables associated with the housing values in Boston.

b) Make some pairwise scatterplots of the predictors (columns) in this data set. Describe your findings.

```
pairs(Boston[c(6,7,10,12,13)])
```



From the paired scatterplot of the variables (rm, age, tax, lstat, and medv) we can see some interesiting findings. Such as rm and medv seeming to be positively associated as when average number of rooms dwelling inecreases, median household value tends to increase. The lstat and medv variables show signs of a non-linear relationship as house prices may decrease at a decreasing rate as the percentage of lower status population increases.

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

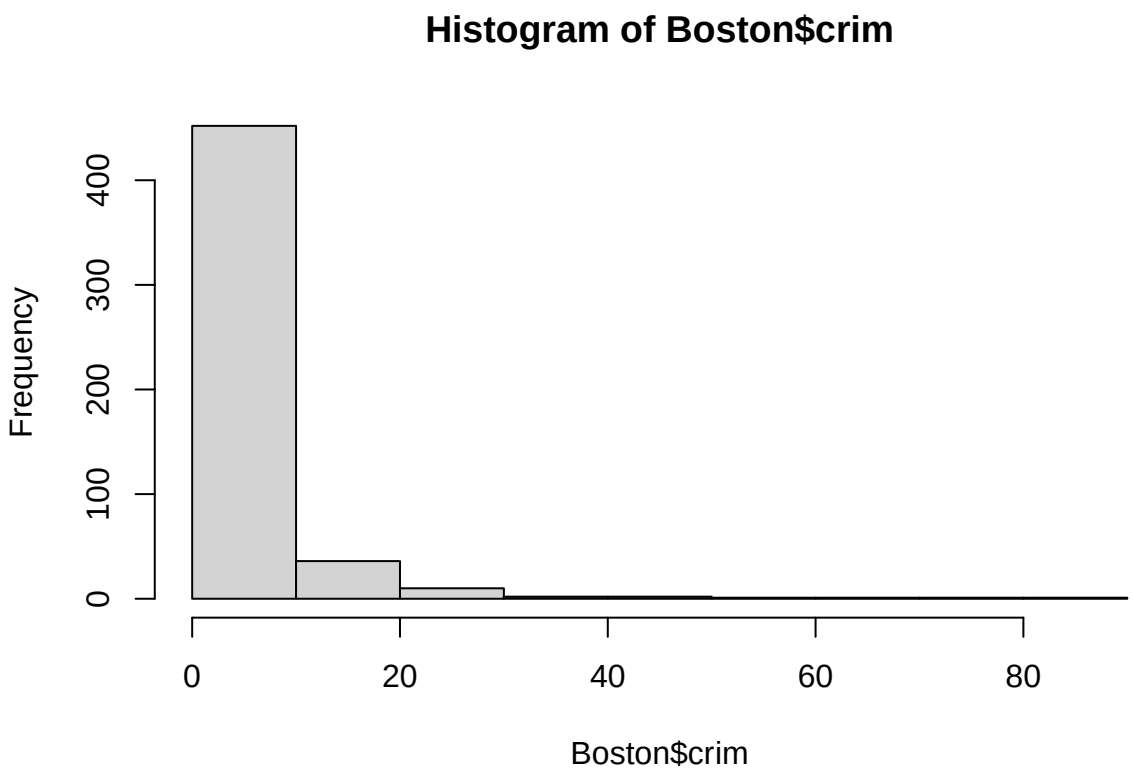
```
correlation <- cor(Boston)

cor_with_crim <- sort(correlation["crim", ], decreasing = TRUE)

cor_with_crim
```

```
##        crim         rad         tax       lstat         nox       indus
##  1.00000000  0.62550515  0.58276431  0.45562148  0.42097171  0.40658341
##         age     ptratio        chas          zn          rm         dis
##  0.35273425  0.28994558 -0.05589158 -0.20046922 -0.21924670 -0.37967009
##        medv
## -0.38830461
```

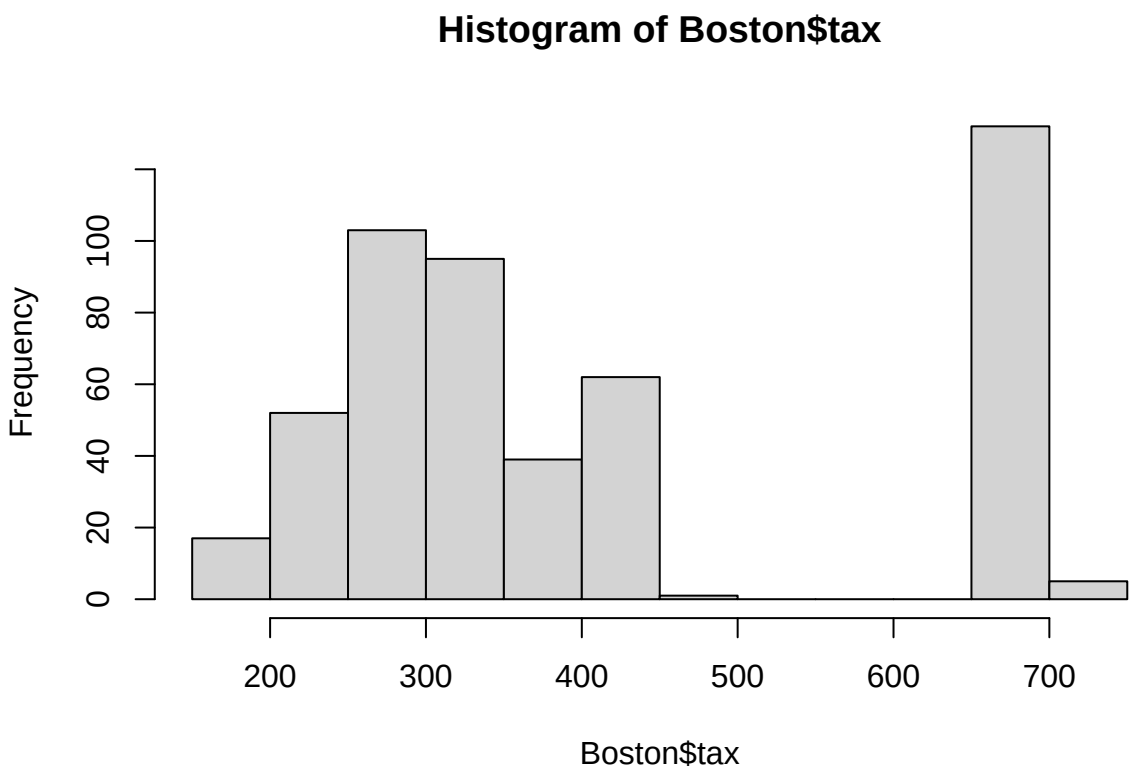
After calculaitng the correlation of variables to per capita crime rate by town (crim) we can see that the only variables that have significant postive correlation with crim are rad (index of accesibility to radial highways), with a correlation of .623 and tax (full value property tax rate per $10,000) with a correlatiion of .583. From this calculation, the data is saying that crime tends to increase as the index of accessibility to radial highways is high for a home and crime tends to increase as property tax rate of the home is high as well.

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

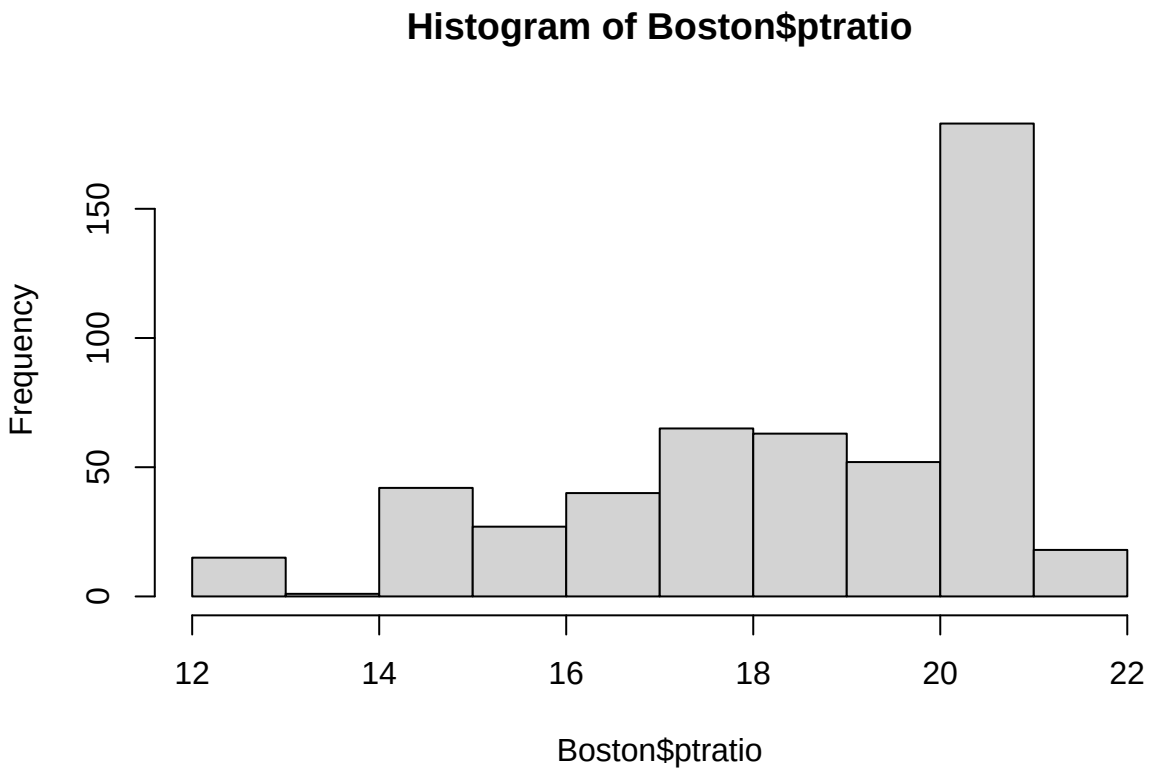
```
hist(Boston$crim)
```




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



```
hist(Boston$ptratio)
```




```
range(Boston$crim)
```

```
## [1]  0.00632 88.97620
```

```
range(Boston$tax)
```

```
## [1] 187 711
```

```
range(Boston$ptratio)
```

```
## [1] 12.6 22.0
```

From the histograms above, it is clear that some of the census tracts of Boston appear to have high tax and pupil-teacher ratios. We can view these high values in the plot as these plots have outliers, points that deviate away from the mean of the of the distribution. Calculating the range also helps us view this variation in cencus tracts as tac having a range of 187 and 711 signifying that a census tract of Boston has a full-value property tax rate of 7.1 million dollars and 1.8 million dollars.

e) How many of the census tracts in this data set bound the Charles river?

```
table(Boston$chas)
```

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

35 of the census tracts in this data set bound the Charles River. In the chas variable, a 1 signifies if tract bounds river.

f) What is the median pupil-teacher ratio among the towns in this data set?

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

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

The median pupil-teacher ratio among the towns in this data set is 19.05.

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

```
min_medv_index <- which.min(Boston$medv)

min_medv_tract <- Boston[min_medv_index, ]
min_medv_tract
```

```
##        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
```

```
summary(Boston)
```

```
##       crim                zn             indus            chas
##  Min.   : 0.00632   Min.   :  0.00   Min.   : 0.46   Min.   :0.00000
##  1st Qu.: 0.08205   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          lstat
##  Min.   : 1.000   Min.   :187.0   Min.   :12.60   Min.   : 1.73
##  1st Qu.: 4.000   1st Qu.:279.0   1st Qu.:17.40   1st Qu.: 6.95
##  Median : 5.000   Median :330.0   Median :19.05   Median :11.36
##  Mean   : 9.549   Mean   :408.2   Mean   :18.46   Mean   :12.65
##  3rd Qu.:24.000   3rd Qu.:666.0   3rd Qu.:20.20   3rd Qu.:16.95
##  Max.   :24.000   Max.   :711.0   Max.   :22.00   Max.   :37.97
##       medv
##  Min.   : 5.00
##  1st Qu.:17.02
##  Median :21.20
##  Mean   :22.53
##  3rd Qu.:25.00
##  Max.   :50.00
```

The census tract of Boston with the lowest median value of owner-occupied homes is the census tract located at the index 399 of the dataset. The values of the other predictors for this census tracts are listed above. After comparing the predictor values for this census tracts, to the overall range for these prdictors, it can be seen that this census tract with the lowest median value of owner-occpied home also has tax, rad, age, ptratio, and lstat values higher than the mean of all the other census tracts.

(h) In this dataset,how many of the census tracts average more than seven rooms per dwelling? More than eight rooms per dwelling? Comment on the census tracts that average more than eight rooms per

dwelling.

```
Boston %>%
  filter(rm > 7) %>%
  summarise(count = n())
```

```
##   count
## 1    64
```

```
Boston %>%
  filter(rm > 8) %>%
  summarise(count = n())
```

```
##   count
## 1    13
```

```
Boston %>%
  filter(rm > 9) %>%
  summarise(count = n())
```

```
##   count
## 1     0
```

64 of the census tracts in this data set average more than seven rooms per dwelling and 13 census tracts average more than 8 rooms per dwelling. There are no census tracts that average more than 8 rooms per dwelling, meaning that in this data set there are nor census tracts in Boston that have homes with 9 rooms.