Individual_assignment1_474154

Yuhan_Xu

2019/8/31

(a)

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

```
library(MASS)
fix(Boston)
```

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

```
dim(Boston)
```

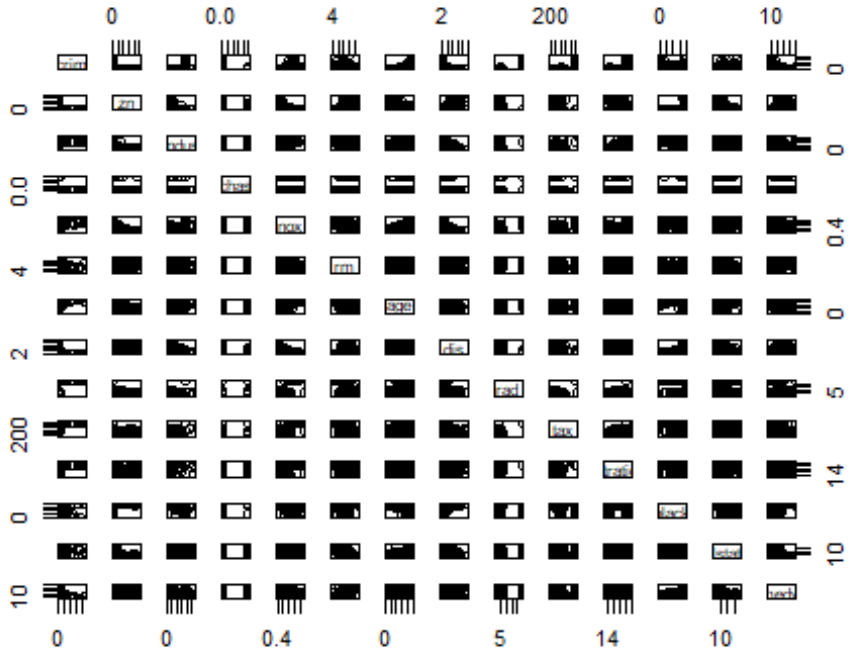
```
## [1] 506  14
```

A: There are 506 rows and 14 columns in this data set. Rows represent the observations, that is the housing values in suburbs of Boston. Columns represent predictors, that is the attributes of these suburbs.

(b)

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

```
pairs(Boston)
```



A: From graphs, we can see that:

Crim correlates with age, dis, rad and tax.

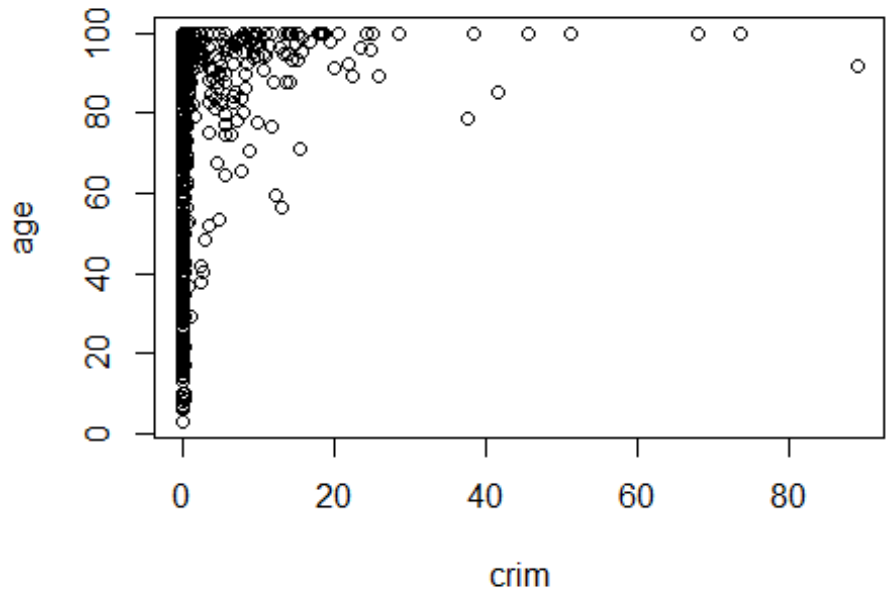
Indus correlates with dis.

Nox correlates with age and dis.

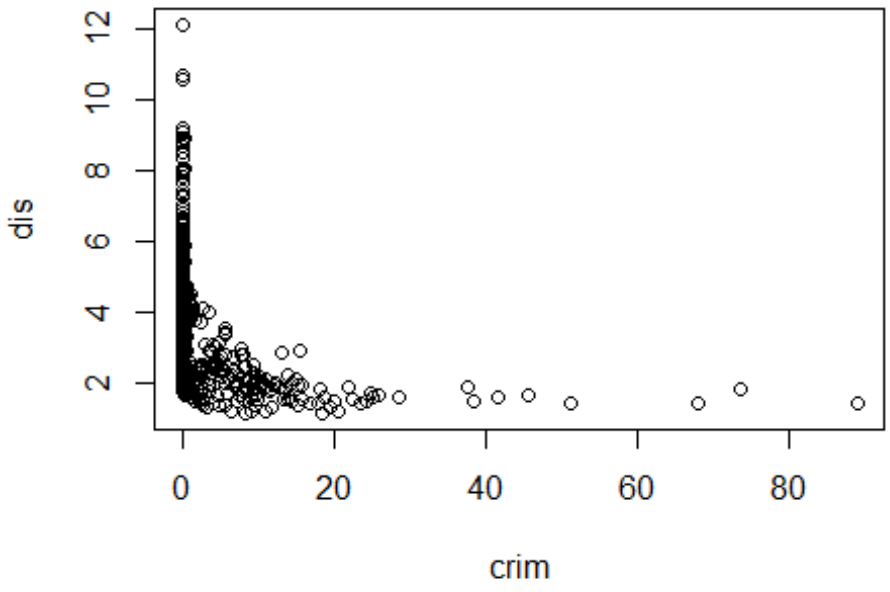
Lstat correlates with medv.

The scatter plots are as follows:

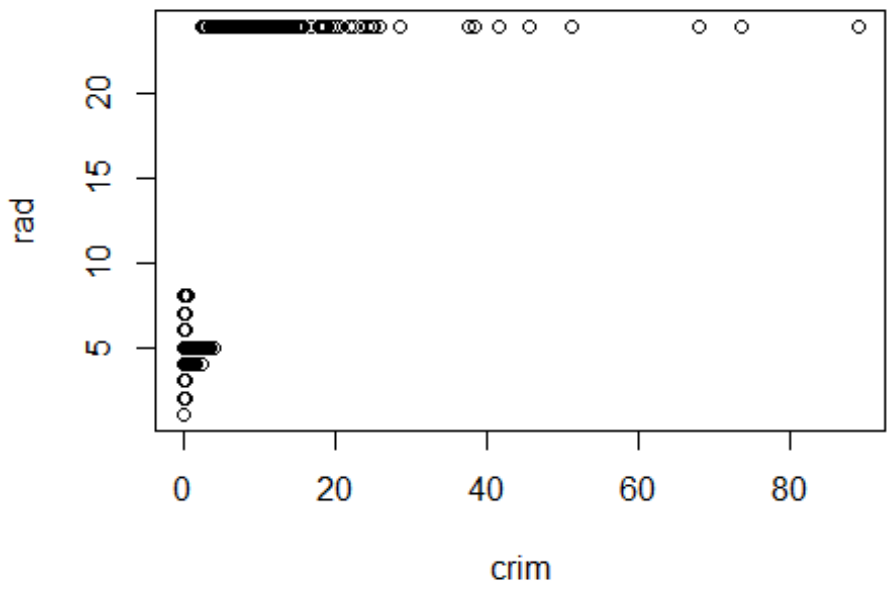
```
library("ISLR")
attach(Boston)
plot(crim, age)
```



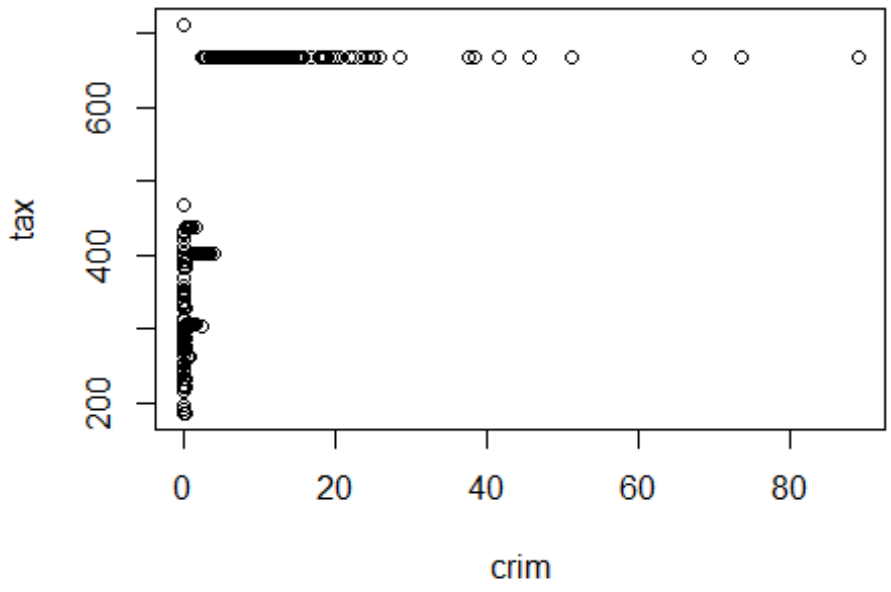

```
plot(crim, dis)
```




```
plot(crim, rad)
```

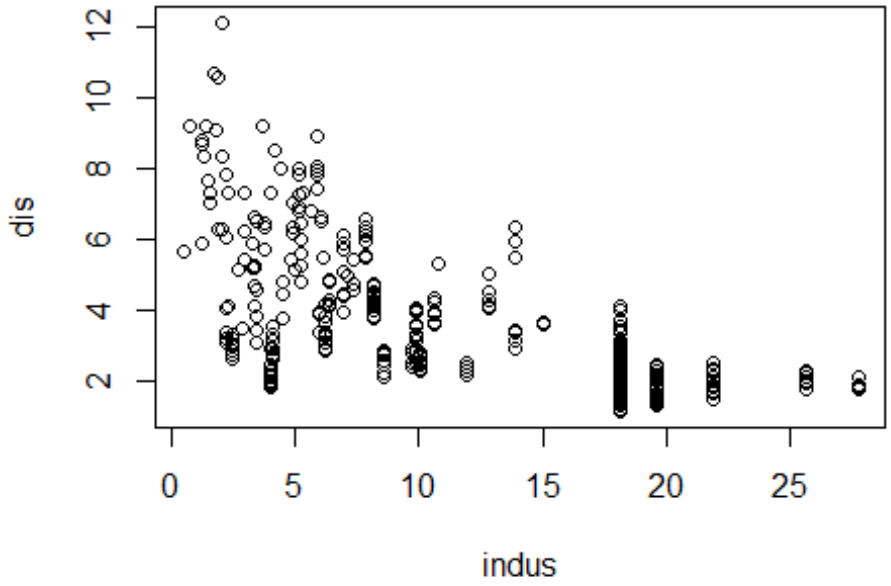



```
plot(crim, tax)
```



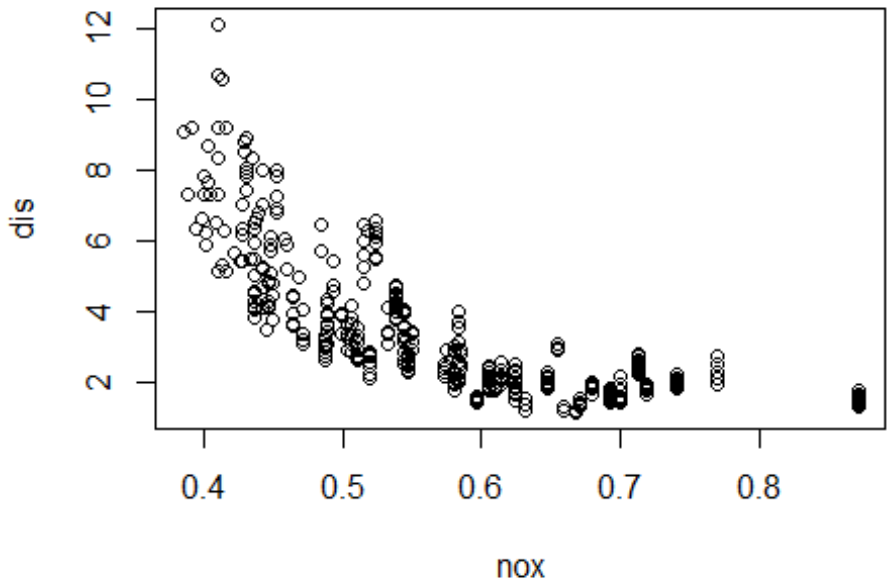
From above scatter plots, we know that crim negatively correlates with dis, and positively correlates with age, rad and tax.

```
plot(indus, dis)
```

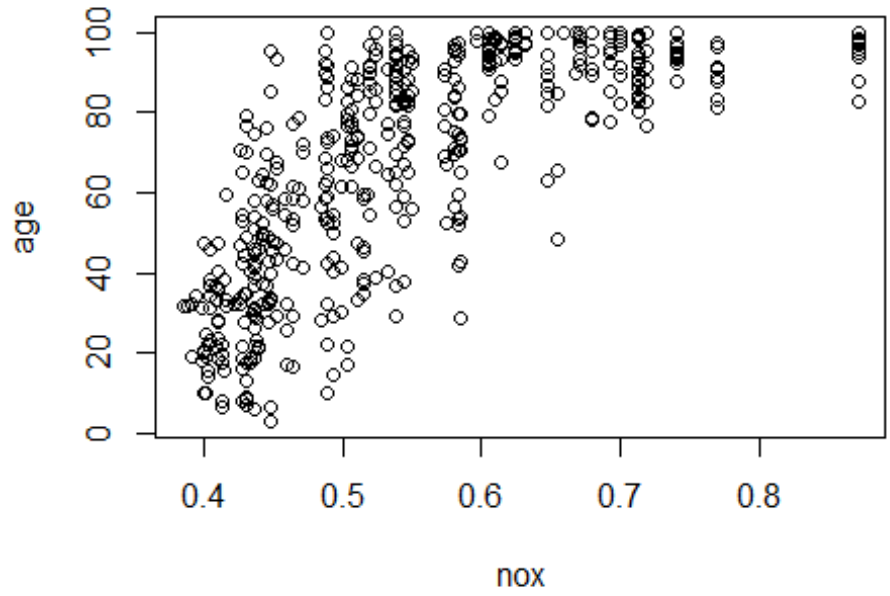


From above scatter plot, we know that indus negatively correlates with dis.

```
plot(nox, dis)
```

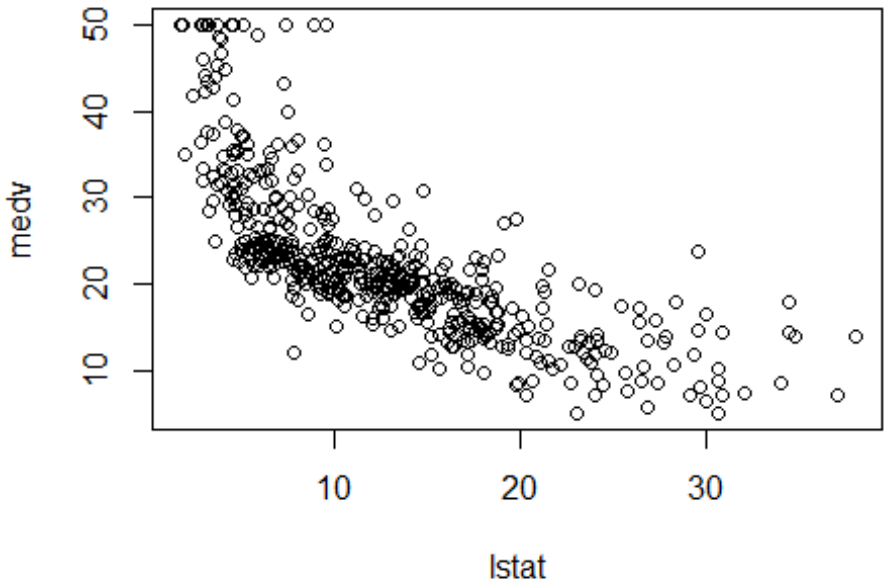

dis
2
4
6
8
10
12
0.4
0.5
0.6
0.7
0.8
nox

```
plot(nox, age)
```


age
0
20
40
60
80
100
0.4
0.5
0.6
0.7
0.8
nox

From above two scatter plots, we know that nox negatively correlates with dis, and positively correlates with age.

```
plot(lstat, medv)
```



From above scatter plot, we know that lstat negatively correlates with medv.

(c)

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

A: From the scatter plots in part (b), we know that age, dis, rad and tax all correlate with crim.

```
cor(crim, age)

## [1] 0.3527343

cor(crim, dis)

## [1] -0.3796701

cor(crim, rad)

## [1] 0.6255051

cor(crim, tax)
```

```
## [1] 0.5827643
```

Specifically, older house results in higher crime rate.

Farther away from the employment centres, crime rate gets lower.

Easier access to radial highways results in higer crime rate.

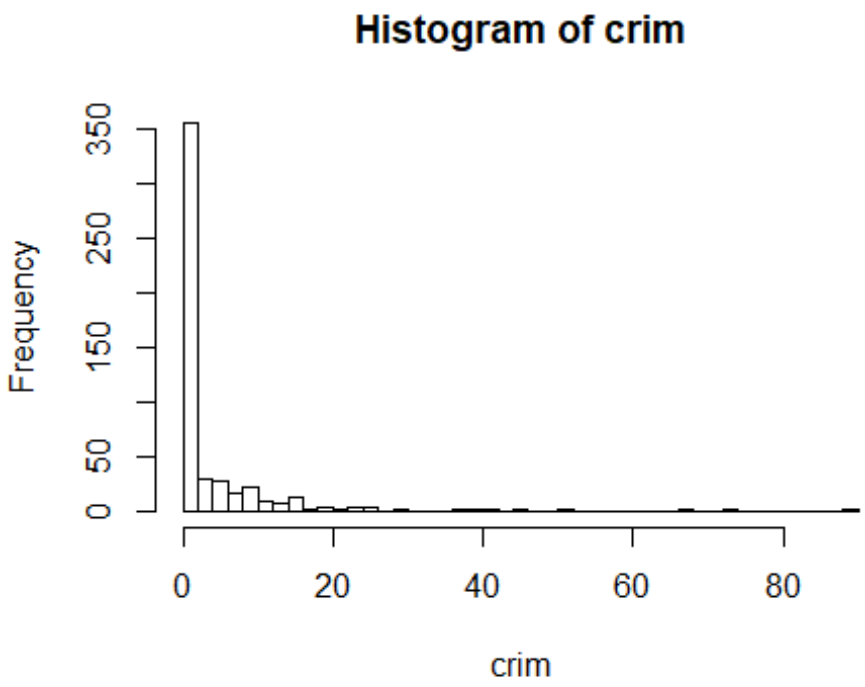
Higher property-tax rate results in higher crime rate.

(d)

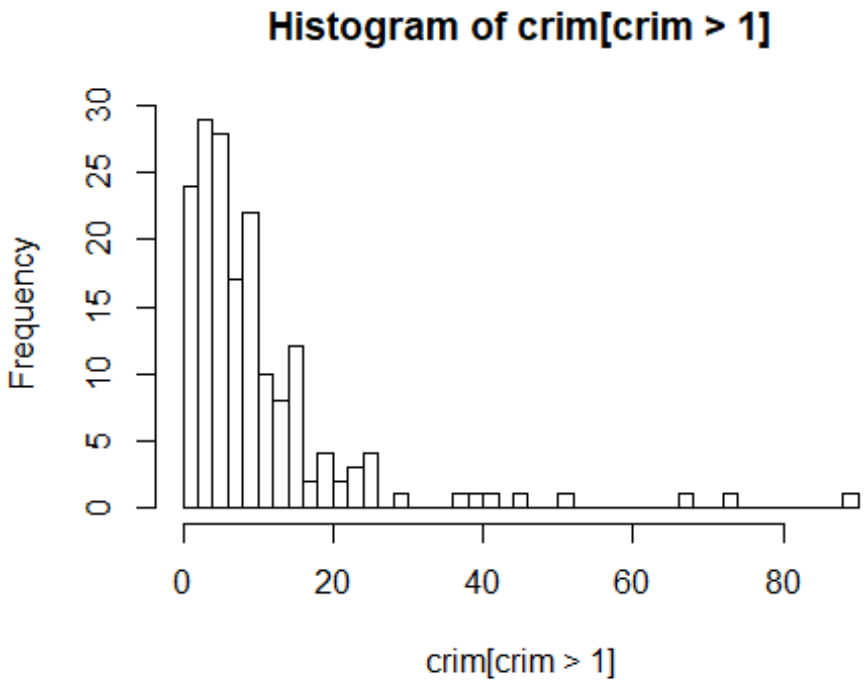
Q: Do any of the suburbs of Boston appear to have particularly high crime rates? Tax rates? Pupil-teacher ratios? Comment on the range of each predictor.

A:

```
hist(crim, breaks = 50)
```

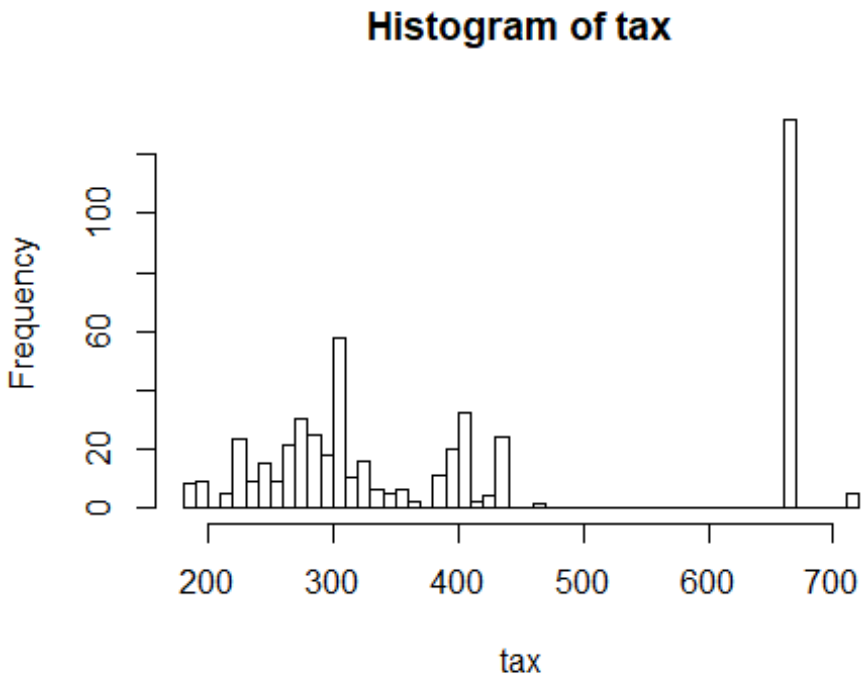



```
hist(crim[crim > 1], breaks = 50)
```



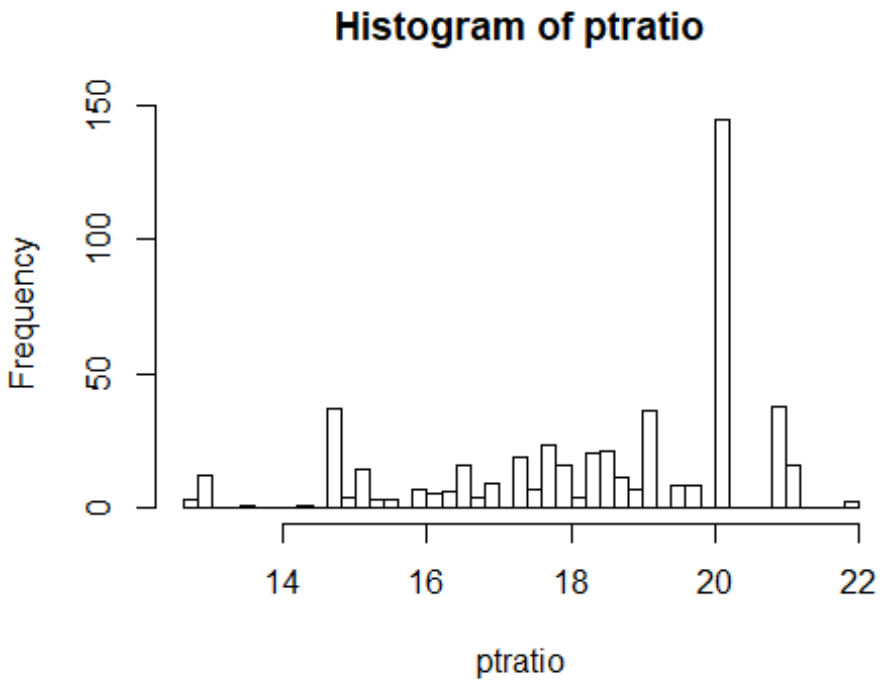
From above histogram, we can see that while most suburbs have a crime rate of around 0, some suburbs have a crime rate of around 40. And some suburbs even have a particularly high crime rate of over 60 and 80.

```
hist(tax, breaks = 50)
```



From above histogram, we can see that while most suburbs have lower tax rate of about 200-400, some suburbs have a particularly high tax rate of over 650.

```
hist(ptratio, breaks = 50)
```



From above histogram, we can see that the pupil-teacher ratio varies from 14 to 22 in different suburbs. There's no particularly high pupil-teacher ratio in certain suburbs. But the proportion of suburbs with a pupil-teacher ratio of around 20 is particularly high.

(e)

Q: How many of the suburbs in this data set bound the Charles river?

```
library("dplyr")
nrow(filter(Boston, chas == 1))
```

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

A: There are 35 suburbs bound the Charles river.

(f)

Q: What is the median pupil-teacher ratio among the towns in this data set?

A:

```
median(ptratio)
```

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

(g)

Q: Which suburb of Boston has the 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.

```
filter(Boston, medv == min(medv))
```

```
##      crim zn indus chas   nox    rm age    dis rad tax ptratio  black
## 1 38.3518  0  18.1    0 0.693 5.453 100 1.4896  24 666    20.2 396.90
## 2 67.9208  0  18.1    0 0.693 5.683 100 1.4254  24 666    20.2 384.97
##   lstat medv
## 1 30.59    5
## 2 22.98    5
```

A: Two suburbs have the lowest median value of of owner-occupied homes. The above two rows contains the values of predictors of these two suburbs.

When comparing to the overall ranges, these two suburbs:

-have higher crime rate (above 3rd quartile)

-have no residential land zoned for lots over 25,000 sq.ft (equal to min)

-have higher proportion of non-retail business acres (equal to 3rd quartile)

-do not bound river

-have higher nitrogen oxides concentration (above 3rd quartile)

-lower average number of rooms per dwelling (below 1st quartile)

-all the owner-occupied units are built prior to 1940 (equal to max)

-are closer to Boston employment centres (below 1st quartile)

-have easiest access to radial highways (equal to max)

-higher property-tax rate (equal to 3rd quartile)

-higher pupil-teacher ratio (equal to 3rd quartile)

-higher proportion of blacks (above 1st quartile and mean, near median)

-larger population of people who have lower status(above 3rd quartile)

```
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
```

(h)

Q: 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.

A:

```
#The number of the suburbs average more than seven rooms per dwelling
nrow(filter(Boston, rm > 7))
```

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

```
#The number of the suburbs average more than eight rooms per dwelling
nrow(filter(Boston, rm > 8))
```

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

64 suburbs have average more than seven rooms per dwelling, 13 have more than eight rooms per dwelling.

These suburbs tend to have lower crime rate, more rooms per dwelling, and smaller population of people who have lower status.

```
summary(filter(Boston, rm > 8))
```

```
##       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(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
```