Individual Assignment 1

Charlie Ling

9/1/2021

```
library(MASS)
View(Boston)
?Boston
```

```
## starting httpd help server ... done
```

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

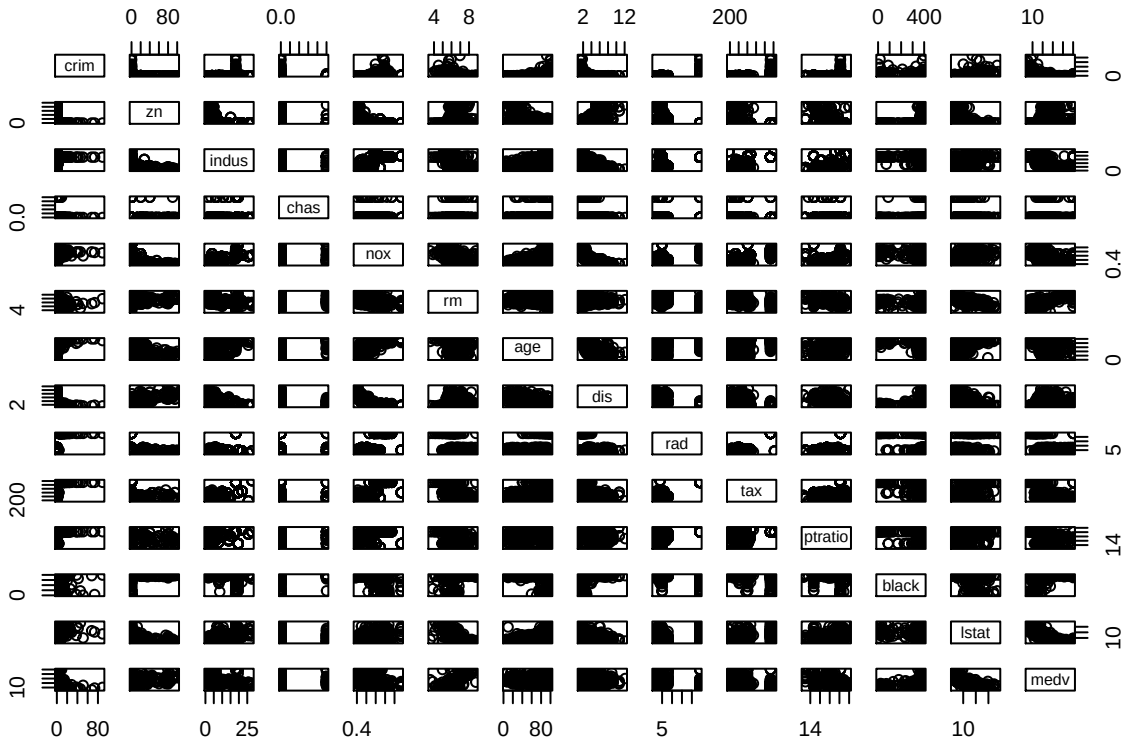
```
dim(Boston)
```

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

```
#506 rows and 14 columns, rows represent observations, columns present features
```

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

```
pairs(Boston)
```



```
m1=lm(nox~dis,data=Boston)
summary(m1)
```

```
##
## Call:
## lm(formula = nox ~ dis, data = Boston)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.12239 -0.05212 -0.01257  0.04391  0.23041
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.715343   0.006796  105.26   <2e-16 ***
## dis         -0.042331   0.001566  -27.03   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07412 on 504 degrees of freedom
## Multiple R-squared:  0.5917, Adjusted R-squared:  0.5909
## F-statistic: 730.4 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
#Buildings closer to  Boston employment centres have higher nitrogen oxides concentration.
```

```
m2=lm(nox~age,data=Boston)
summary(m2)
```

```
##
## Call:
## lm(formula = nox ~ age, data = Boston)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18717 -0.05312 -0.01182  0.04722  0.27407
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.348204   0.009267   37.57   <2e-16 ***
## age         0.003011   0.000125   24.08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07909 on 504 degrees of freedom
## Multiple R-squared:  0.535,  Adjusted R-squared:  0.5341
## F-statistic:   580 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
#Older buildings have higher nitrogen oxides concentration.
```

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

```
summary(lm(crim~tax,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ tax, data = Boston)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.513  -2.738  -0.194   1.065  77.696
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.528369   0.815809  -10.45   <2e-16 ***
## tax          0.029742   0.001847   16.10   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.997 on 504 degrees of freedom
## Multiple R-squared:  0.3396, Adjusted R-squared:  0.3383
## F-statistic: 259.2 on 1 and 504 DF,  p-value: < 2.2e-16
```

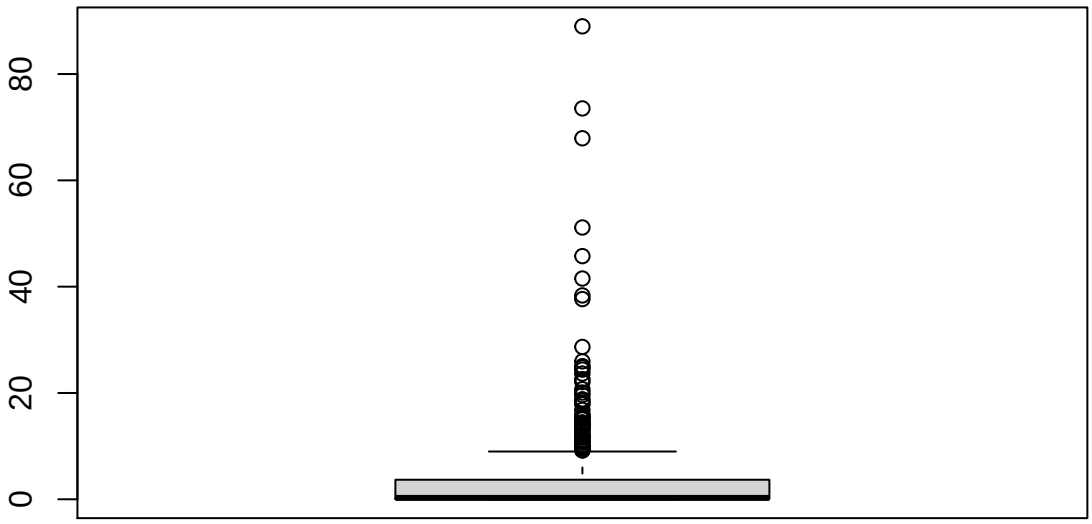
```
#Tax. High tax makes people's life harder, maybe pushes them commit crimes.
```

(d) 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.

```
sapply(Boston,range)
```

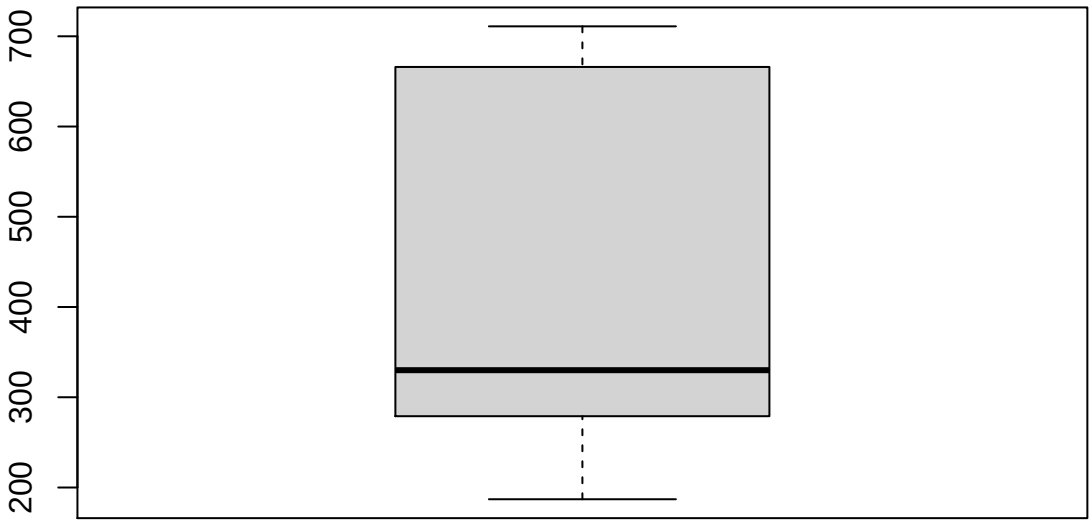
```
##          crim  zn indus chas   nox    rm   age     dis rad tax ptratio  black
## [1,]  0.00632   0  0.46    0 0.385 3.561   2.9  1.1296   1 187    12.6   0.32
## [2,] 88.97620 100 27.74    1 0.871 8.780 100.0 12.1265  24 711    22.0 396.90
##      lstat medv
## [1,]  1.73    5
## [2,] 37.97   50
```

```
boxplot(Boston$crim)
```

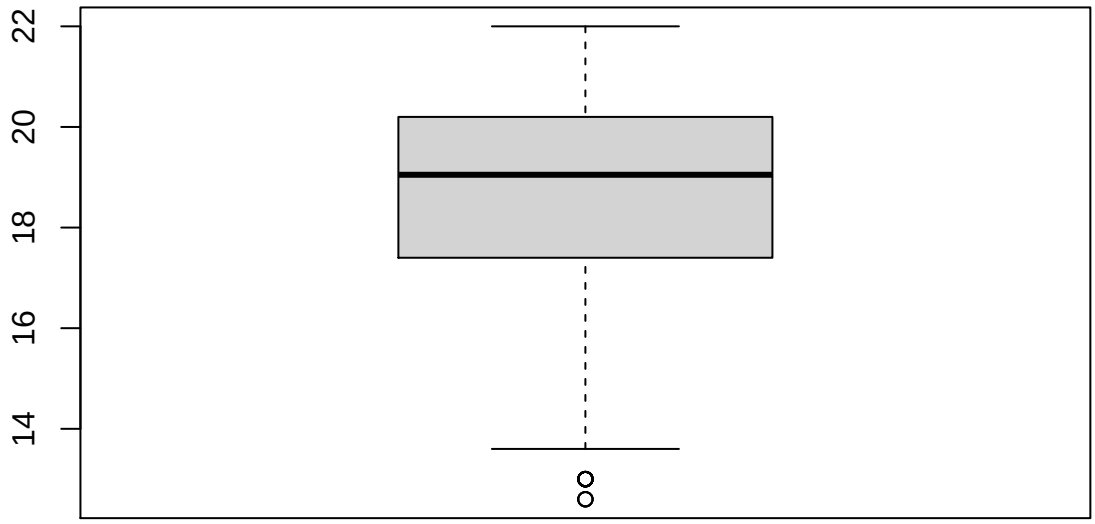



```
#Some suburbs have particularly high crime rates because there are outliers in the crime boxplot.
```

```
boxplot(Boston$tax)
```




```
boxplot(Boston$ptratio)
```



(e) How many of the suburbs in this data set bound the Charles river?

```
sum(Boston$chas==1)
```

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

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

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

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

(g) 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.

```
n=which.min(Boston$medv)
n
```

```
## [1] 399
```

```
Boston[n,1:13]
```

```
##        crim zn indus chas   nox    rm age    dis rad tax ptratio black lstat
## 399 38.3518  0  18.1    0 0.693 5.453 100 1.4896  24 666    20.2 396.9 30.59
```

```
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          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
```

```
# Most of those values(such as crim, indus, nox, age, rad, tax, ptratio, black,
# lstat)are above the 3rd quantile of overall ranges for those predictors, while rm,
# dis are reletively lower.
```

(h) 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.

```
sum(Boston$rm>7)
```

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

```
sum(Boston$rm>8)
```

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

```
nums=which(Boston$rm>8)
Boston[nums,1:14]
```

```
##        crim zn indus chas    nox    rm  age    dis rad tax ptratio  black lstat
## 98  0.12083  0  2.89    0 0.4450 8.069 76.0 3.4952   2 276    18.0 396.90  4.21
## 164 1.51902  0 19.58    1 0.6050 8.375 93.9 2.1620   5 403    14.7 388.45  3.32
## 205 0.02009 95  2.68    0 0.4161 8.034 31.9 5.1180   4 224    14.7 390.55  2.88
## 225 0.31533  0  6.20    0 0.5040 8.266 78.3 2.8944   8 307    17.4 385.05  4.14
## 226 0.52693  0  6.20    0 0.5040 8.725 83.0 2.8944   8 307    17.4 382.00  4.63
## 227 0.38214  0  6.20    0 0.5040 8.040 86.5 3.2157   8 307    17.4 387.38  3.13
## 233 0.57529  0  6.20    0 0.5070 8.337 73.3 3.8384   8 307    17.4 385.91  2.47
## 234 0.33147  0  6.20    0 0.5070 8.247 70.4 3.6519   8 307    17.4 378.95  3.95
## 254 0.36894 22  5.86    0 0.4310 8.259  8.4 8.9067   7 330    19.1 396.90  3.54
## 258 0.61154 20  3.97    0 0.6470 8.704 86.9 1.8010   5 264    13.0 389.70  5.12
## 263 0.52014 20  3.97    0 0.6470 8.398 91.5 2.2885   5 264    13.0 386.86  5.91
## 268 0.57834 20  3.97    0 0.5750 8.297 67.0 2.4216   5 264    13.0 384.54  7.44
```

```
## 365 3.47428  0 18.10    1 0.7180 8.780 82.9 1.9047  24 666    20.2 354.55  5.29
##     medv
## 98  38.7
## 164 50.0
## 205 50.0
## 225 44.8
## 226 50.0
## 227 37.6
## 233 41.7
## 234 48.3
## 254 42.8
## 258 50.0
## 263 48.8
## 268 50.0
## 365 21.9
```

```
# Those suburbs that average more than eight rooms per dwelling have low crime
# rate, low lstat and high median value of owner-occupied homes.
```