

personal assignment 1

```
#student ID: 474084
#student name: Yiqing Zhang

#(a)
library(MASS)
?Boston
nrow(Boston)

## [1] 506
ncol(Boston)

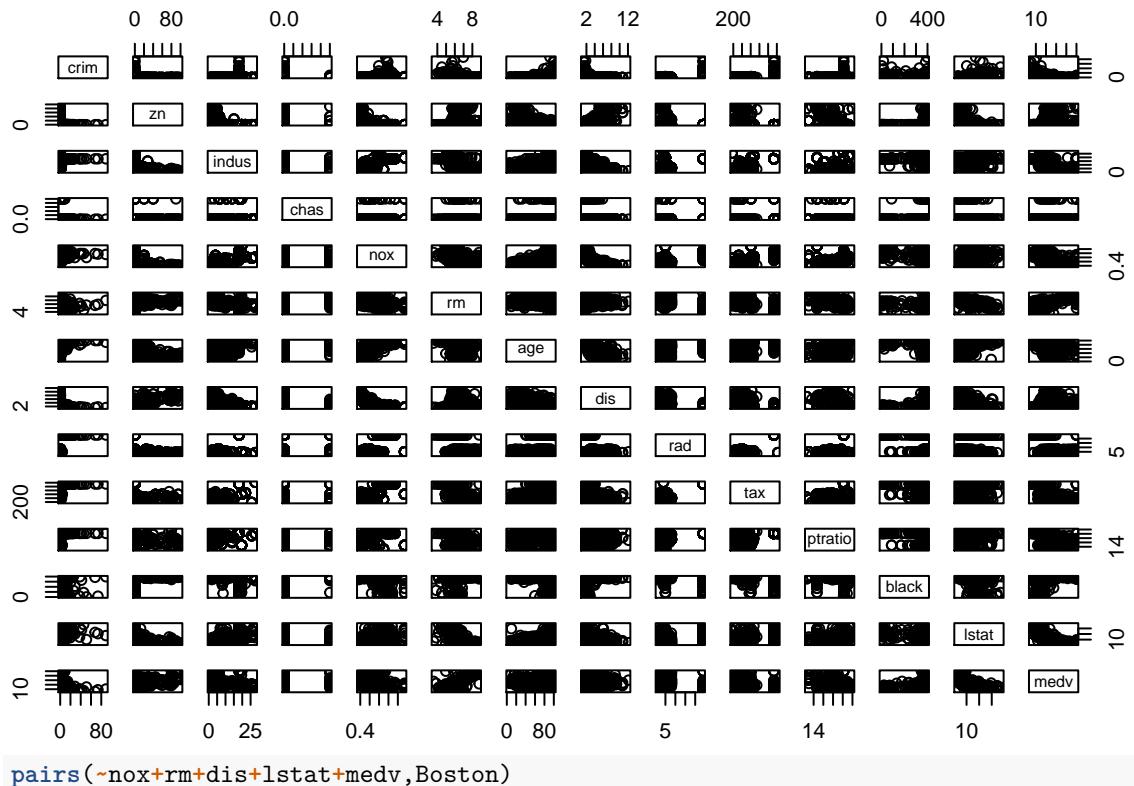
## [1] 14
head(Boston)

##      crim  zn  indus  chas   nox    rm   age    dis rad tax ptratio  black
## 1 0.00632 18  2.31     0 0.538 6.575 65.2 4.0900    1 296 15.3 396.90
## 2 0.02731  0  7.07     0 0.469 6.421 78.9 4.9671    2 242 17.8 396.90
## 3 0.02729  0  7.07     0 0.469 7.185 61.1 4.9671    2 242 17.8 392.83
## 4 0.03237  0  2.18     0 0.458 6.998 45.8 6.0622    3 222 18.7 394.63
## 5 0.06905  0  2.18     0 0.458 7.147 54.2 6.0622    3 222 18.7 396.90
## 6 0.02985  0  2.18     0 0.458 6.430 58.7 6.0622    3 222 18.7 394.12
##      lstat medv
## 1 4.98 24.0
## 2 9.14 21.6
## 3 4.03 34.7
## 4 2.94 33.4
## 5 5.33 36.2
## 6 5.21 28.7

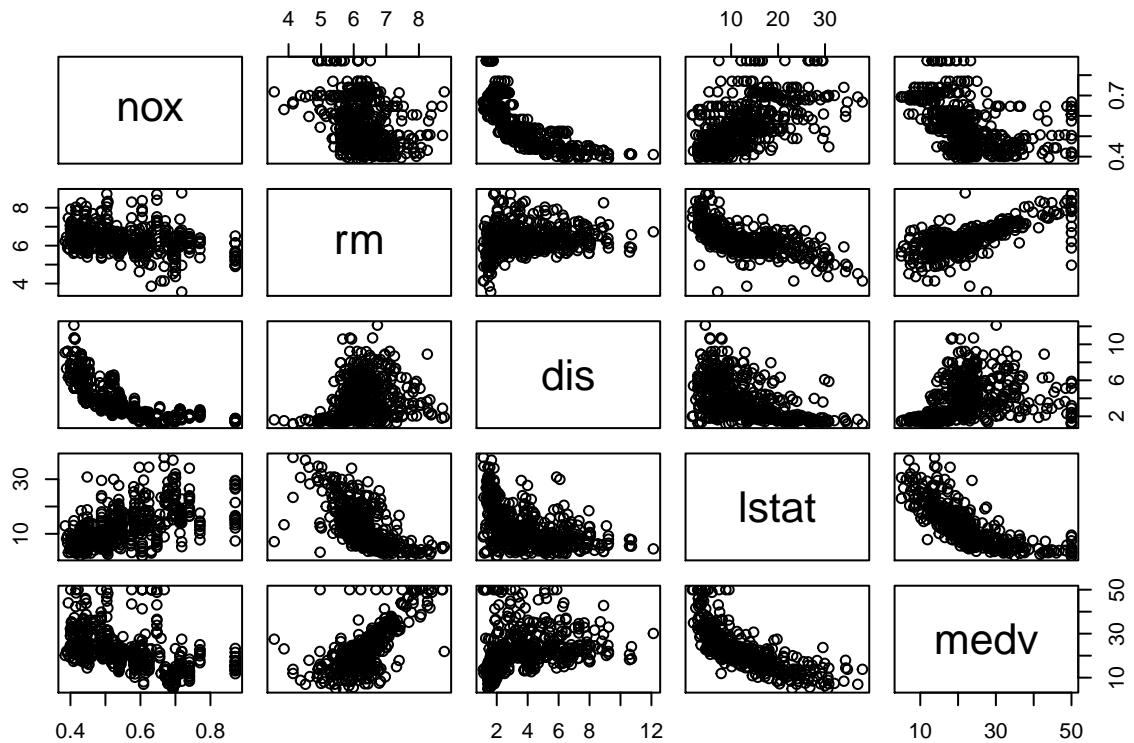
#There are 14 columns and 506 rows.
#And the rows represent 506 different suburbs in Boston, and the columns represent 14 different
#features of every area, which are as follows.

#crim: per capita crime rate by town.
#zn: proportion of residential land zoned for lots over 25,000 sq.ft.
#indus: proportion of non-retail business acres per town.
#chas: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).
#nox: nitrogen oxides concentration (parts per 10 million).
#rm: average number of rooms per dwelling.
#age: proportion of owner-occupied units built prior to 1940.
#dis: weighted mean of distances to five Boston employment centres.
#rad: index of accessibility to radial highways.
#tax: full-value property-tax rate per \$10,000.
#ptratio: pupil-teacher ratio by town.
#black: 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town.
#lstat: lower status of the population (percent).
#medv: median value of owner-occupied homes in \$1000s.

#(b)
pairs(Boston)
```

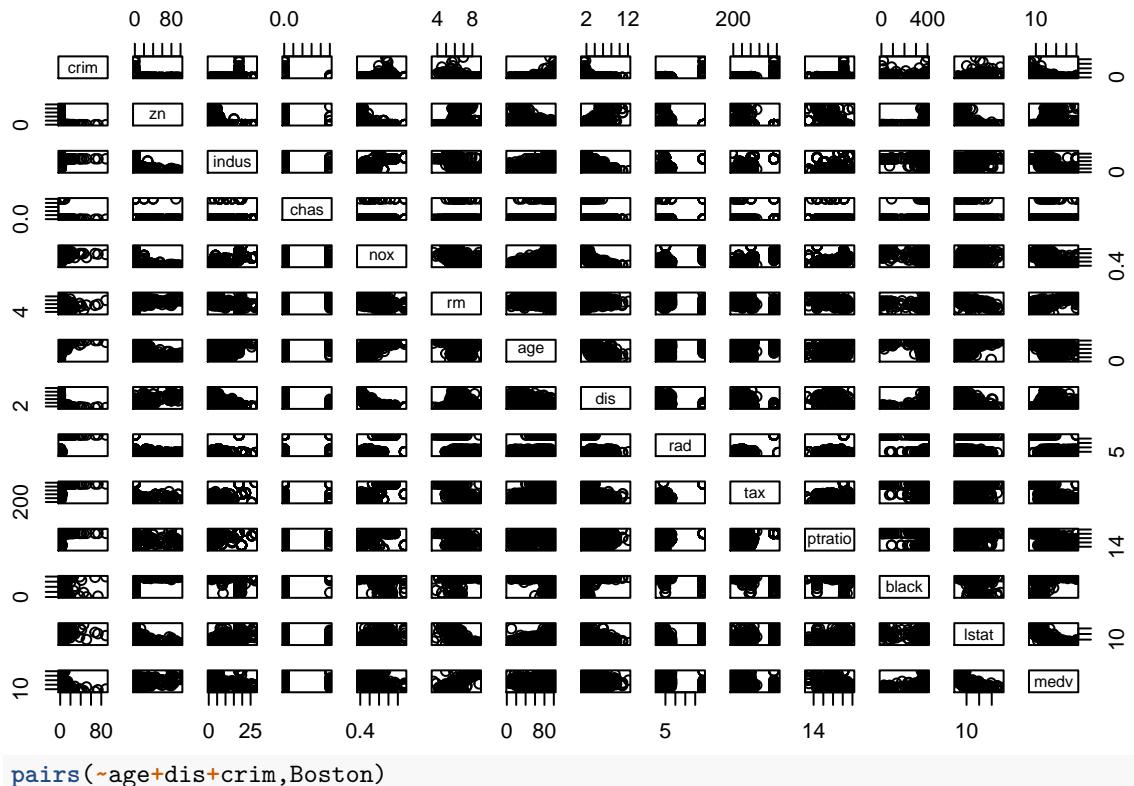


```
pairs(~nox+rm+dis+lstat+medv, Boston)
```

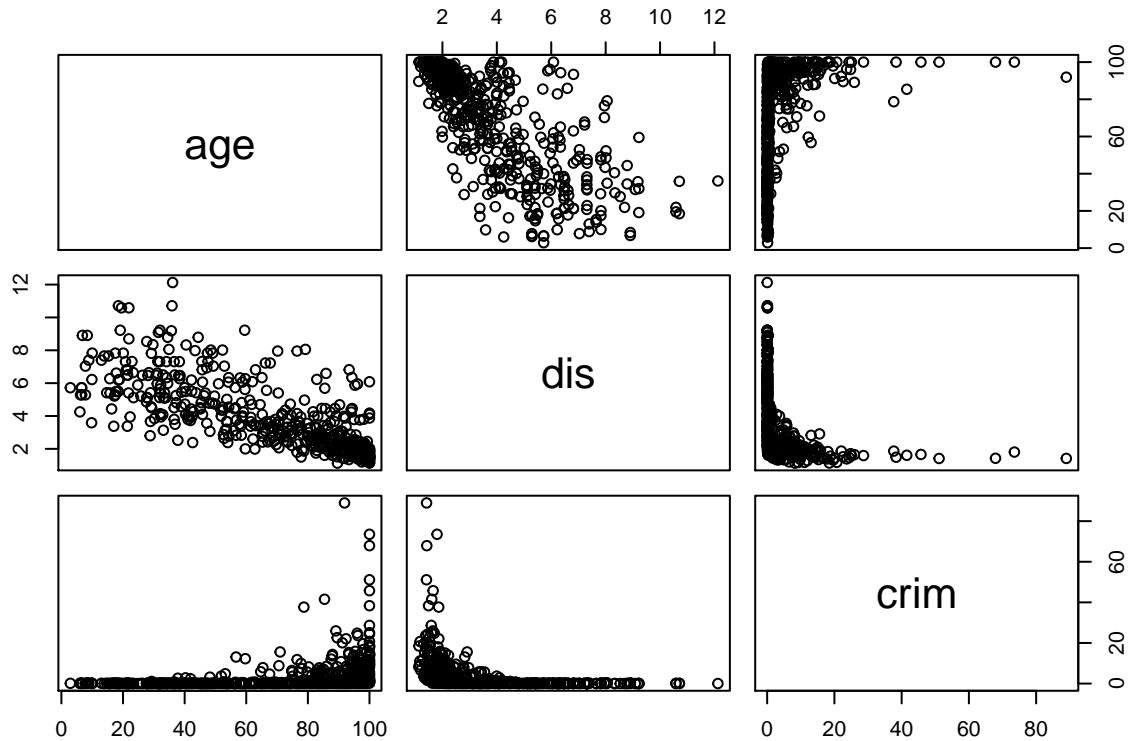


#According to the scatterpoint graphs, there might be a negative correlation between nox and dis; lstat and medv. There might be a positive correlation between medv and rm.

```
#(c)
pairs(Boston)
```



```
pairs(~age+dis+crim,Boston)
```



```
# According to the scatterplot graphs, the age and the distance of them are associated with
#the crime rate. When the proportion of owner-occupied units built prior
#to 1940 goes up to a certain number, the crime rate also goes up rapidly.
#Also, when the weighted mean of distances to five Boston employment
#centers is very small, the crime rate is extremely high, when the
```

```
#distances goes up, the crime rate almost becomes a line at a low value.
```

```
 #(d)
```

```
sapply(Boston[,c(1,10,11)],range)
```

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

```
#There is a suburb having extremely high crime rate than
#others in Boston, which reaches 88.97.
#And the range of tax rate is 187 to 711,
#the pupil-teacher ratio can be up to 22, compared to
#the lowest 12.6.
```

```
 #(e)
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following object is masked from 'package:MASS':
##   select
## The following objects are masked from 'package:stats':
##   filter, lag
## The following objects are masked from 'package:base':
##   intersect, setdiff, setequal, union
chas_river=filter(Boston,chas==1)
nrow(chas_river)
```

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

```
#There are 35 suburbs bond the Charles river.
```

```
 #(f)
```

```
attach(Boston)
median(ptratio)
```

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

```
#The median of pupil-teacher ratio by town is 19.05.
```

```
 #(g)
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
## The following object is masked from 'package:base':
##   date
library(dplyr)
min(medv)
```

```

## [1] 5
medv_low=Boston[medv==5,]
medv_low

##      crim zn indus chas   nox     rm age     dis rad tax ptratio black
## 399 38.3518 0 18.1    0 0.693 5.453 100 1.4896 24 666    20.2 396.90
## 406 67.9208 0 18.1    0 0.693 5.683 100 1.4254 24 666    20.2 384.97
## lstat medv
## 399 30.59    5
## 406 22.98    5
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

#suburb 399 and 406 have the lowest median value of
#owner-occupied homes in $1000.
#and other predictors of suburb 399 and 406 show that these
#two suburbs have higher crime rates, and higher proportion of
#old houses and non-retail business acres per town. those suburbs
#have rather high proportion of blacks by town and lower status of
#the population.

nrow(filter(Boston,rm>7))

## [1] 64
nrow(filter(Boston,rm>8))

```

```

## [1] 13
rm_high=Boston[rm>8,]
rm_high

##      crim zn indus chas    nox     rm   age    dis rad tax ptratio black
## 98 0.12083 0 2.89 0 0.4450 8.069 76.0 3.4952 2 276 18.0 396.90
## 164 1.51902 0 19.58 1 0.6050 8.375 93.9 2.1620 5 403 14.7 388.45
## 205 0.02009 95 2.68 0 0.4161 8.034 31.9 5.1180 4 224 14.7 390.55
## 225 0.31533 0 6.20 0 0.5040 8.266 78.3 2.8944 8 307 17.4 385.05
## 226 0.52693 0 6.20 0 0.5040 8.725 83.0 2.8944 8 307 17.4 382.00
## 227 0.38214 0 6.20 0 0.5040 8.040 86.5 3.2157 8 307 17.4 387.38
## 233 0.57529 0 6.20 0 0.5070 8.337 73.3 3.8384 8 307 17.4 385.91
## 234 0.33147 0 6.20 0 0.5070 8.247 70.4 3.6519 8 307 17.4 378.95
## 254 0.36894 22 5.86 0 0.4310 8.259 8.4 8.9067 7 330 19.1 396.90
## 258 0.61154 20 3.97 0 0.6470 8.704 86.9 1.8010 5 264 13.0 389.70
## 263 0.52014 20 3.97 0 0.6470 8.398 91.5 2.2885 5 264 13.0 386.86
## 268 0.57834 20 3.97 0 0.5750 8.297 67.0 2.4216 5 264 13.0 384.54
## 365 3.47428 0 18.10 1 0.7180 8.780 82.9 1.9047 24 666 20.2 354.55
##      lstat medv
## 98 4.21 38.7
## 164 3.32 50.0
## 205 2.88 50.0
## 225 4.14 44.8
## 226 4.63 50.0
## 227 3.13 37.6
## 233 2.47 41.7
## 234 3.95 48.3
## 254 3.54 42.8
## 258 5.12 50.0
## 263 5.91 48.8
## 268 7.44 50.0
## 365 5.29 21.9

#There are 64 suburbs having average number of rooms per
#dwelling more than 7, and 13 suburbs more than 8.
#The suburbs with average number more than 8 are 98,164,05,225,226,227,233,234,254,258

```