Take Home Exam

Michael Zhang

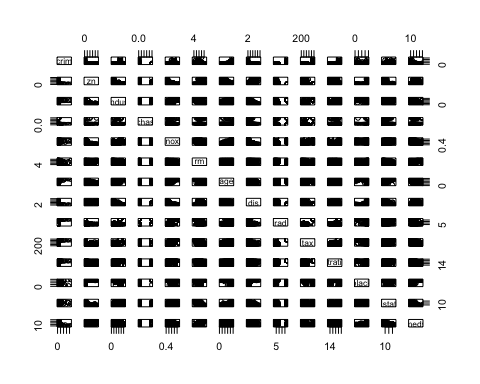
# Book Problems

## Chapter 2, #10

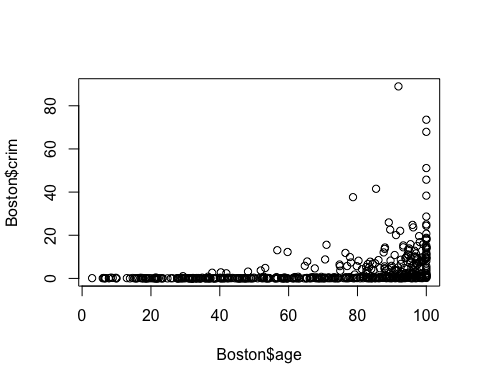
library(MASS)

There are 506 rows and 14 columns. The rows represent observations (housing values in Boston suburbs), and the columns contain features (e.g. median house value, per capita crime rate, etc.).

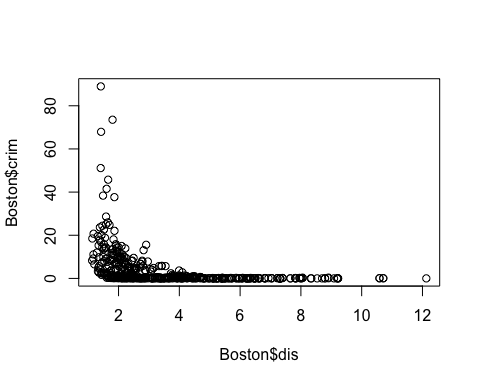
pairs(Boston)

 We see a fairly strong negative correlation between lstat and medv. We also see a fairly strong positive correlation between rm and medv and a similarly negative correlation between rm and lstat. Additionally, we see a negative correlation between dis and lstat. There are other correlations between the predictors, but these are some of the strongest that we see from the pairwise scatterplots.

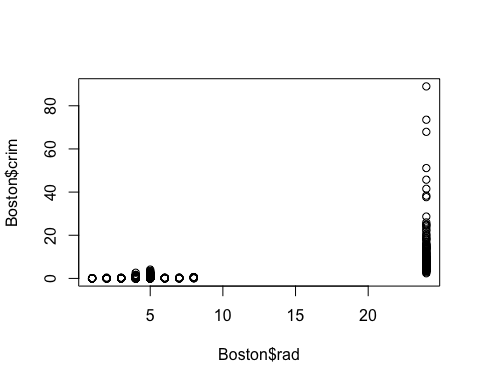
plot(Boston$age, Boston$crim)



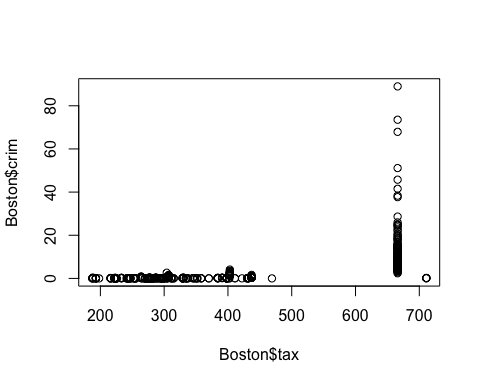
plot(Boston$dis, Boston$crim)



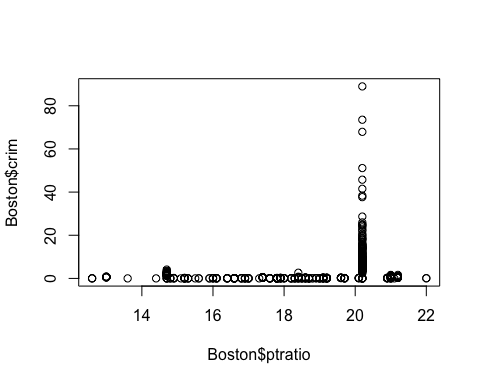
plot(Boston$rad, Boston$crim)



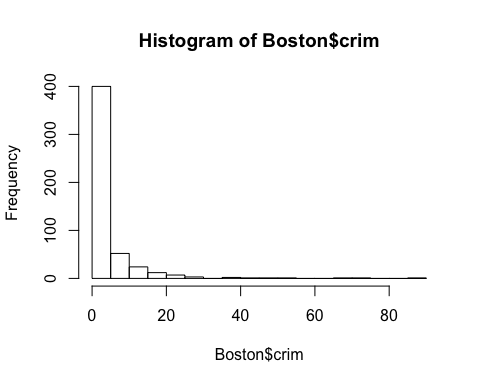
plot(Boston$tax, Boston$crim)



plot(Boston$ptratio, Boston$crim)

 Yes, there are predictors associated with crime rate. \* The older the home, the more crime \* The farther away from Boston employment centers, the less crime \* The more accessible to radial highways, the more crime \* The higher the property tax, the more crime \* The higher the student-teacher ratio, the more crime

hist(Boston$crim, 25)



median(Boston$crim)

## [1] 0.25651

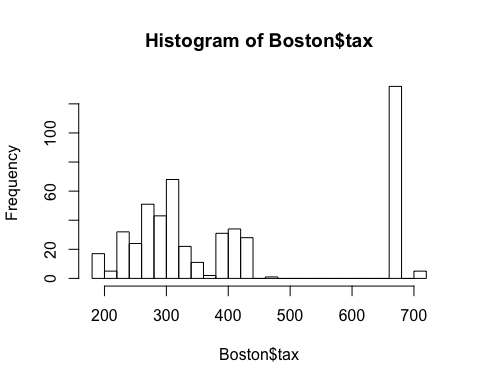
range(Boston$crim)

## [1] 0.00632 88.97620

quantile(Boston$crim, 0.95)

## 95%   
## 15.78915

hist(Boston$tax, 25)



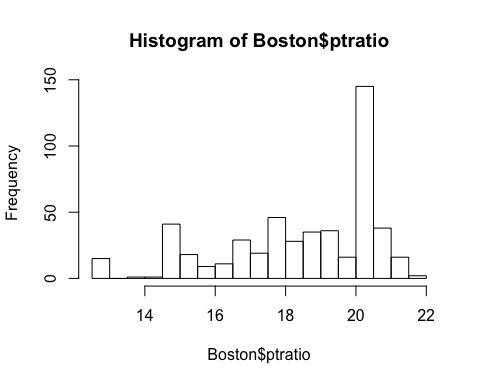
median(Boston$tax)

## [1] 330

range(Boston$tax)

## [1] 187 711

hist(Boston$ptratio, 25)

 Yes, the median crime rate is 0.26, but the crime rate ranges from 0.01 to 88.98. Most cities have low crime rates, and the 5% of suburbs with the highest crime rate have crime rates larger than 15.78.

Tax rate is similar in that the median tax rate is 330, but there is a big jump between suburbs with lower tax rates and suburbs with tax rates higher than 650.

The pupil-teacher ratio distribution among the suburbs is pretty evenly distributed with the exception of a peak of suburbs with a pupil-teacher ratio of ~20.

sum(Boston$chas)

## [1] 35

35 surburbs out of 506 bound the Charles River

median(Boston$ptratio)

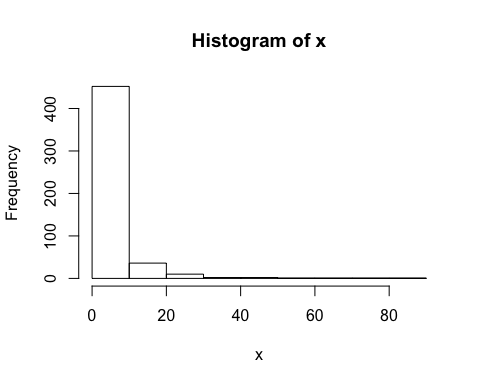
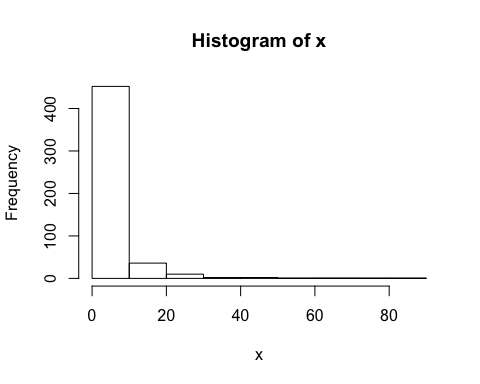
## [1] 19.05

The median pupil-teacher ratio is 19.05.

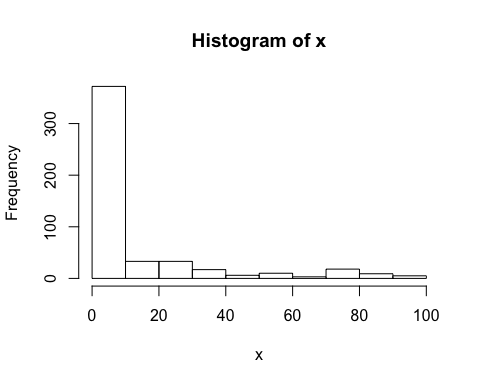
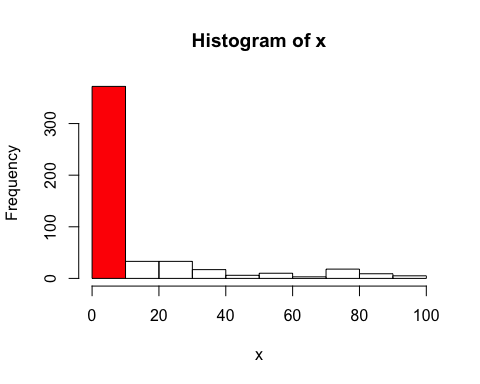
subset(Boston, medv == min(Boston$medv))

## 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

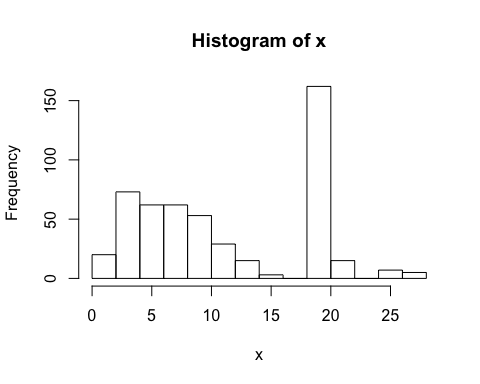
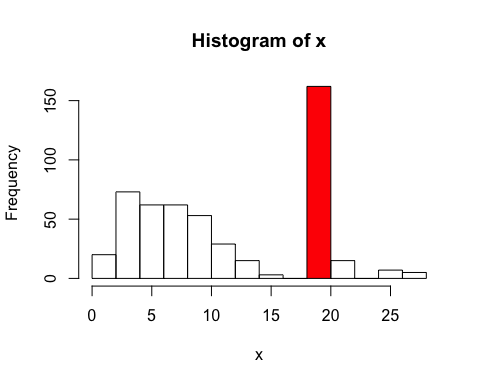
highlight <- function(x, value, col.value, col=NA, ...){  
 hst <- hist(x, ...)  
 idx <- findInterval(value, hst$breaks)  
 cols <- rep(col, length(hst$counts))  
 cols[idx] <- col.value  
 hist(x, col=cols, ...)  
}  
highlight(Boston$crim, Boston$crim[Boston$medv == min(Boston$medv)], "red")

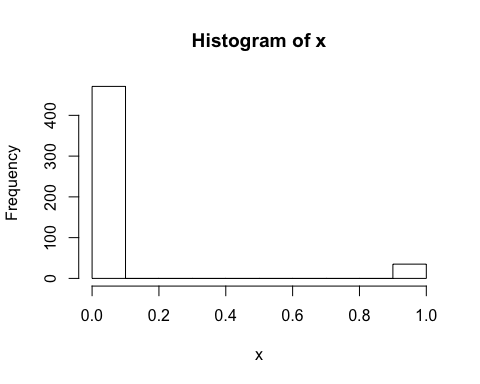
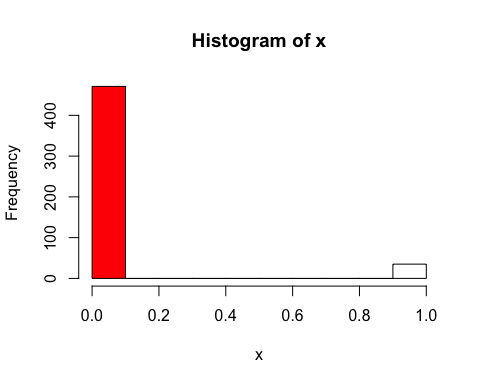
highlight(Boston$zn, Boston$zn[Boston$medv == min(Boston$medv)], "red")

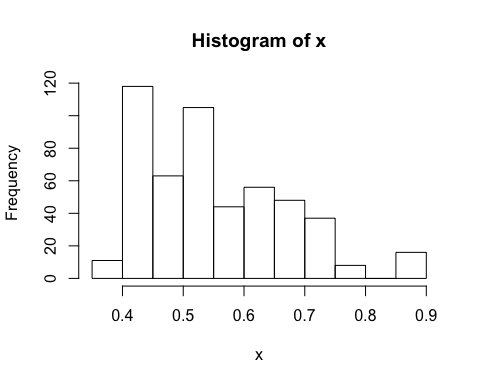
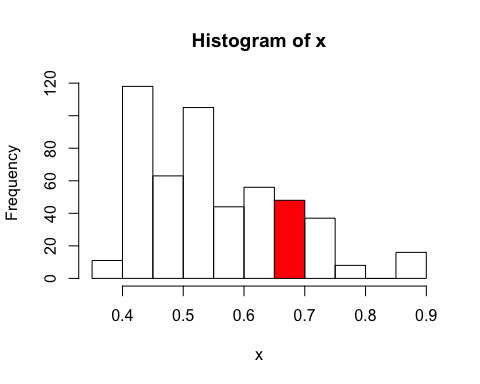
highlight(Boston$indus, Boston$indus[Boston$medv == min(Boston$medv)], "red")

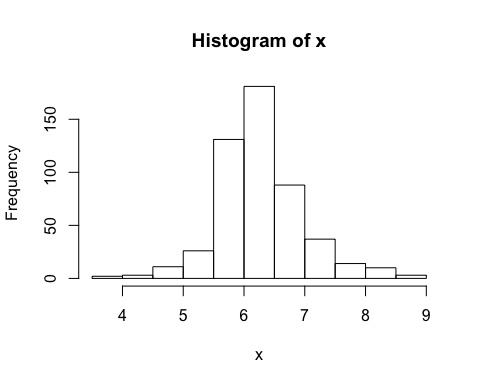
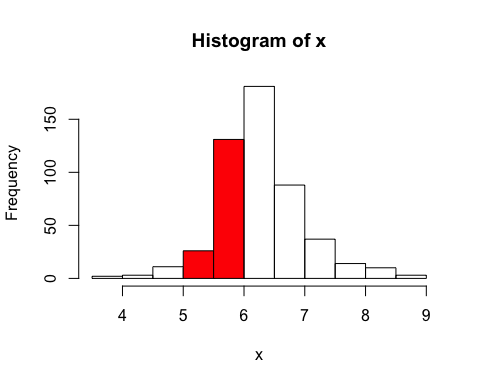
highlight(Boston$chas, Boston$chas[Boston$medv == min(Boston$medv)], "red")

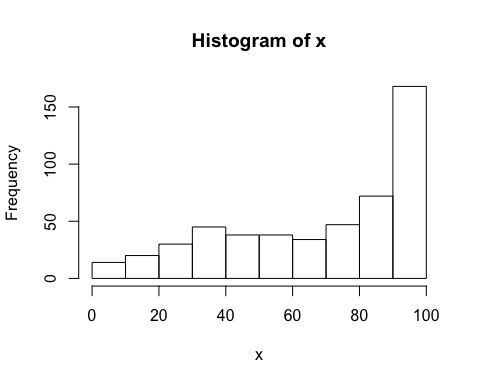
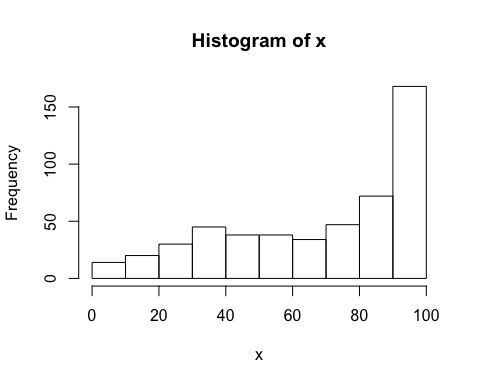
highlight(Boston$nox, Boston$nox[Boston$medv == min(Boston$medv)], "red")

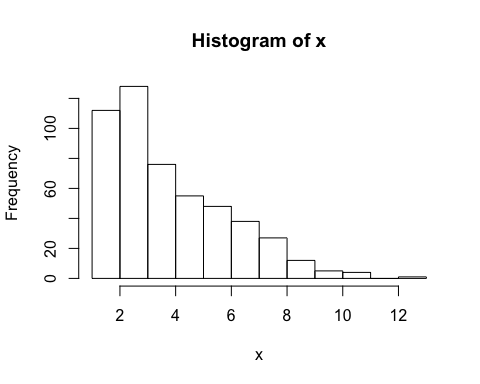
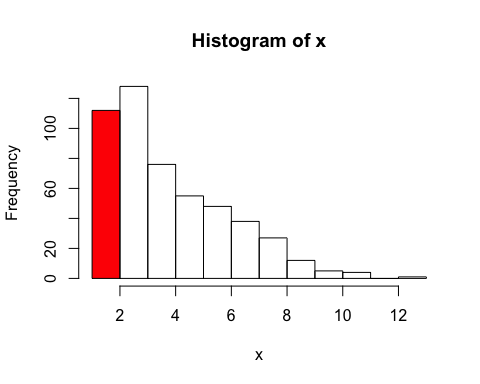
highlight(Boston$rm, Boston$rm[Boston$medv == min(Boston$medv)], "red")

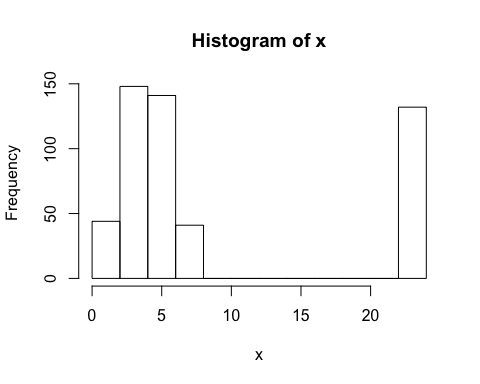
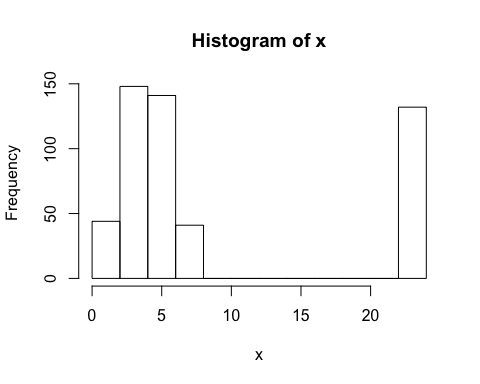
highlight(Boston$age, Boston$age[Boston$medv == min(Boston$medv)], "red")

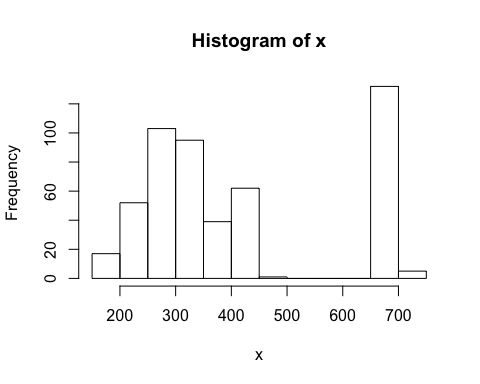
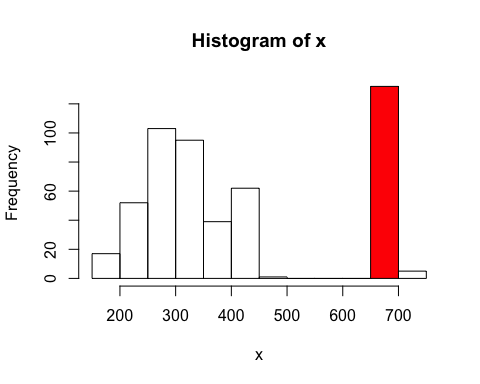
highlight(Boston$dis, Boston$dis[Boston$medv == min(Boston$medv)], "red")

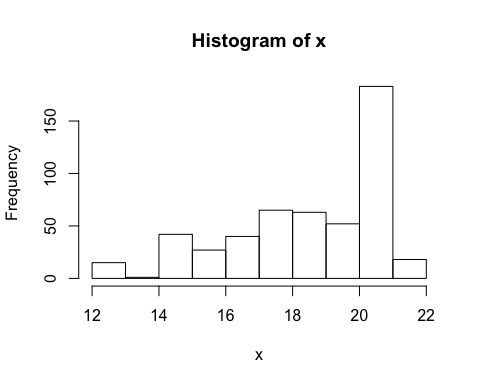
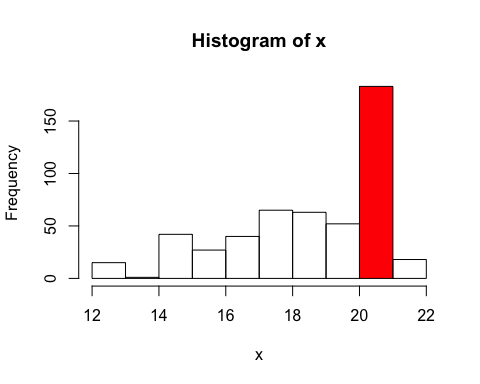
highlight(Boston$rad, Boston$rad[Boston$medv == min(Boston$medv)], "red")

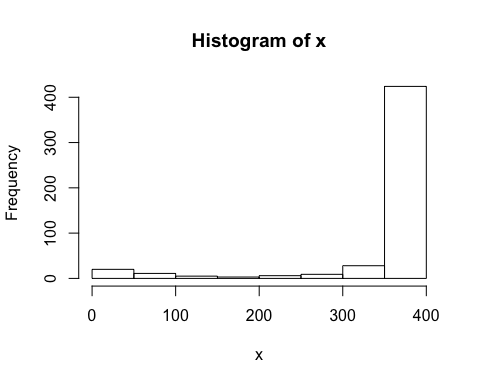
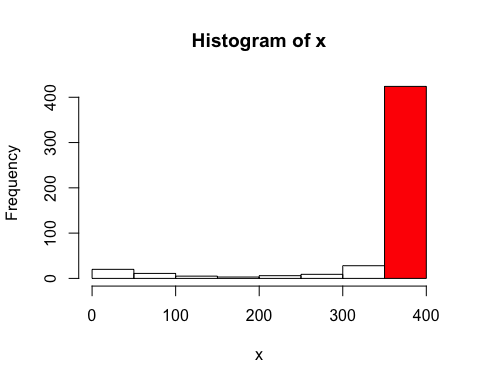
highlight(Boston$tax, Boston$tax[Boston$medv == min(Boston$medv)], "red")

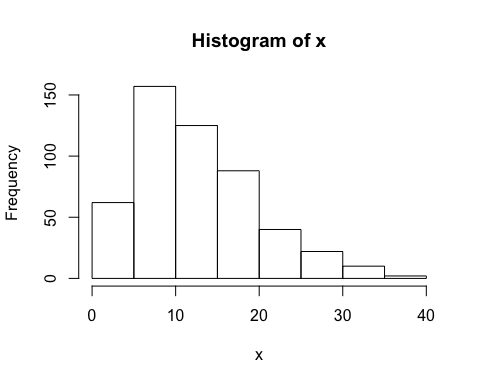
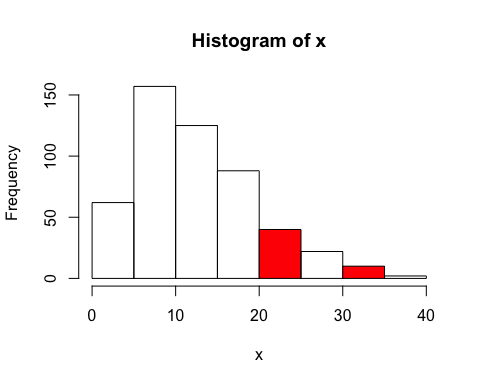
highlight(Boston$ptratio, Boston$ptratio[Boston$medv == min(Boston$medv)], "red")

highlight(Boston$black, Boston$black[Boston$medv == min(Boston$medv)], "red")

highlight(Boston$lstat, Boston$lstat[Boston$medv == min(Boston$medv)], "red")

  High crime, low proportion of large lots, large proportion of non-retail business, not bound by Charles River, larger nitrogen oxide concentration, oldest homes, closest to employment centers, most accessible to radial highways, higher property taxes, higher student-teacher ratios, higher proportion of blacks, and higher proportion of lower status.

nrow(subset(Boston, rm > 7))

## [1] 64

nrow(subset(Boston, rm > 8))

## [1] 13

summary(subset(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

There are 64 suburbs with more than 7 rooms per dwelling, and 13 suburbs with more than 8 rooms per dwelling.

The suburbs with more than 8 rooms per dwelling have less crime, higher median value, and lower lstat.

## Chapter