Predictive\_Modeling\_Week\_1\_Excercise\_1

Amol Gote

7/11/2020

# Read File

college <- read.csv("data/College.csv")  
#college

# Read Fix College name column

#rownames(college) = college[,1]  
college\_orig <- college  
college =college [,-1]  
#fix(college)  
#college

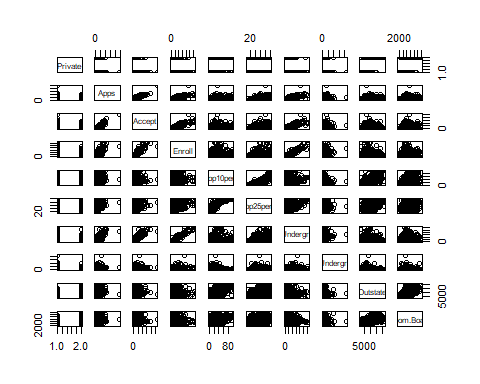
# Summary

summary(college)

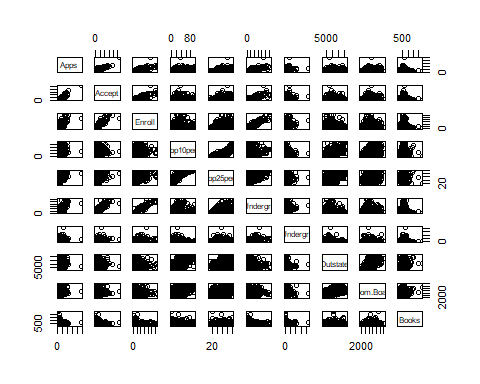
## Private Apps Accept Enroll Top10perc   
## No :212 Min. : 81 Min. : 72 Min. : 35 Min. : 1.00   
## Yes:565 1st Qu.: 776 1st Qu.: 604 1st Qu.: 242 1st Qu.:15.00   
## Median : 1558 Median : 1110 Median : 434 Median :23.00   
## Mean : 3002 Mean : 2019 Mean : 780 Mean :27.56   
## 3rd Qu.: 3624 3rd 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   
## Min. : 10.00   
## 1st Qu.: 53.00   
## Median : 65.00   
## Mean : 65.46   
## 3rd Qu.: 78.00   
## Max. :118.00

# pairs() function

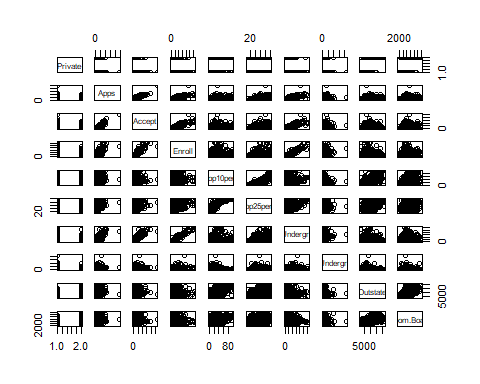
pairs(college[,1:10])



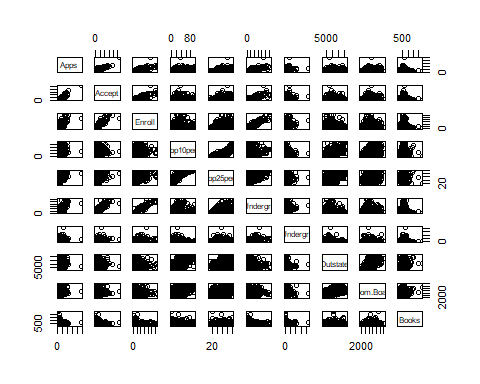
pairs(∼ Apps + Accept + Enroll + Top10perc + Top25perc + F.Undergrad +   
 P.Undergrad + Outstate + Room.Board + Books, college)



pairs(college[,1:10])

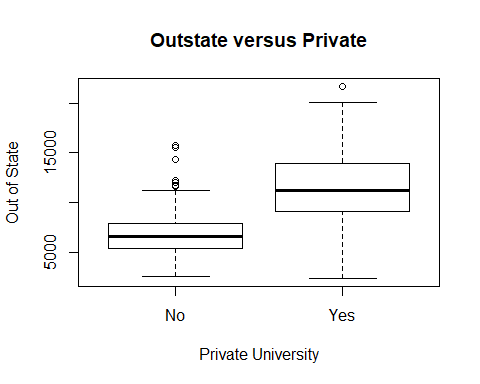


pairs(∼ Apps + Accept + Enroll + Top10perc + Top25perc + F.Undergrad +   
 P.Undergrad + Outstate + Room.Board + Books, college)



# plot function

plot(college$Private, college$Outstate,   
 xlab = "Private University", ylab ="Out of State", main = "Outstate versus Private")



# Elite

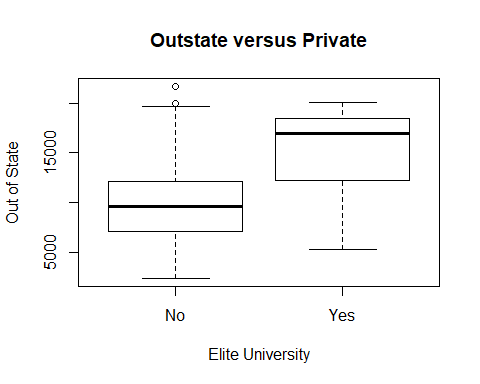
Elite = rep("No",nrow(college))  
Elite[college$Top10perc > 50]= "Yes"  
Elite = as.factor (Elite)  
college = data.frame(college, Elite)  
summary(college)

## 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

summary(college$Elite)

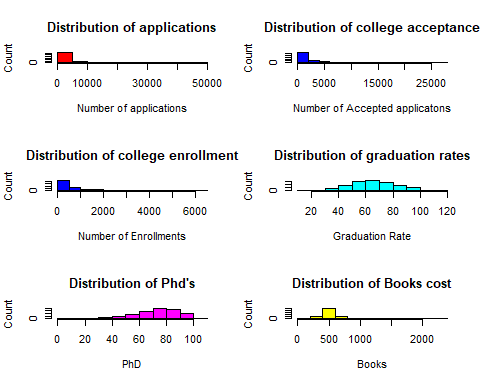
## No Yes   
## 699 78

plot(college$Elite, college$Outstate,   
 xlab = "Elite University", ylab ="Out of State", main = "Outstate versus Private")



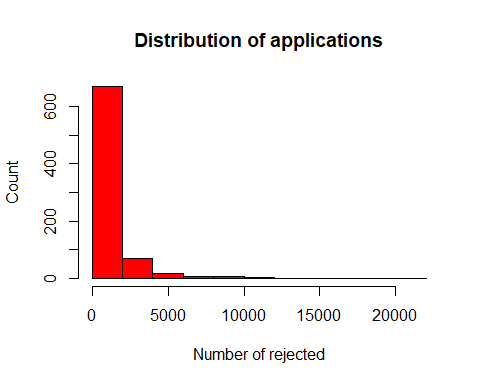
# Histograms

par(mfrow=c(3,2))  
hist(college$Apps, col = 2, xlab = "Number of applications", ylab = "Count", main = "Distribution of applications")  
hist(college$Accept, col = 4, xlab = "Number of Accepted applicatons", ylab = "Count", main = "Distribution of college acceptance")  
hist(college$Enroll, col = 4, xlab = "Number of Enrollments", ylab = "Count", main = "Distribution of college enrollment")  
hist(college$Grad.Rate, col = 5, xlab = "Graduation Rate", ylab = "Count", main = "Distribution of graduation rates")  
hist(college$PhD, col = 6, xlab = "PhD", ylab = "Count", main = "Distribution of Phd's")  
hist(college$Books, col = 7, xlab = "Books", ylab = "Count", main = "Distribution of Books cost")



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

college$rejected <- college$Apps - college$Accept  
hist(college$rejected, col = 2, xlab = "Number of rejected", ylab = "Count", main = "Distribution of applications")



summary(college$rejected)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0 131.0 326.0 982.8 1066.0 21764.0

colleges\_rejected\_5000 <- college %>%  
 filter(college$rejected > 5000)  
nrow(colleges\_rejected\_5000)

## [1] 24

colleges\_rejected\_20000 <- college %>%  
 filter(college$rejected > 20000)  
colleges\_rejected\_20000

## Private Apps Accept Enroll Top10perc Top25perc F.Undergrad P.Undergrad  
## 1 No 48094 26330 4520 36 79 21401 3712  
## Outstate Room.Board Books Personal PhD Terminal S.F.Ratio perc.alumni Expend  
## 1 7410 4748 690 2009 90 95 19.5 19 10474  
## Grad.Rate Elite rejected  
## 1 77 No 21764

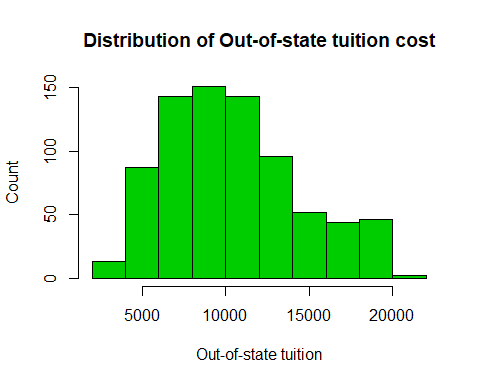
colleges\_with\_high\_reject <- college\_orig %>%  
 filter(Apps == 48094)  
colleges\_with\_high\_reject

## X Private Apps Accept Enroll Top10perc Top25perc  
## 1 Rutgers at New Brunswick No 48094 26330 4520 36 79  
## F.Undergrad P.Undergrad Outstate Room.Board Books Personal PhD Terminal  
## 1 21401 3712 7410 4748 690 2009 90 95  
## S.F.Ratio perc.alumni Expend Grad.Rate  
## 1 19.5 19 10474 77

summary(college$Outstate)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2340 7320 9990 10441 12925 21700

hist(college$Outstate, col = 3, xlab = "Out-of-state tuition", ylab = "Count", main = "Distribution of Out-of-state tuition cost")



1. Difference between number of applications and accepted indicates how many applications have been rejected.
2. Average rejected applications is 928 and median is 326.
3. There are 24 colleges with number of applications rejected > 5000.
4. “Rutgers at New Brunswick” college has the highest rejected applications as well as highest number of applications which is 48090.
5. College out state tuition is evenly distributed with mean and median around 10000

# Problem 10

# Boston dataset info

#?Boston  
#Boston  
nrow(Boston)

## [1] 506

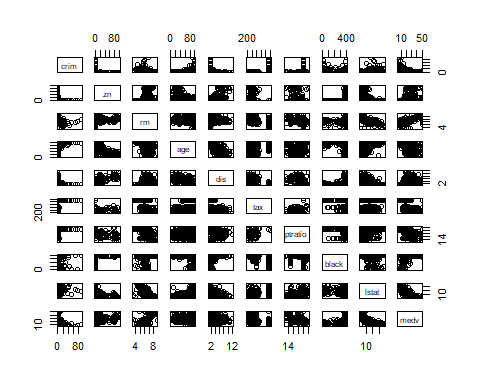
ncol(Boston)

## [1] 14

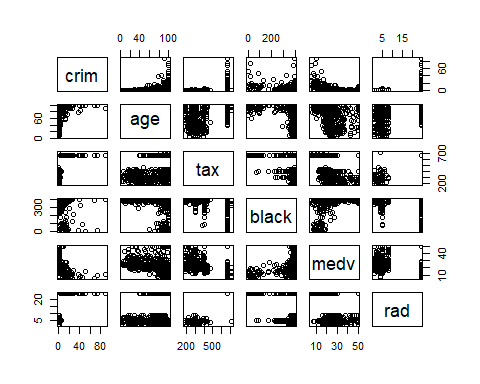
1. Dataset has 506 rows
2. Dataset has 14 columns.
3. Each column represents variable for which an observation was made in 506 neighborhoods of Boston.These variables include
   1. crim - per capita crime rate by town.
   2. zn - proportion of residential land zoned for lots over 25,000 sq.ft.
   3. indus - proportion of non-retail business acres per town.
   4. chas - Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).
   5. nox - nitrogen oxides concentration (parts per 10 million).
   6. rm - average number of rooms per dwelling.
   7. age - proportion of owner-occupied units built prior to 1940.
   8. dis - weighted mean of distances to five Boston employment centres.
   9. rad - index of accessibility to radial highways.
   10. tax - full-value property-tax rate per $10,000.
   11. ptratio - pupil-teacher ratio by town.
   12. black - 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town.
   13. lstat - lower status of the population (percent).
   14. medv - median value of owner-occupied homes in $1000s.

# Pairwise scatterplots of the predictors

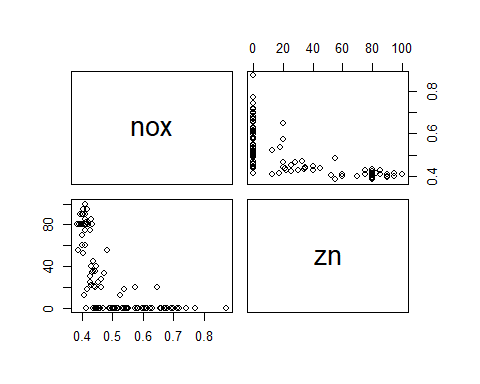
pairs(∼ crim + zn +   
 + rm  
 + age   
 + dis   
 + tax   
 + ptratio  
 + black  
 + lstat  
 + medv, Boston  
 )



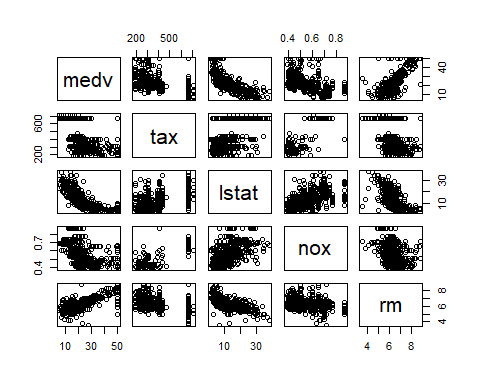
pairs(crim ~ age + tax + black + medv + rad, Boston)



pairs(nox ~ zn, Boston)



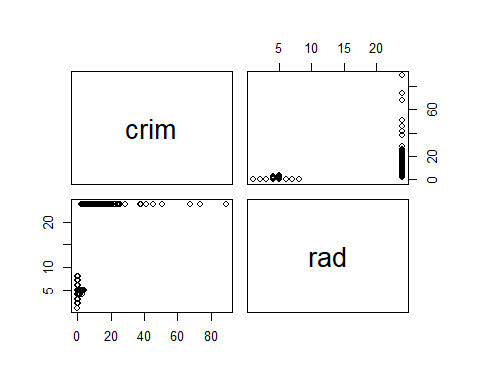
pairs(medv ~ tax + lstat + nox + rm, Boston)



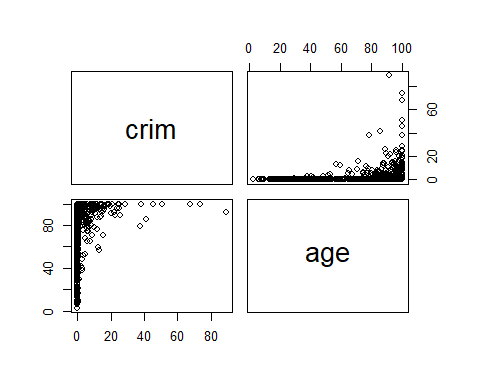
1. Where Proportion of residential land zoned is less the nox levels have increased.
2. As lsat (lower status of the population) increases then medv(median value of owner-occupied homes) decreases.
3. As nox (lower status of the population) increases then medv(median value of owner-occupied homes) decreases.
4. As rm (average number of rooms per dwelling) increases then medv(median value of owner-occupied homes) increases.

# predictors associated with per capita crime rate?

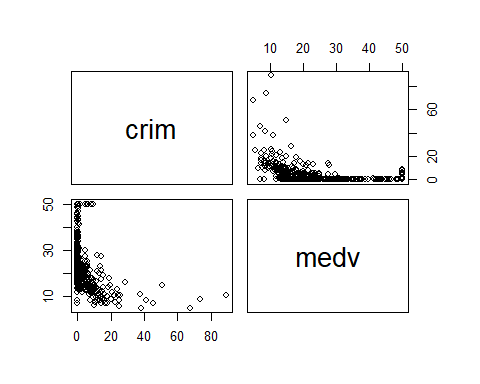
pairs(crim ~ rad, Boston)



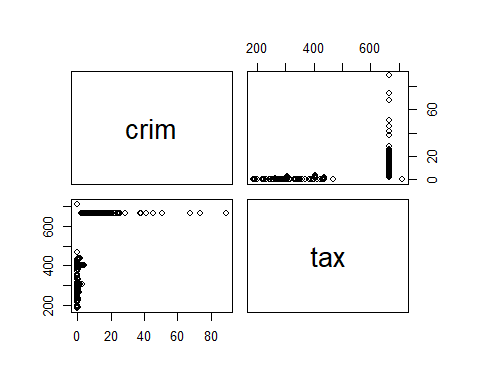
pairs(crim ~ age, Boston)



pairs(crim ~ medv, Boston)



pairs(crim ~ tax, Boston)



1. A higher values of rad (index of accessibility to radial highways) contain the highest level of crime rate.
2. As the age (proportion of owner-occupied units built prior to 1940) has increased the crime rate has increased.
3. crime rate is higher in the lower median value of owner-occupied homes.
4. As tax rate increase, crime rate increases.

# suburbs of Boston appear to have particularly high crime rates? Tax rates? Pupil-teacher ratios

summary(Boston$crim)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00632 0.08204 0.25651 3.61352 3.67708 88.97620

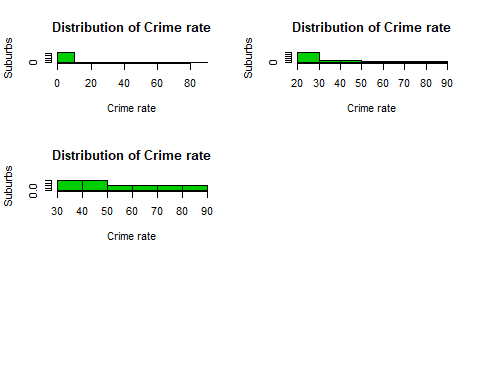
crimeGreateThan20 <- Boston$crim[Boston$crim>20]  
crimeGreateThan30 <- Boston$crim[Boston$crim>30]  
length(crimeGreateThan20)

## [1] 18

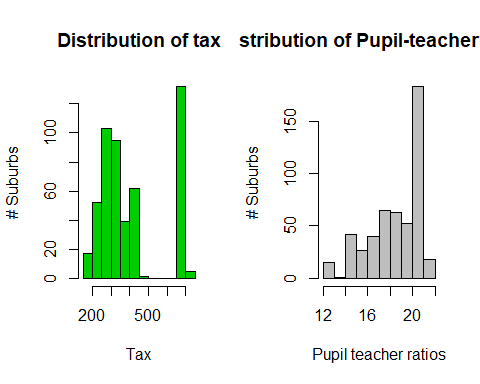
length(crimeGreateThan30)

## [1] 8

par(mfrow=c(3,2))  
hist(Boston$crim, col = 3, xlab = "Crime rate", ylab = "Suburbs", main = "Distribution of Crime rate")  
highCrime <- Boston$crim[Boston$crim>20]  
hist(highCrime, col = 3, xlab = "Crime rate", ylab = "Suburbs", main = "Distribution of Crime rate")  
highCrime <- Boston$crim[Boston$crim>30]  
hist(highCrime, col = 3, xlab = "Crime rate", ylab = "Suburbs", main = "Distribution of Crime rate")



par(mfrow=c(1,2))  
hist(Boston$tax, col = 3, xlab = "Tax", ylab = "# Suburbs", main = "Distribution of tax")  
hist(Boston$ptratio, col = 8, xlab = "Pupil teacher ratios", ylab = "# Suburbs", main = "Distribution of Pupil-teacher ratios")



summary(Boston$tax)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 187.0 279.0 330.0 408.2 666.0 711.0

summary(Boston$ptratio)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 12.60 17.40 19.05 18.46 20.20 22.00

length(Boston$crim[Boston$tax>600])

## [1] 137

length(Boston$crim[Boston$ptratio>20])

## [1] 201

1. Most cities have less crime rate with average crime rate = 3.6.
2. There are 18 suburbs with crime rate greater than 20 and 8 suburbs with crime rate greater than 30.
3. Some suburbs have the highest tax with value greater than 600, there are 137 suburbs like these. These suburbs are beyond the 3rd quartile values.
4. Some suburbs have the highest Pupil-teacher ratios with value greater than 20, there are 201 suburbs like these. These suburbs are beyond the 3rd quartile values.

# Suburbs in this data set bound the Charles river

suburbs\_bound\_to\_river <- Boston %>%  
 filter(chas == 1)  
nrow(suburbs\_bound\_to\_river)

## [1] 35

1. Number of Suburbs bound to the Charles river is 35.

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

median(Boston$ptratio)

## [1] 19.05

1. Median pupil-teacher ratio is 19.05.

# Which suburb of Boston has lowest median value of owneroccupied homes?

lowestMedianValue <- Boston %>%  
 filter(medv == min(Boston$medv))  
lowestMedianValue

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

1. There are 2 suburbs with with lowest median value of 5.

compare\_summary\_dataset <- function(dataset\_orig,dataset\_target) {  
 sum1 <- apply(dataset\_orig,2,summary) %>% data.frame()   
 sum2 <- apply(dataset\_target,2,summary) %>% data.frame()   
  
 names(sum1) <- paste0(names(sum1),"1")  
 names(sum2) <- paste0(names(sum2),"2")  
  
 final <- cbind(sum1,sum2)  
  
 final1 <- t(final)   
  
 final2 <- final1[order(row.names(final1)), ]  
  
 final\_1 <- t(final2) %>% data.frame()  
 final\_1  
}

# 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.

t(compare\_summary\_dataset(Boston,lowestMedianValue))

## Min. 1st Qu. Median Mean 3rd Qu. Max.  
## age1 2.90000 45.025000 77.50000 68.57490119 94.075000 100.0000  
## age2 100.00000 100.000000 100.00000 100.00000000 100.000000 100.0000  
## black1 0.32000 375.377500 391.44000 356.67403162 396.225000 396.9000  
## black2 384.97000 387.952500 390.93500 390.93500000 393.917500 396.9000  
## chas1 0.00000 0.000000 0.00000 0.06916996 0.000000 1.0000  
## chas2 0.00000 0.000000 0.00000 0.00000000 0.000000 0.0000  
## crim1 0.00632 0.082045 0.25651 3.61352356 3.677083 88.9762  
## crim2 38.35180 45.744050 53.13630 53.13630000 60.528550 67.9208  
## dis1 1.12960 2.100175 3.20745 3.79504269 5.188425 12.1265  
## dis2 1.42540 1.441450 1.45750 1.45750000 1.473550 1.4896  
## indus1 0.46000 5.190000 9.69000 11.13677866 18.100000 27.7400  
## indus2 18.10000 18.100000 18.10000 18.10000000 18.100000 18.1000  
## lstat1 1.73000 6.950000 11.36000 12.65306324 16.955000 37.9700  
## lstat2 22.98000 24.882500 26.78500 26.78500000 28.687500 30.5900  
## medv1 5.00000 17.025000 21.20000 22.53280632 25.000000 50.0000  
## medv2 5.00000 5.000000 5.00000 5.00000000 5.000000 5.0000  
## nox1 0.38500 0.449000 0.53800 0.55469506 0.624000 0.8710  
## nox2 0.69300 0.693000 0.69300 0.69300000 0.693000 0.6930  
## ptratio1 12.60000 17.400000 19.05000 18.45553360 20.200000 22.0000  
## ptratio2 20.20000 20.200000 20.20000 20.20000000 20.200000 20.2000  
## rad1 1.00000 4.000000 5.00000 9.54940711 24.000000 24.0000  
## rad2 24.00000 24.000000 24.00000 24.00000000 24.000000 24.0000  
## rm1 3.56100 5.885500 6.20850 6.28463439 6.623500 8.7800  
## rm2 5.45300 5.510500 5.56800 5.56800000 5.625500 5.6830  
## tax1 187.00000 279.000000 330.00000 408.23715415 666.000000 711.0000  
## tax2 666.00000 666.000000 666.00000 666.00000000 666.000000 666.0000  
## zn1 0.00000 0.000000 0.00000 11.36363636 12.500000 100.0000  
## zn2 0.00000 0.000000 0.00000 0.00000000 0.000000 0.0000

1. Age of lowest median value of owner-occupied homes is higher compared to overall dataset.
2. Proportion of blacks is same as that of median of overall dataset, mean of these 2 suburbs is on higher side.
3. Crime rate is on the higher side; 2 suburbs have crime rate of 38 and 67 with average of 53 which is considerably higher than the overall dataset.
4. Lower status of the population (percent) is greater than mean and median of overall dataset.
5. nox (nitrogen oxides concentration (parts per 10 million)) is on higher side.
6. ptratio (pupil-teacher ratio) falls in that tower of > 20 in the above histogram.
7. Tax is on the higher side compared to overall dataset and it falls in the tower > 600 histogram.

# Suburbs average more than seven rooms per dwelling

dwelling\_7 <- Boston %>%  
 filter(rm > 7)  
nrow(dwelling\_7)

## [1] 64

# Suburbs average more than eight rooms per dwelling

dwelling\_8 <- Boston %>%  
 filter(rm > 8)  
nrow(dwelling\_8)

## [1] 13

summary(dwelling\_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

1. Number of suburbs with more than seven rooms per dwelling is 64.
2. Number of suburbs with more than eight rooms per dwelling is 13.

# Suburbs that average more than eight rooms per dwelling compared to overall dataset.

t(compare\_summary\_dataset(Boston,dwelling\_8))

## Min. 1st Qu. Median Mean 3rd Qu. Max.  
## age1 2.90000 45.025000 77.50000 68.57490119 94.075000 100.00000  
## age2 8.40000 70.400000 78.30000 71.53846154 86.500000 93.90000  
## black1 0.32000 375.377500 391.44000 356.67403162 396.225000 396.90000  
## black2 354.55000 384.540000 386.86000 385.21076923 389.700000 396.90000  
## chas1 0.00000 0.000000 0.00000 0.06916996 0.000000 1.00000  
## chas2 0.00000 0.000000 0.00000 0.15384615 0.000000 1.00000  
## crim1 0.00632 0.082045 0.25651 3.61352356 3.677083 88.97620  
## crim2 0.02009 0.331470 0.52014 0.71879538 0.578340 3.47428  
## dis1 1.12960 2.100175 3.20745 3.79504269 5.188425 12.12650  
## dis2 1.80100 2.288500 2.89440 3.43019231 3.651900 8.90670  
## indus1 0.46000 5.190000 9.69000 11.13677866 18.100000 27.74000  
## indus2 2.68000 3.970000 6.20000 7.07846154 6.200000 19.58000  
## lstat1 1.73000 6.950000 11.36000 12.65306324 16.955000 37.97000  
## lstat2 2.47000 3.320000 4.14000 4.31000000 5.120000 7.44000  
## medv1 5.00000 17.025000 21.20000 22.53280632 25.000000 50.00000  
## medv2 21.90000 41.700000 48.30000 44.20000000 50.000000 50.00000  
## nox1 0.38500 0.449000 0.53800 0.55469506 0.624000 0.87100  
## nox2 0.41610 0.504000 0.50700 0.53923846 0.605000 0.71800  
## ptratio1 12.60000 17.400000 19.05000 18.45553360 20.200000 22.00000  
## ptratio2 13.00000 14.700000 17.40000 16.36153846 17.400000 20.20000  
## rad1 1.00000 4.000000 5.00000 9.54940711 24.000000 24.00000  
## rad2 2.00000 5.000000 7.00000 7.46153846 8.000000 24.00000  
## rm1 3.56100 5.885500 6.20850 6.28463439 6.623500 8.78000  
## rm2 8.03400 8.247000 8.29700 8.34853846 8.398000 8.78000  
## tax1 187.00000 279.000000 330.00000 408.23715415 666.000000 711.00000  
## tax2 224.00000 264.000000 307.00000 325.07692308 307.000000 666.00000  
## zn1 0.00000 0.000000 0.00000 11.36363636 12.500000 100.00000  
## zn2 0.00000 0.000000 0.00000 13.61538462 20.000000 95.00000

1. Crime rate for suburbs with more than eight rooms per dwelling is on lower side compared to overall dataset.
2. lstat (lower status of the population (percent)) is lower than the overall population.
3. medv (median value of owner-occupied homes) is on higher side than overall dataset.
4. tax (full-value property-tax rate per $10,000) mean is lower than that of the overall dataset.