Step 1

The library Boston provides data for housing values in Suburbs of Boston. The library contains 506 rows and 14 columns. The columns of this data frame include per capita crime(crim) rate by town, pupil-teach(ptratio) ratio by town, and others.

```
library(MASS)
data(Boston)
str(Boston)
```

```
## 'data.frame':    506 obs. of  14 variables:
##  $ crim   : num  0.00632 0.02731 0.02729 0.03237 0.06905 ...
##  $ zn     : num  18 0 0 0 0 0 12.5 12.5 12.5 12.5 ...
##  $ indus  : num  2.31 7.07 7.07 2.18 2.18 2.18 7.87 7.87 7.87 7.87 ...
##  $ chas   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ nox    : num  0.538 0.469 0.469 0.458 0.458 0.458 0.524 0.524 0.524 0.524 ...
##  $ rm     : num  6.58 6.42 7.18 7 7.15 ...
##  $ age    : num  65.2 78.9 61.1 45.8 54.2 58.7 66.6 96.1 100 85.9 ...
##  $ dis    : num  4.09 4.97 4.97 6.06 6.06 ...
##  $ rad    : int  1 2 2 3 3 3 5 5 5 5 ...
##  $ tax    : num  296 242 242 222 222 222 311 311 311 311 ...
##  $ ptratio: num  15.3 17.8 17.8 18.7 18.7 18.7 15.2 15.2 15.2 15.2 ...
##  $ black  : num  397 397 393 395 397 ...
##  $ lstat  : num  4.98 9.14 4.03 2.94 5.33 ...
##  $ medv   : num  24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...
```

Step 2

```
head(Boston, n=5) #First five rows
```

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

```
tail(Boston, n=2) #Last two rows
```

```
##        crim zn indus chas   nox    rm  age    dis rad tax ptratio  black lstat
## 505 0.10959  0 11.93    0 0.573 6.794 89.3 2.3889   1 273      21 393.45  6.48
## 506 0.04741  0 11.93    0 0.573 6.030 80.8 2.5050   1 273      21 396.90  7.88
##     medv
## 505 22.0
## 506 11.9
```

```
Boston[5,] #5th row
```

```
##      crim zn indus chas   nox    rm  age    dis rad tax ptratio black lstat
## 5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7 396.9  5.33
##   medv
## 5 36.2
```

```
head(Boston[1], n=5) #First 5 rows of column 1
```

```
##      crim
## 1 0.00632
## 2 0.02731
## 3 0.02729
## 4 0.03237
## 5 0.06905
```

```
names(Boston) #Displays variable names
```

```
##  [1] "crim"    "zn"      "indus"   "chas"    "nox"     "rm"      "age"
##  [8] "dis"     "rad"     "tax"     "ptratio" "black"   "lstat"   "medv"
```

Step 3

```
mean(Boston$crim) #Mean of crime rate statistics
```

```
## [1] 3.613524
```

```
median(Boston$crim) #Median of crime rate statistics
```

```
## [1] 0.25651
```

```
range(Boston$crim) #Range of crime rate statistics where first value is min and second is max
```

```
## [1]  0.00632 88.97620
```

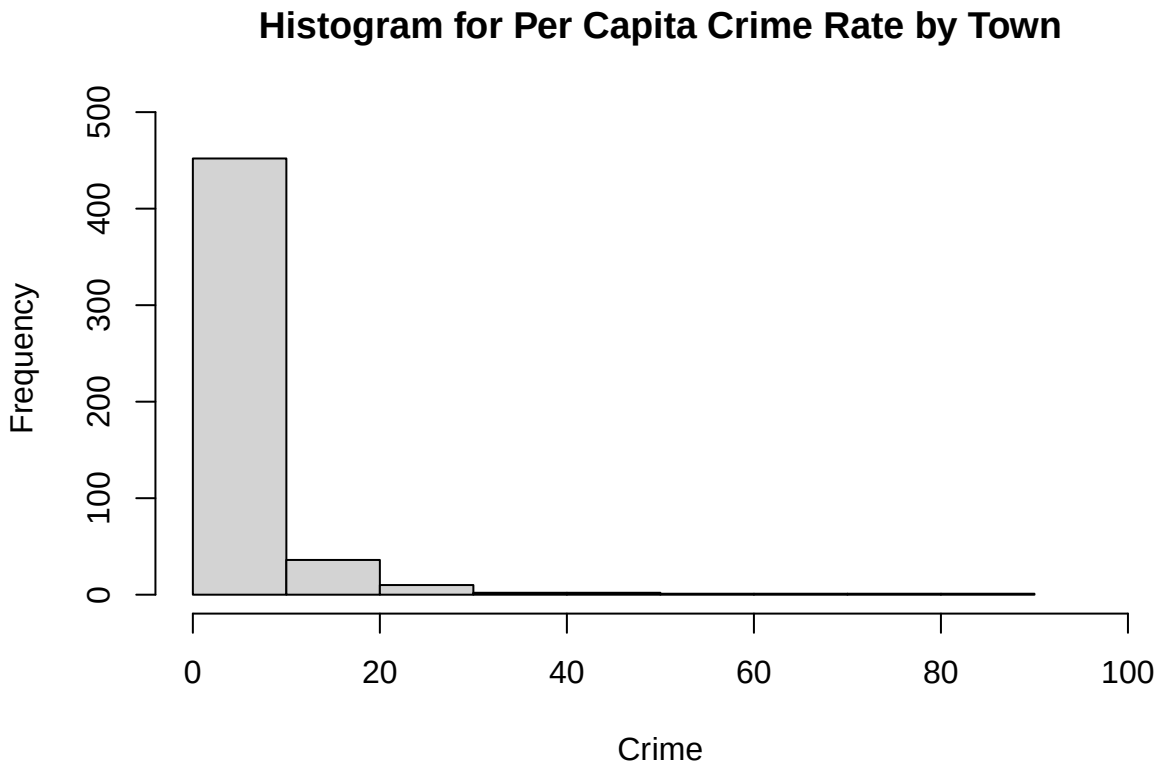
```
max(Boston[,1])-min(Boston[,1]) #Mathematical range of crime column
```

```
## [1] 88.96988
```

Step 4

Below is a histogram graph for crime part of the Boston library data frame. This graph tells us that majority of rows fall into the range of 20 and below. In other words, majority of neighborhoods have 20 or less crimes per thousand people.

```
hist(Boston$crim,
     main="Histogram for Per Capita Crime Rate by Town",
     xlab ="Crime",
     ylim = c(0, 490),
     xlim = c(0, 100))
```



Step 5

In a relationship between crime rate and median value of homes, the negative correlation value shows that there is a moderate inverse relationship between the two. Which means that in places where the crime rate is high, the median value of homes is moderately low. Therefore, using crime rate statistics to predict the median value of a house is likely to improve its accuracy.

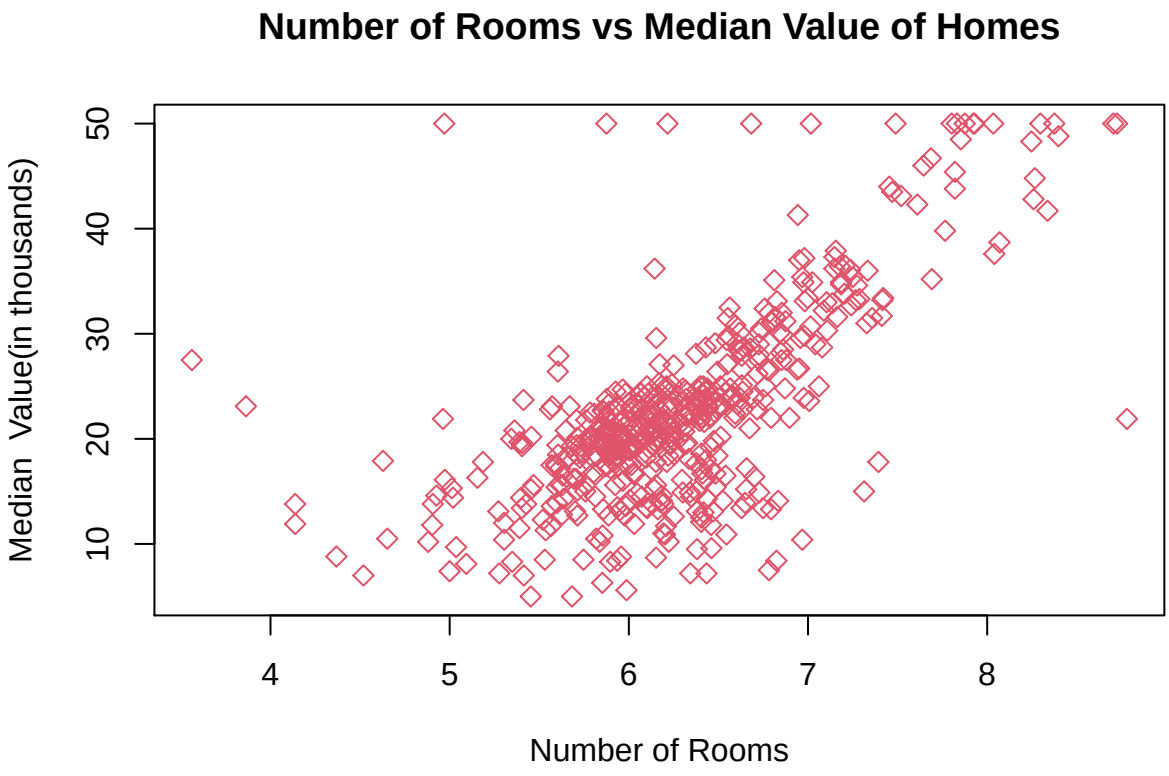
```
cor(Boston$crim, Boston$medv) #Correlation value between crime rate and median value of homes.
```

```
## [1] -0.3883046
```

Step 6

From the graph below we can derive to a conclusion that in most cases the value of a house is higher if it has more rooms. This is confirmed by the correlation coefficient. It is moderately likely that number of rooms determines the house's value.

```r
x <- Boston$rm
y <- Boston$medv
plot(x, y,
     main = "Number of Rooms vs Median Value of Homes",
     xlab = "Number of Rooms",
     ylab = "Median  Value(in thousands)",
     pch = 5,
     col = 258)
```




```r
cor(x, y) #Correlation coefficient
```

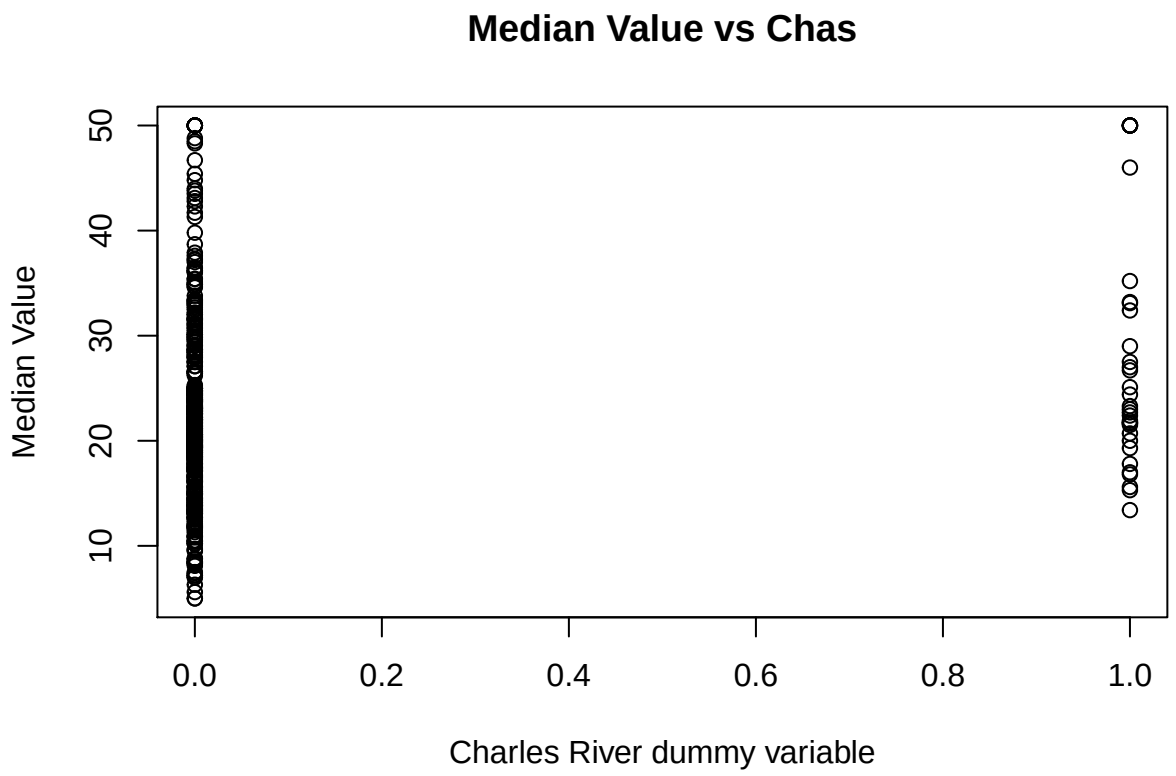
```
## [1] 0.6953599
```

Step 7

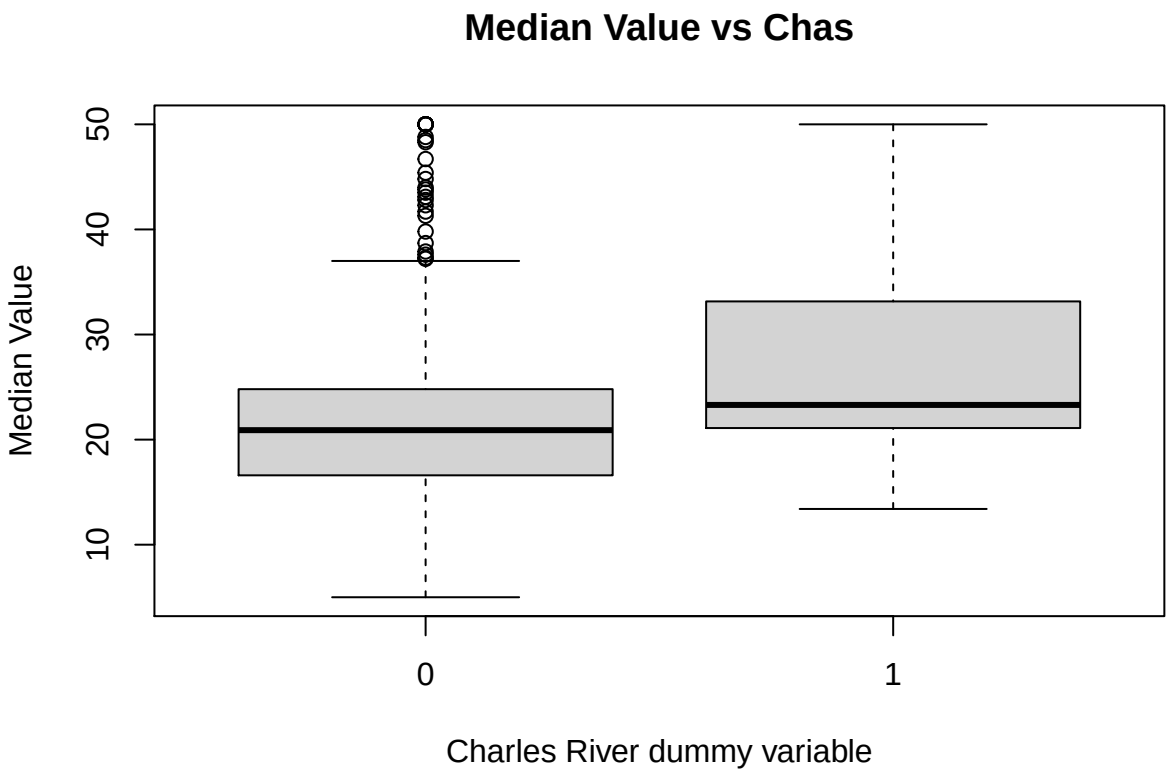
While the first graph depicts the raw data relationship between median value of houses and Charles River's dummy value, the second graph shows the concentration of data. By looking at the second graph, we are able to make the conclusion that if an area in which the neighborhood is located in tracts a river, the median value of houses is slightly higher compared to those that are not.The second graph also shows the outliers(values that fall outside of other values of the data set).

```r
if (is.factor(Boston$chas)==FALSE) {
  print('Variable chas is not a factor')
  plot(Boston$chas, Boston$medv,
       main = "Median Value vs Chas",
       xlab = "Charles River dummy variable",
       ylab = "Median Value",
       )
}
```

```
## [1] "Variable chas is not a factor"
```




```r
z <- as.factor(Boston$chas)
plot(z, Boston$medv,
       main = "Median Value vs Chas",
       xlab = "Charles River dummy variable",
       ylab = "Median Value",
       )
```



Step 8

A higher index of accessibility to radial highways means better accessibility, and vice versa. Rad is an integer variable with minimum value of 1, median value of 5, mean value of 9.549, and maximum value of 24. By looking at the unique values of the library we observe that there are instances of index values of 1, 2, 3, 4, 5, 6, 7, 8, and 24.

```
summary(Boston$rad) #Summary of the rad column
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   4.000   5.000   9.549  24.000  24.000
```

```
unique(Boston$rad) #Unique values in rad column
```

```
## [1]  1  2  3  5  4  8  6  7 24
```

```
rad_sum <- sum(Boston$rad == 24) #Number of neighborhoods with rad index of 24
cat(rad_sum, "Instances")
```

```
## 132 Instances
```

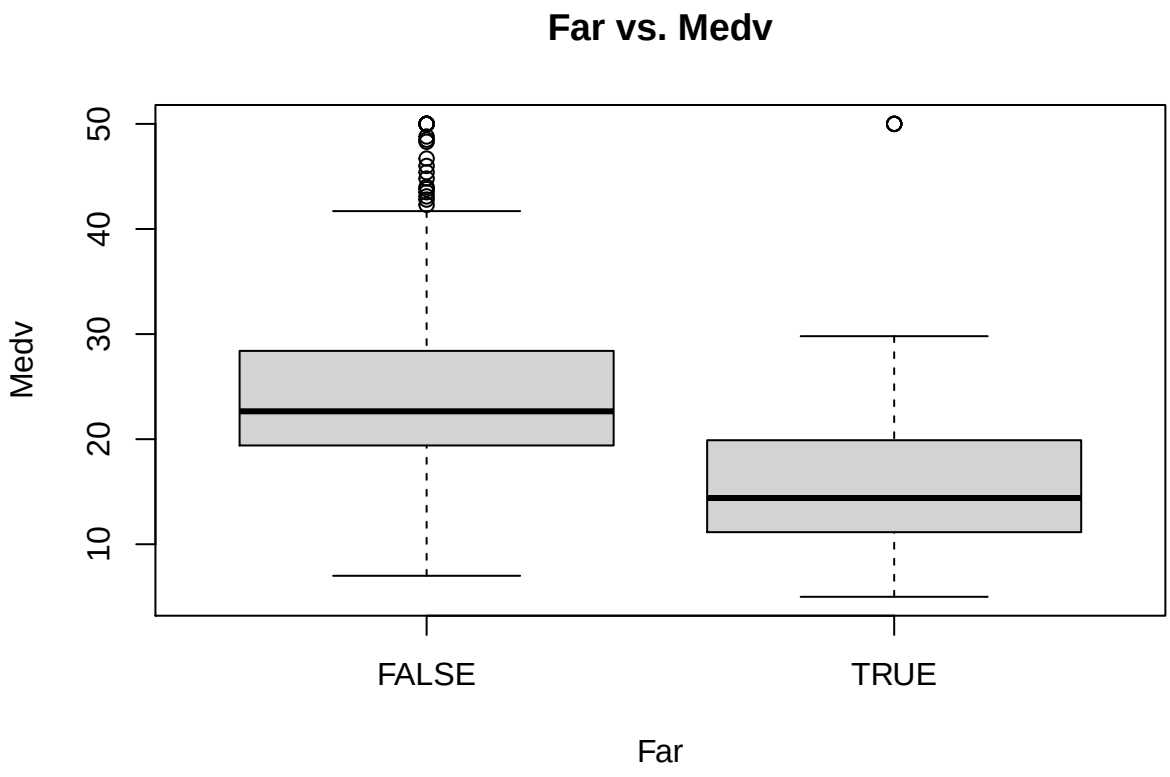
```
n_rows <- nrow(Boston)
cat((rad_sum/n_rows)*100, "percent of all neighborhoods.")
```

```
## 26.08696 percent of all neighborhoods.
```

Step 9

The graph below shows the relationship between houses with better and worse access to radial highways, and median value of houses. From this graph we can tell that if a neighborhood has a higher access to radial highways, the median value of houses is lower compared to those with lower.

```
Boston$far <- ifelse(Boston$rad == 24,TRUE, FALSE)
Boston$far <- as.factor(Boston$far)
plot(Boston$far,
     Boston$medv,
     main = 'Far vs. Medv',
     xlab = 'Far',
     ylab = 'Medv')
```



Step 10

The neighborhood with the highest median value is on row 162. The crime rate per capita at this neighborhood is below city average. At 7.489 rooms per house, this neighborhood has above average and more than

3/4 of the whole city room count per house. This neighborhood is also placed at the lowest percentage of lower status of the population in the neighborhood.

```
q10data <- Boston[,c(1,6,13,14)] #Split required columns into a different table
summary(q10data) #Get the summary of required columns
```

```
##       crim                rm            lstat            medv
##  Min.   : 0.00632   Min.   :3.561   Min.   : 1.73   Min.   : 5.00
##  1st Qu.: 0.08205   1st Qu.:5.886   1st Qu.: 6.95   1st Qu.:17.02
##  Median : 0.25651   Median :6.208   Median :11.36   Median :21.20
##  Mean   : 3.61352   Mean   :6.285   Mean   :12.65   Mean   :22.53
##  3rd Qu.: 3.67708   3rd Qu.:6.623   3rd Qu.:16.95   3rd Qu.:25.00
##  Max.   :88.97620   Max.   :8.780   Max.   :37.97   Max.   :50.00
```

```
q10data[which.max(q10data$medv),] #Find and print the data for the highest median value neighborhood
```

```
##        crim    rm lstat medv
## 162 1.46336 7.489  1.73   50
```

Step 11

```
write.csv(Boston[,c(6, 14)], file ="data.csv", row.names = FALSE) #Exporting only rm and medv columns o
```