Demonstrate the program using Regression Tree.R

Harini G

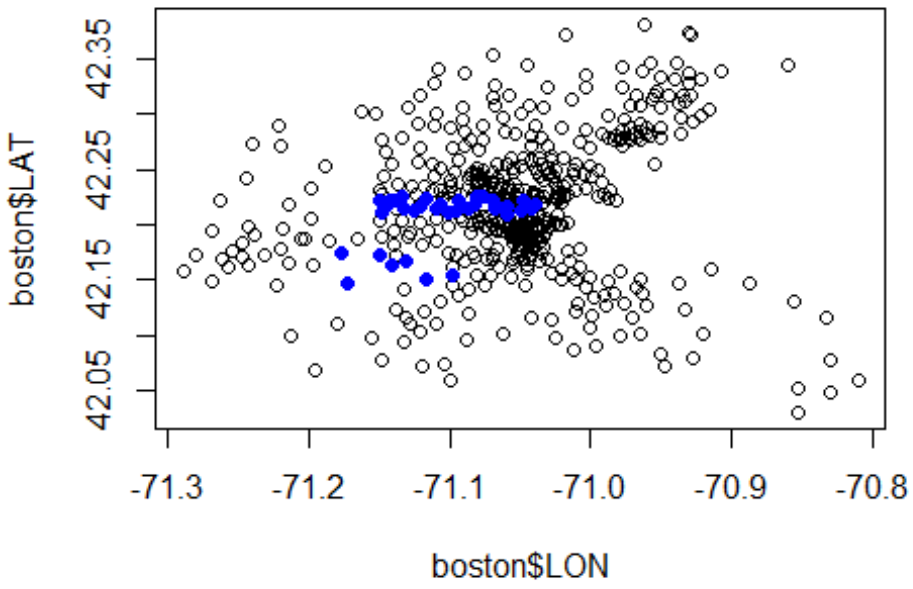
```
#Name:Harini G


#1.Download the dataset BOSTON.csv
boston=read.csv("D:/Harini(christ unniversity)/2nd sem
subjects/R/boston.csv")
head(boston)
```

```
##         TOWN TRACT      LON     LAT MEDV    CRIM ZN INDUS CHAS   NOX    RM
AGE
## 1     Nahant  2011 -70.9550 42.2550 24.0 0.00632 18  2.31    0 0.538 6.575
65.2
## 2 Swampscott  2021 -70.9500 42.2875 21.6 0.02731  0  7.07    0 0.469 6.421
78.9
## 3 Swampscott  2022 -70.9360 42.2830 34.7 0.02729  0  7.07    0 0.469 7.185
61.1
## 4 Marblehead  2031 -70.9280 42.2930 33.4 0.03237  0  2.18    0 0.458 6.998
45.8
## 5 Marblehead  2032 -70.9220 42.2980 36.2 0.06905  0  2.18    0 0.458 7.147
54.2
## 6 Marblehead  2033 -70.9165 42.3040 28.7 0.02985  0  2.18    0 0.458 6.430
58.7
##      DIS RAD TAX PTRATIO
## 1 4.0900   1 296    15.3
## 2 4.9671   2 242    17.8
## 3 4.9671   2 242    17.8
## 4 6.0622   3 222    18.7
## 5 6.0622   3 222    18.7
## 6 6.0622   3 222    18.7
```

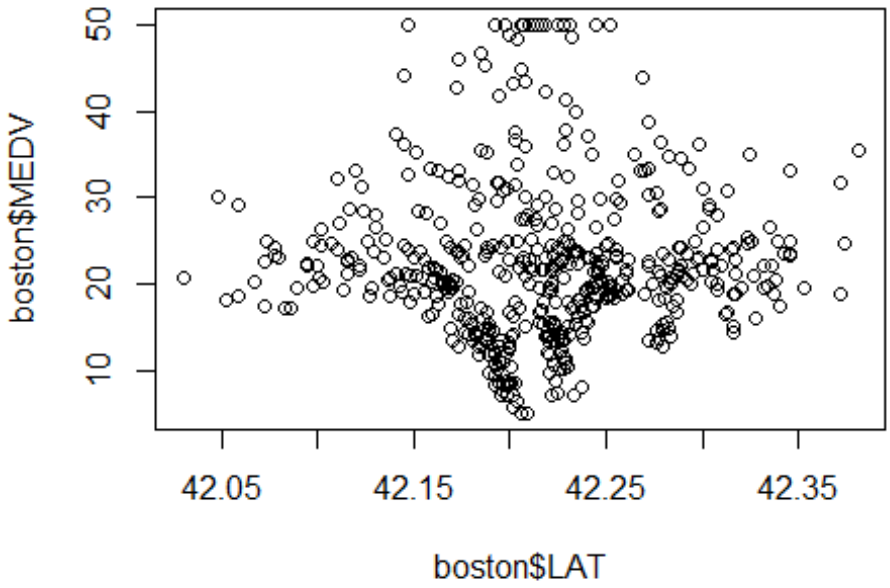
```
#2.MEDV is the output /target variable i.e price of the house to be predicted
#the target variable is continious. so, its best to use regression model.
summary(boston$MEDV)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.00   17.02   21.20   22.53   25.00   50.00
```

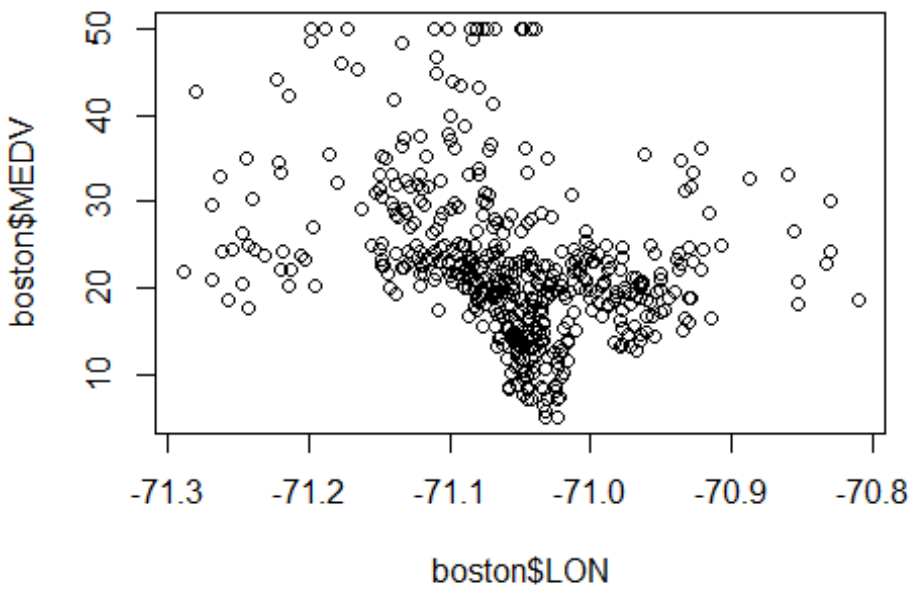
```
#3. Using the plot commands, plot the latitude and longitude of each of our
census tracts
plot(boston$LON,boston$LAT)

#4. Show all the points that lie along the Charles River in a blue colour.
points(boston$LON[boston$CHAS==1], boston$LAT[boston$CHAS==1], col="blue",
pch=19)
```



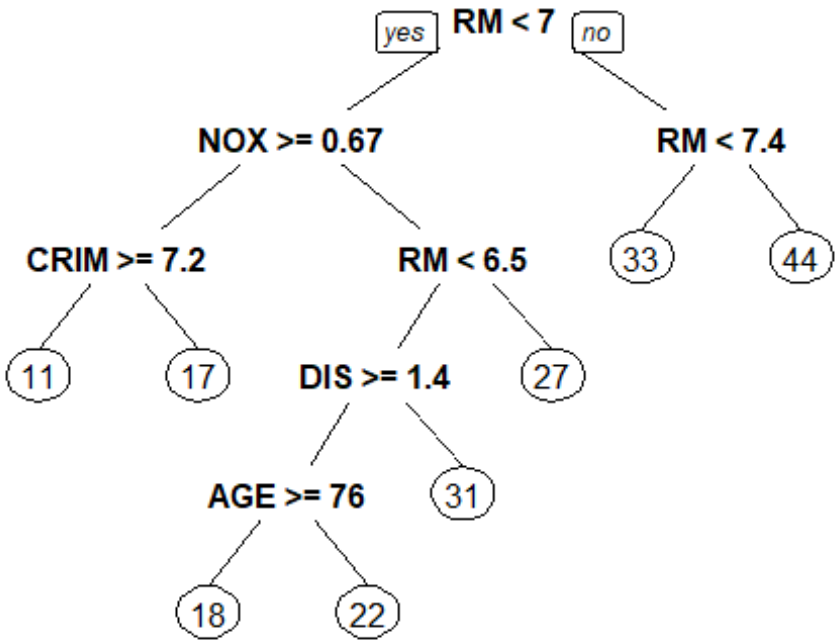

```
#5.Apply Linear Regression by plotting the relationship between latitude and
house prices and the longitude and the house prices.
plot(boston$LAT, boston$MEDV)
```



```
plot(boston$LON, boston$MEDV)
```




```
#6.Apply Regression Tree to the problem and draw conclusions from it.
#install.packages("rpart")
library(rpart)

## Warning: package 'rpart' was built under R version 4.0.4

#install.packages("rpart.plot")
library(rpart.plot)

## Warning: package 'rpart.plot' was built under R version 4.0.4

#install.packages("caTools")
library(caTools)

## Warning: package 'caTools' was built under R version 4.0.4

set.seed(100)
split = sample.split(boston$MEDV, SplitRatio = 0.7)
train = subset(boston, split==TRUE)
test = subset(boston, split==FALSE)
tree = rpart(MEDV ~ LAT + LON + CRIM + ZN + INDUS + CHAS + NOX + RM + AGE +
DIS + RAD + TAX + PTRATIO, data=train)
prp(tree)
```


yes
RM < 7
no
NOX >= 0.67
RM < 7.4
CRIM >= 7.2
RM < 6.5
33
44
11
17
DIS >= 1.4
27
AGE >= 76
31
18
22