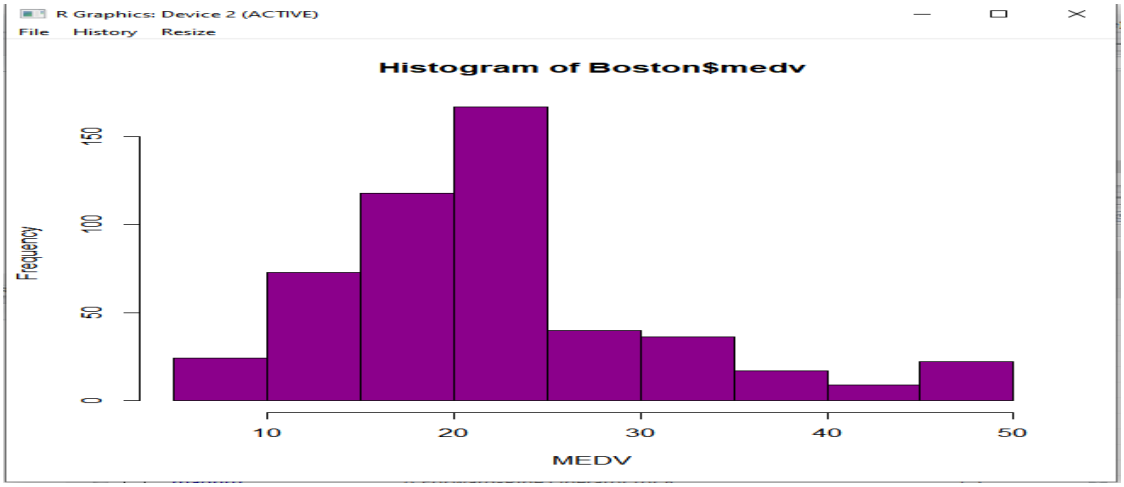
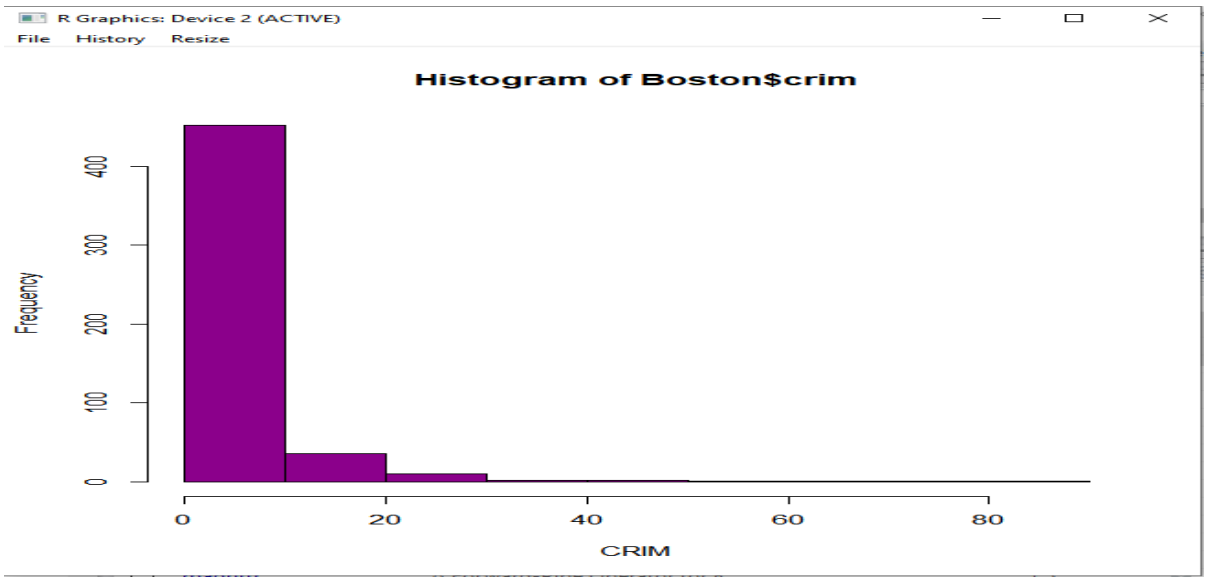
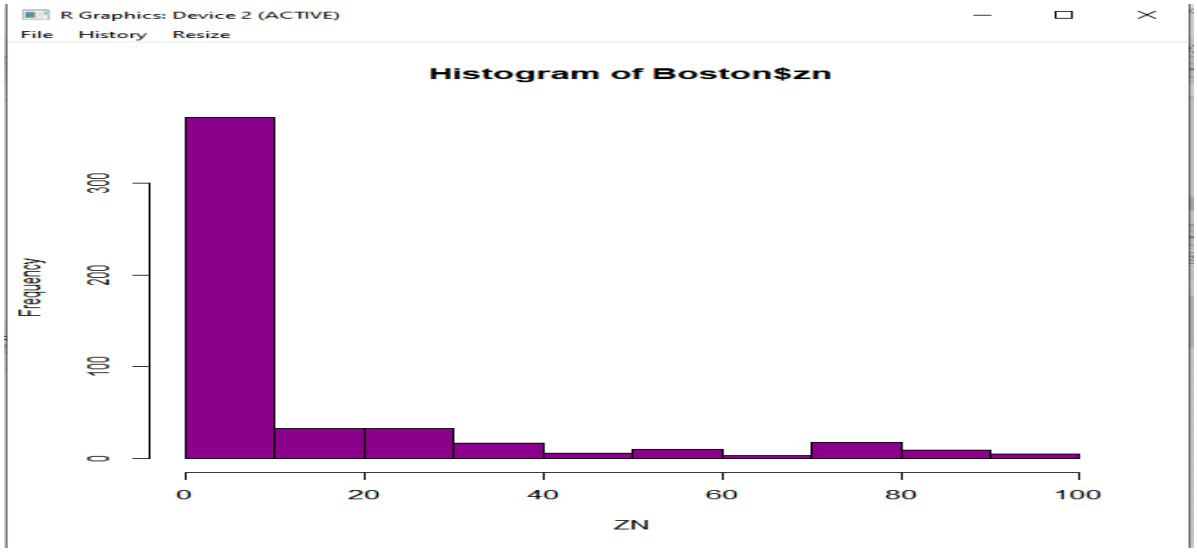
Report

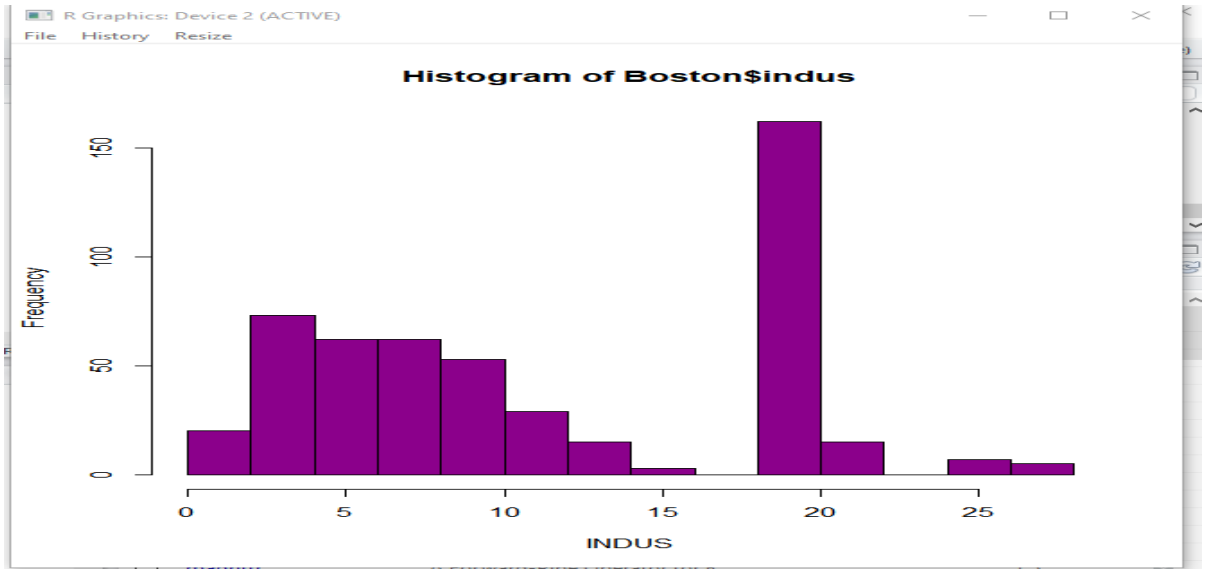
Question 2.

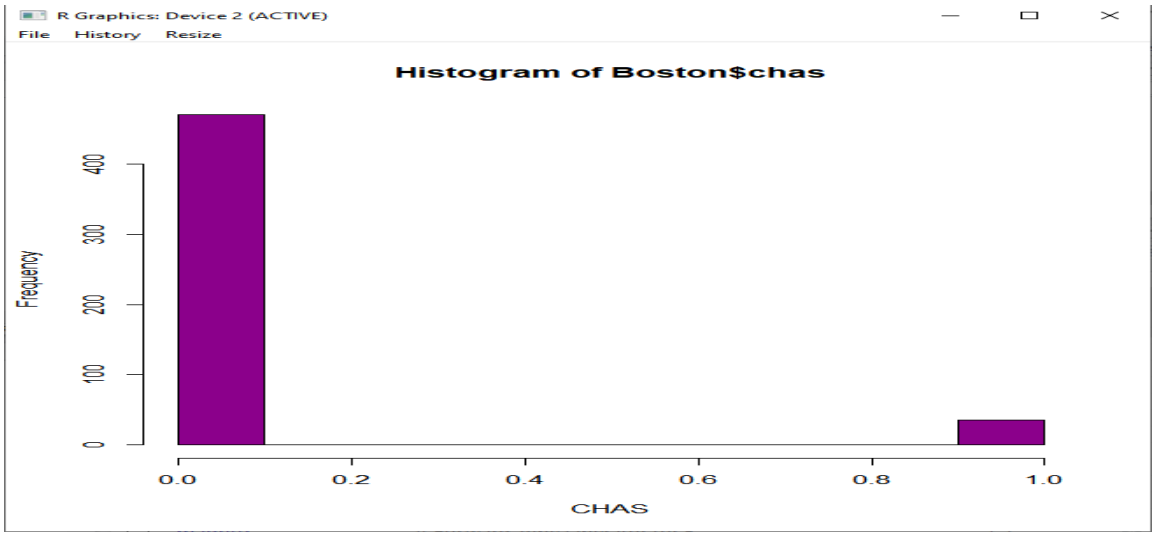
a)

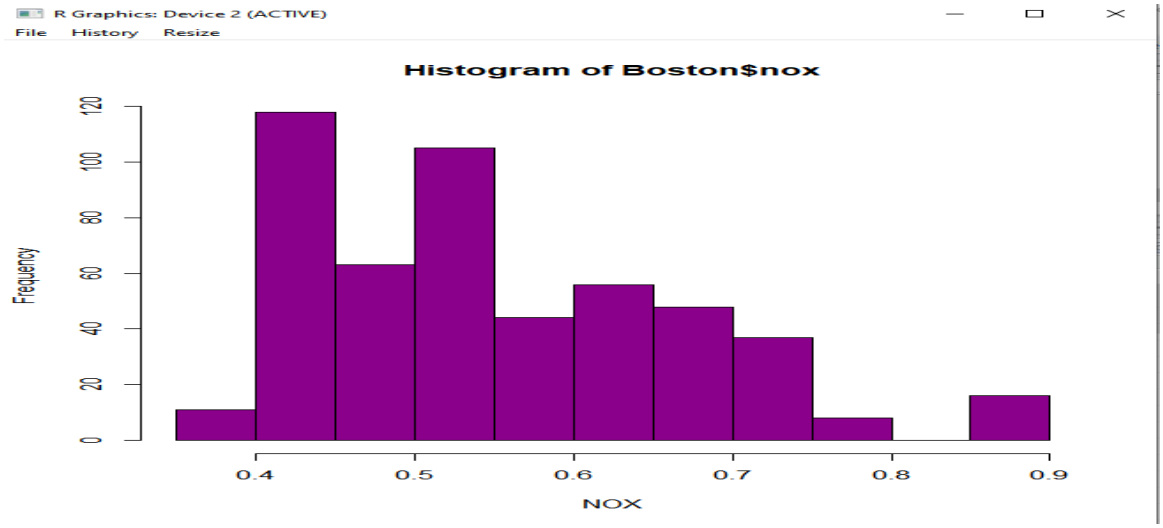


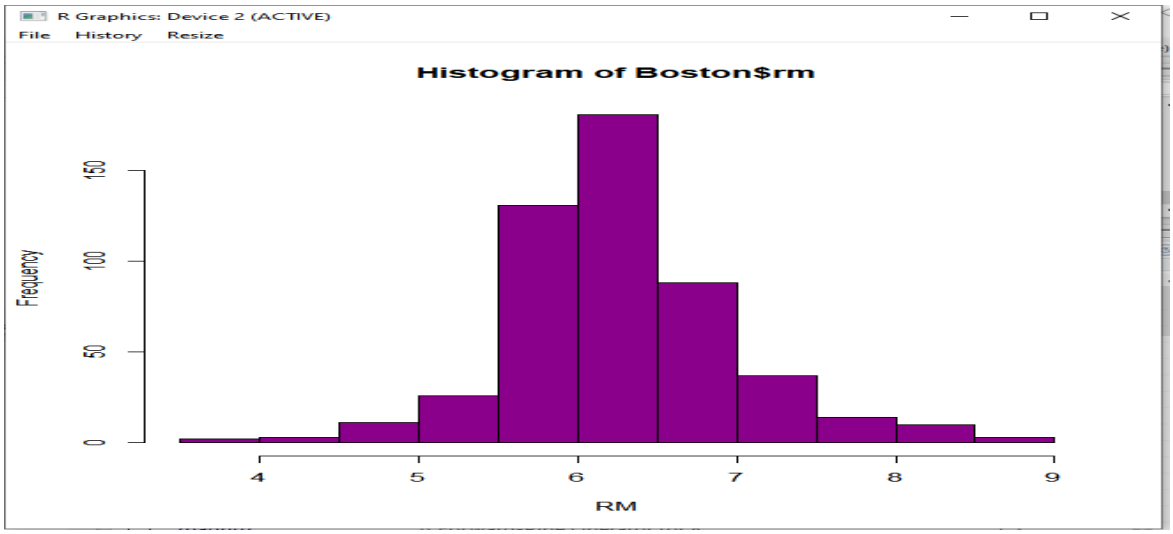


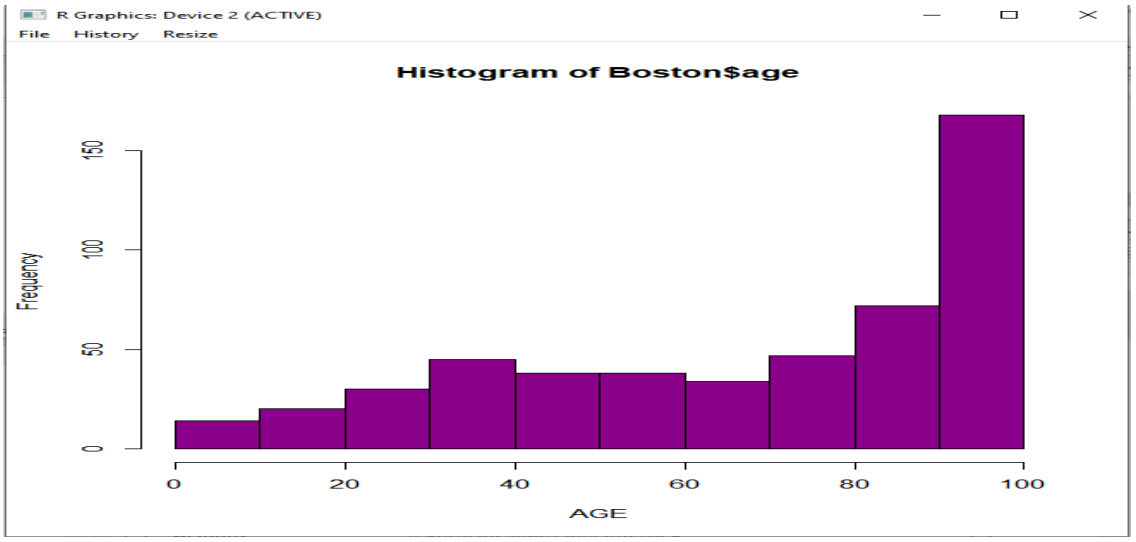

R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$zn
Frequency
0
100
200
300
0
20
40
60
80
100
ZN

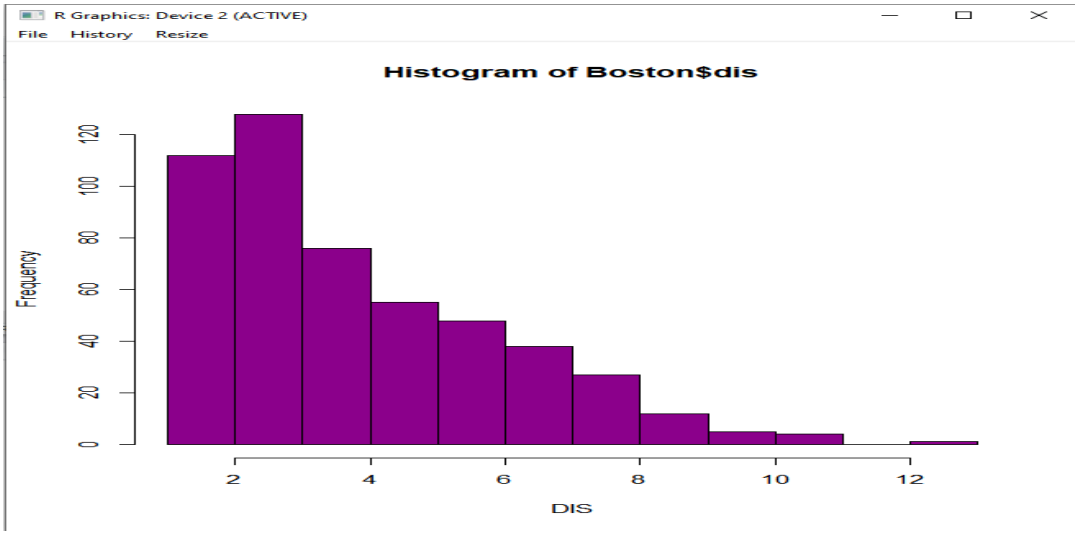

R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$indus
Frequency
0
50
100
150
0
5
10
15
20
25
INDUS

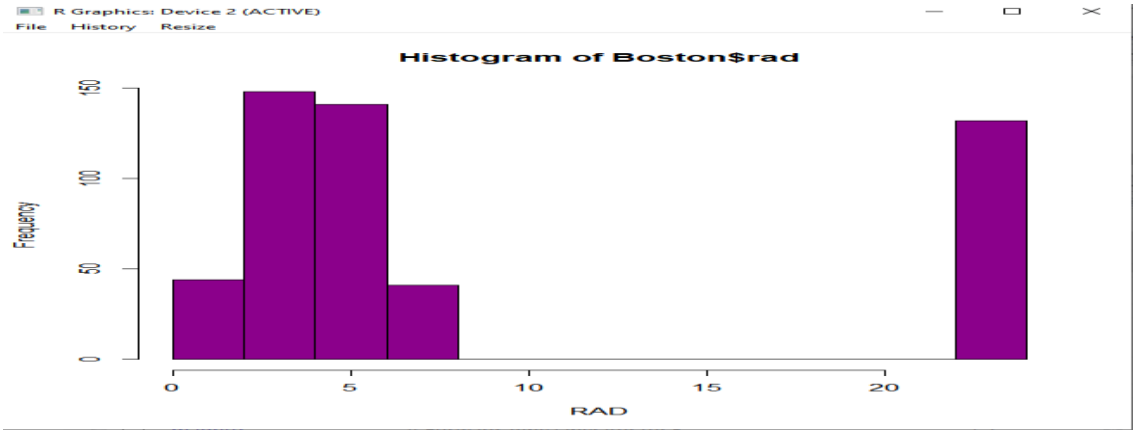

R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$chas
Frequency
0
100
200
300
400
0.0
0.2
0.4
0.6
0.8
1.0
CHAS

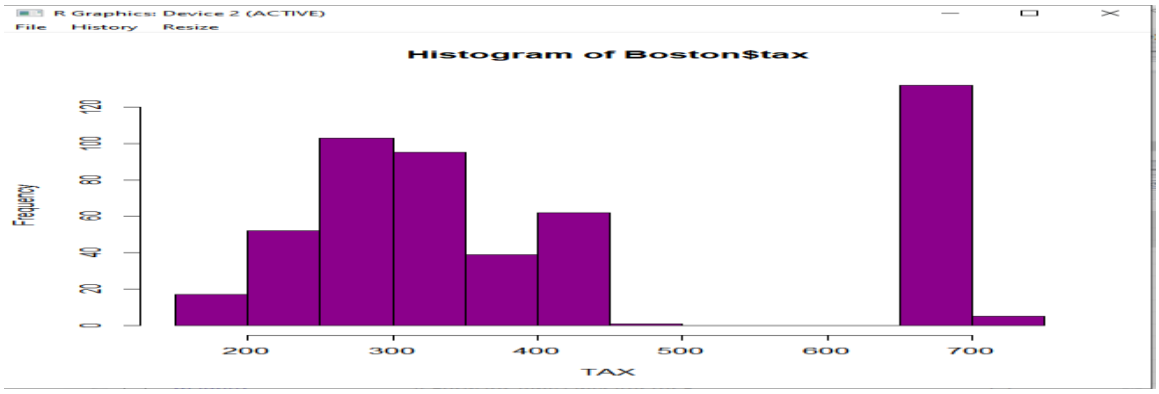

R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$nox
Frequency
0
20
40
60
80
100
120
0.4
0.5
0.6
0.7
0.8
0.9
NOX

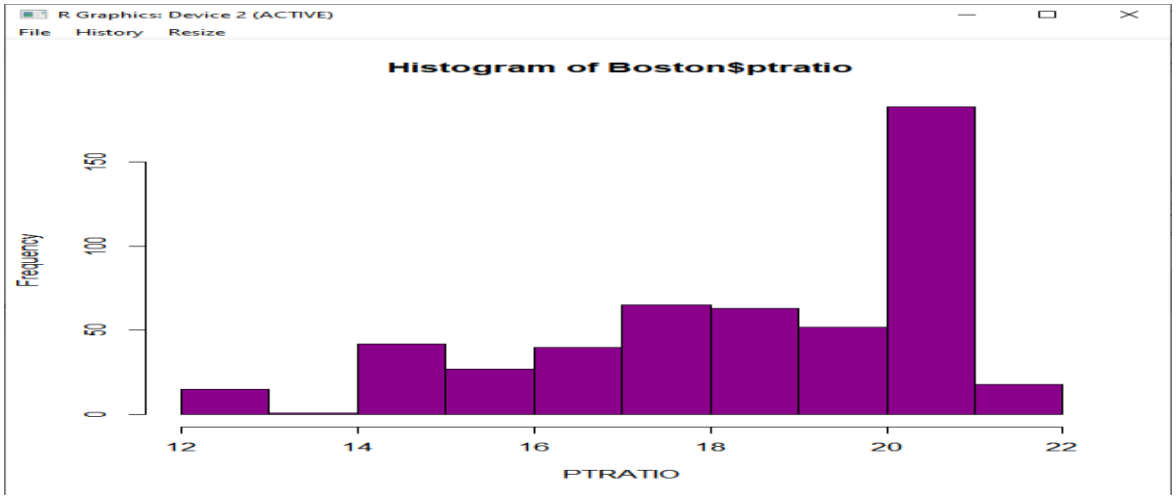

R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$rm
Frequency
0
50
100
150
4
5
6
7
8
9
RM

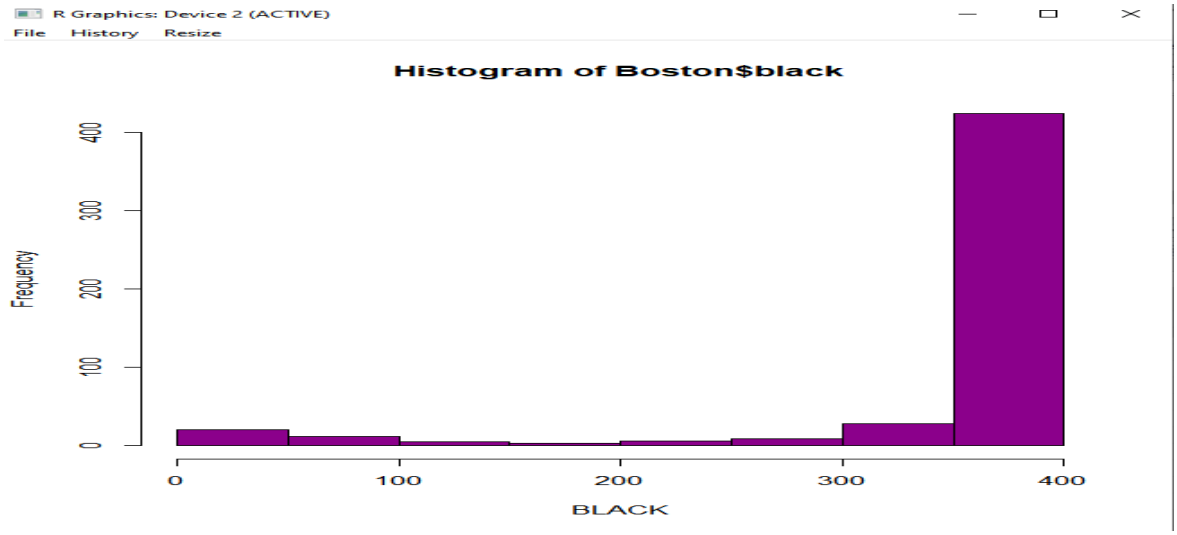

R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$age
Frequency
0
50
100
150
0
20
40
60
80
100
AGE

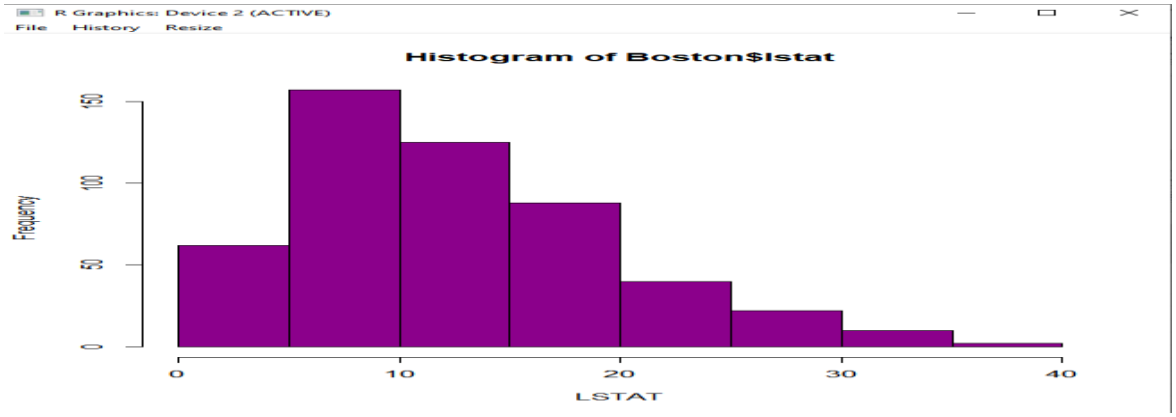

R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$dis
Frequency
0
20
40
60
80
100
120
2
4
6
8
10
12
DIS


R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$rad
Frequency
0
50
100
150
0
5
10
15
20
RAD

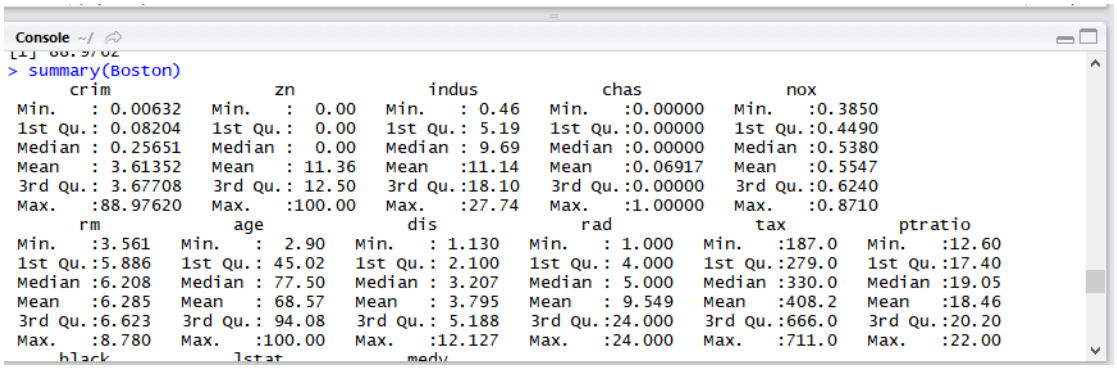

R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$tax
Frequency
0
20
40
60
80
100
120
200
300
400
500
600
700
TAX

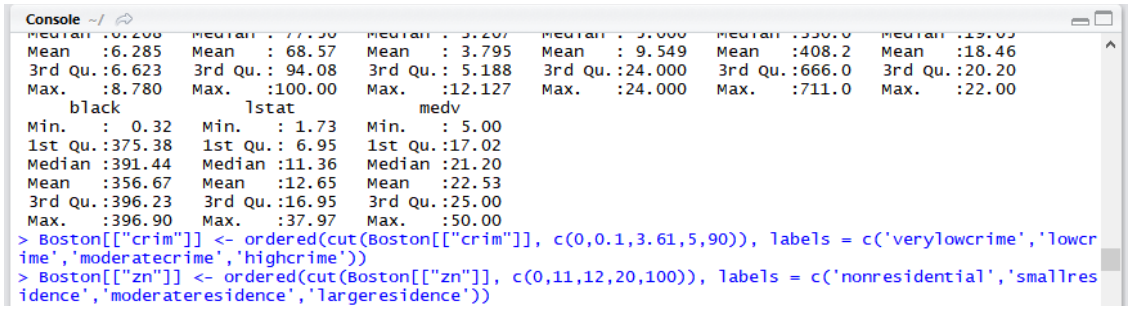

R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$ptratio
Frequency
0
50
100
150
12
14
16
18
20
22
PTRATIO


R Graphics: Device 2 (ACTIVE)
File History Resize
Histogram of Boston$black
Frequency
0
100
200
300
400
0
100
200
300
400
BLACK



Values for grouping categories were chosen based on mean , quartile, maximum and mininum values of the column data

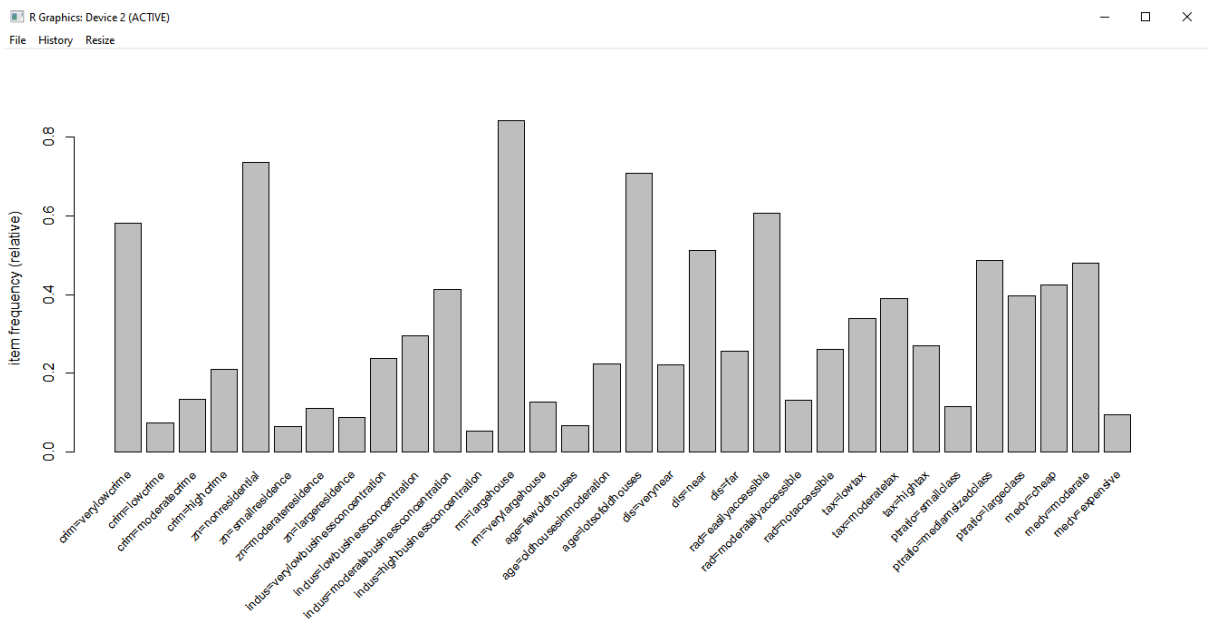




b) Item frequency Plot of binary incidence matrix of Boston housing data is as follows :

Parameters considered :

1. Support = 0.05



After Applying apriori algorithm with following parameters:

Support = 0.02, confidence = 0.8

c)

```
Console ~/
> summary(rulesLowCrimeNearCity)
set of 0 rules

> summary(rules_LowCrime)
set of 2648 rules

rule length distribution (lhs + rhs):sizes
  2   3   4   5   6   7   8   9  10
  9 116 444 798 760 395 108  17   1

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2.00    5.00    5.00    5.47    6.00   10.00

summary of quality measures:
    support          confidence          lift            count
 Min.   :0.02174   Min.   :0.8000   Min.   :1.377   Min.   : 11.0
 1st Qu.:0.02767   1st Qu.:1.0000   1st Qu.:1.721   1st Qu.: 14.0
 Median :0.03755   Median :1.0000   Median :1.721   Median : 19.0
 Mean   :0.05020   Mean   :0.9776   Mean   :1.683   Mean   : 25.4
 3rd Qu.:0.05534   3rd Qu.:1.0000   3rd Qu.:1.721   3rd Qu.: 28.0
 Max.   :0.47036   Max.   :1.0000   Max.   :1.721   Max.   :238.0

mining info:
      data ntransactions support confidence
 boston_tr           506    0.02        0.8
>
```

Student should choose house away from his work as crime rate decreases.

d)

```
> summary(rulesLowPupil_TeacherRatio)
set of 296 rules

rule length distribution (lhs + rhs):sizes
 3  4  5  6  7  8  9 10
 6 29 64 84 69 34  9  1

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.000   5.000   6.000   6.095   7.000  10.000

summary of quality measures:
    support         confidence          lift            count
 Min.   :0.02174   Min.   :0.8000   Min.   :6.979   Min.   :11.00
 1st Qu.:0.02372   1st Qu.:0.9333   1st Qu.:8.143   1st Qu.:12.00
 Median :0.02767   Median :1.0000   Median :8.724   Median :14.00
 Mean   :0.03149   Mean   :0.9636   Mean   :8.407   Mean   :15.93
 3rd Qu.:0.03360   3rd Qu.:1.0000   3rd Qu.:8.724   3rd Qu.:17.00
 Max.   :0.05929   Max.   :1.0000   Max.   :8.724   Max.   :30.00

mining info:
      data ntransactions support confidence
 boston_tr           506    0.02        0.8
>
```

```
> inspect(head(sort(rulesLowPupil_TeacherRatio, by ='lift'),n = 6))
    lhs                                                rhs                     support confidence     lift count
[1] {crim=lowcrime,
     zn=smallresidence}                             => {ptratio=smallclass} 0.02371542          1 8.724138    12
[2] {crim=lowcrime,
     indus=verylowbusinessconcentration}            => {ptratio=smallclass} 0.02371542          1 8.724138    12
[3] {crim=lowcrime,
     tax=lowtax}                                    => {ptratio=smallclass} 0.02371542          1 8.724138    12
[4] {crim=lowcrime,
     zn=smallresidence,
     indus=verylowbusinessconcentration}            => {ptratio=smallclass} 0.02371542          1 8.724138    12
[5] {crim=lowcrime,
     zn=smallresidence,
     tax=lowtax}                                    => {ptratio=smallclass} 0.02371542          1 8.724138    12
[6] {crim=lowcrime,
     zn=smallresidence,
     rad=easilyaccessible}                          => {ptratio=smallclass} 0.02371542          1 8.724138    12
>
```

He would get small residence for low crime . for getting school with low teacher-pupil ratio they would get small residence. Chances of getting a good residence is very low.

Question 3

A random data that resembles the original data with different column ‘flag’ . A flag was 0 for reference data and 1 for original data.

Decision Tree model was built.