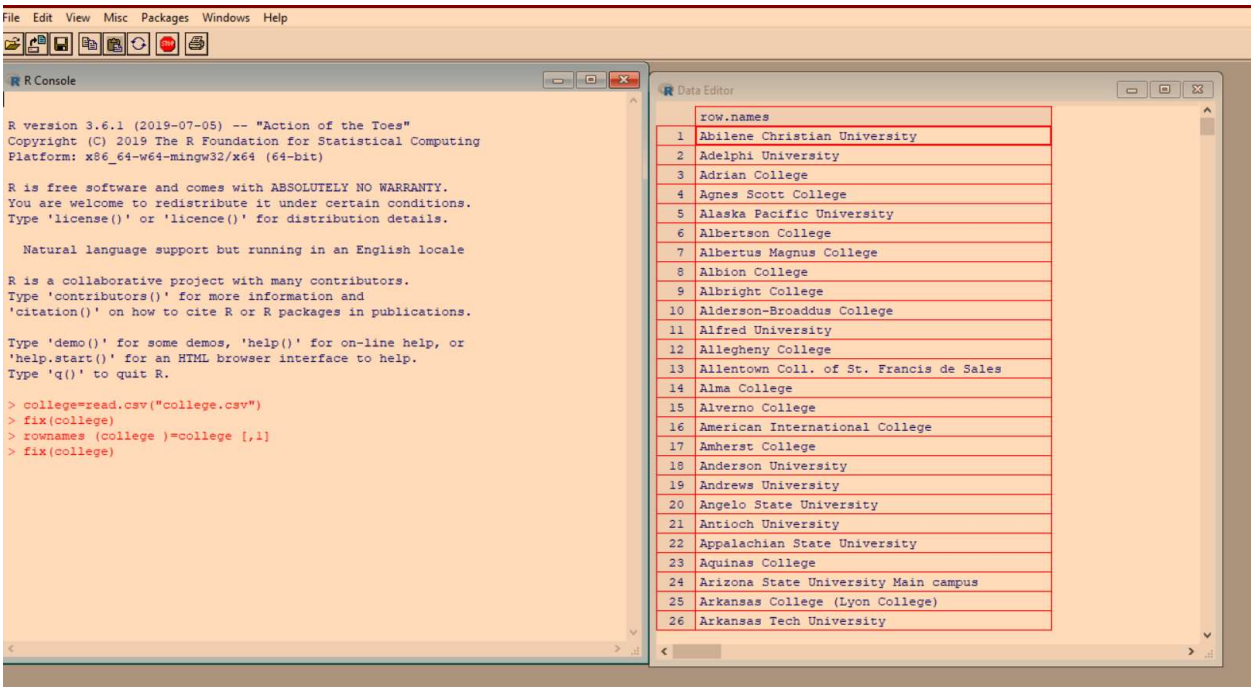
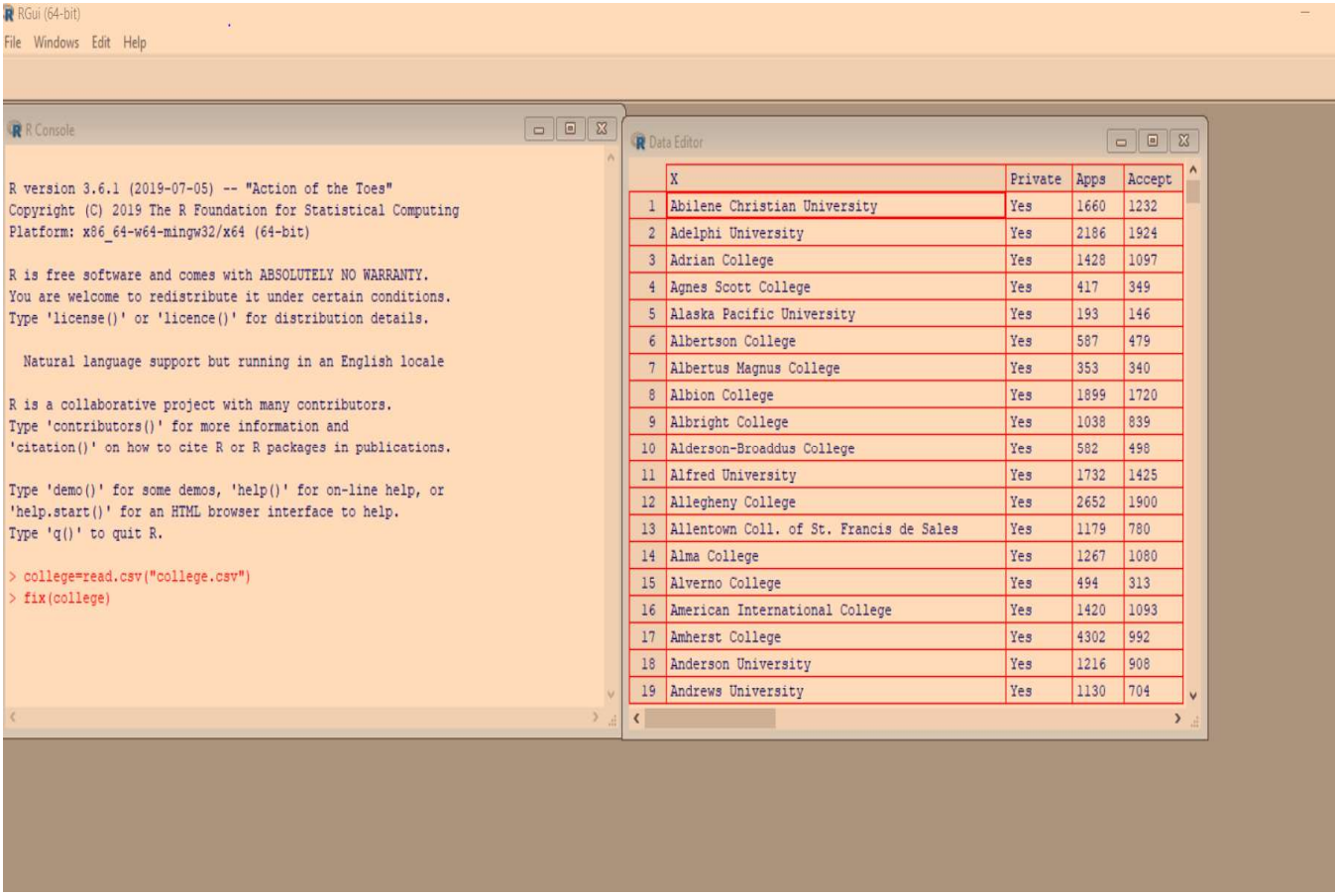
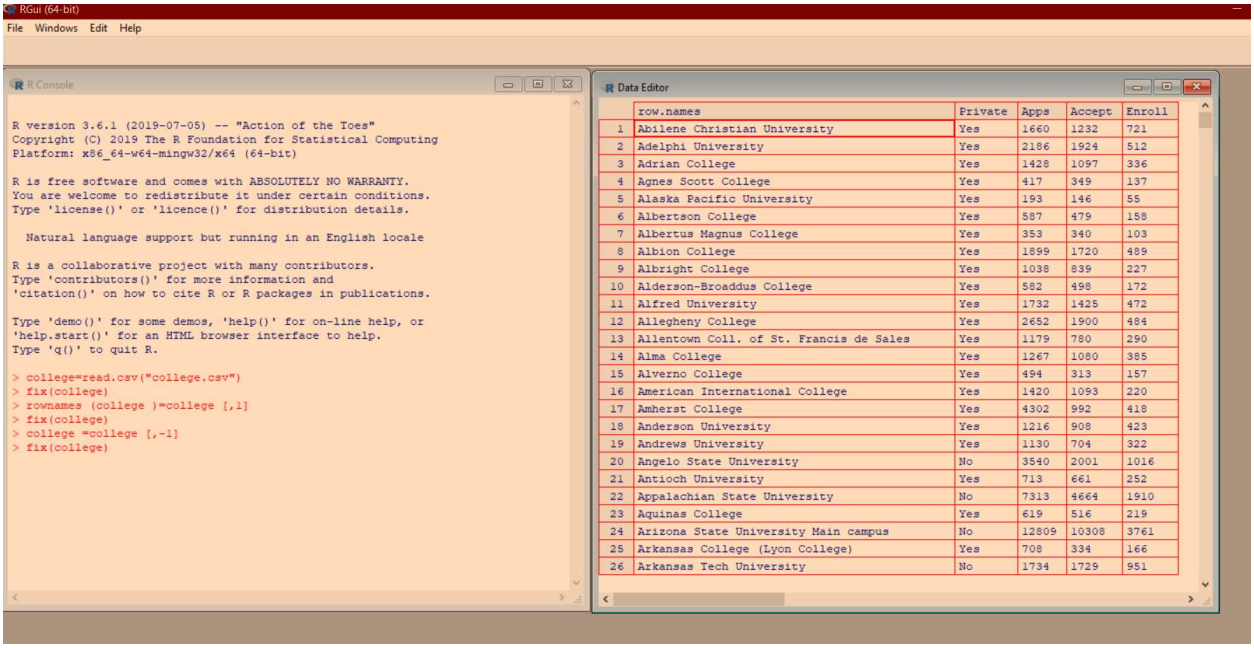
1) A) and B)

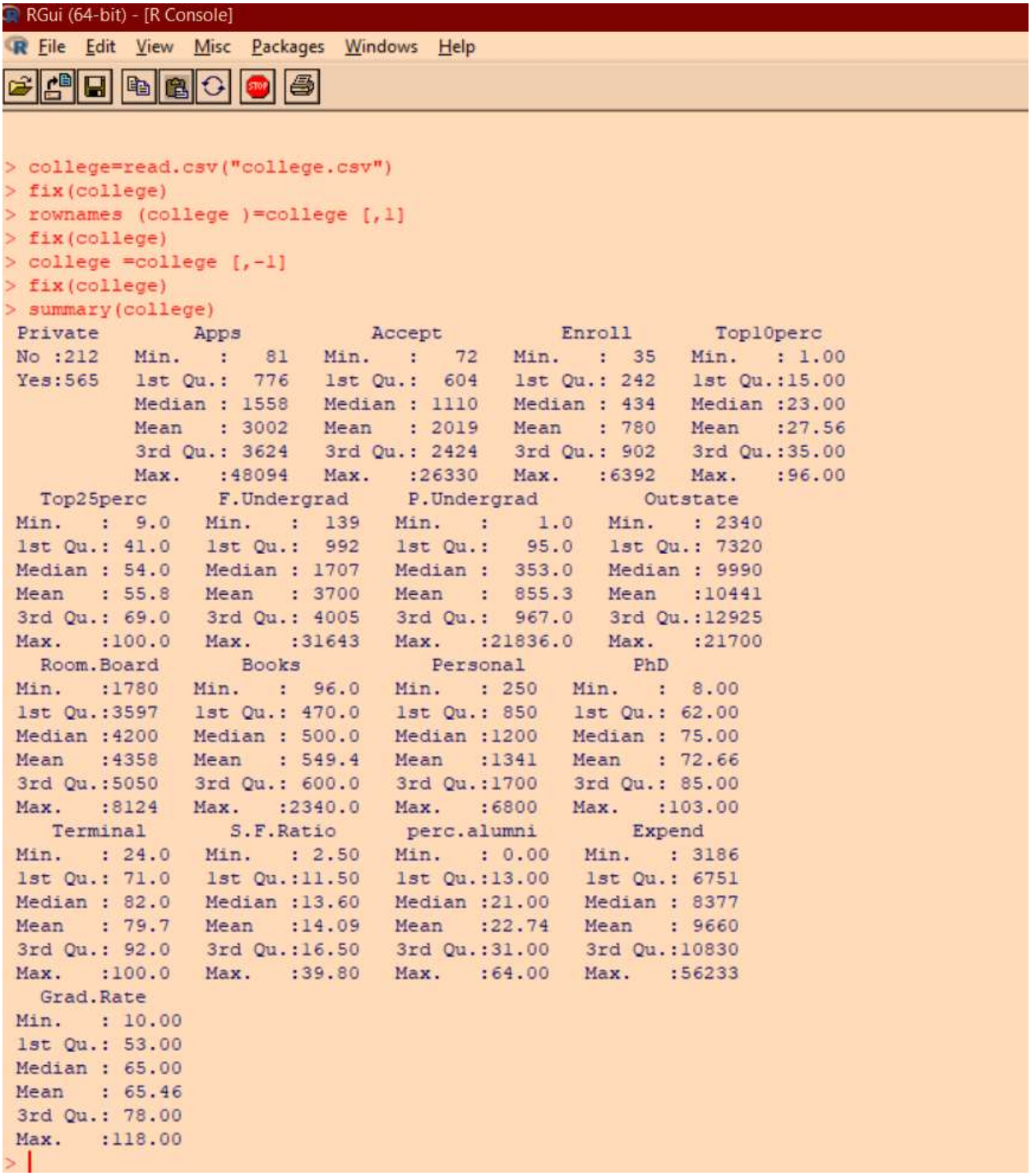

File Edit View Misc Packages Windows Help
R Console
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> college=read.csv("college.csv")
> fix(college)
> rownames (college)=college [,1]
> fix(college)
Data Editor
row.names
1 Abilene Christian University
2 Adelphi University
3 Adrian College
4 Agnes Scott College
5 Alaska Pacific University
6 Albertson College
7 Albertus Magnus College
8 Albion College
9 Albright College
10 Alderson-Broaddus College
11 Alfred University
12 Allegheny College
13 Allentown Coll. of St. Francis de Sales
14 Alma College
15 Alverno College
16 American International College
17 Amherst College
18 Anderson University
19 Andrews University
20 Angelo State University
21 Antioch University
22 Appalachian State University
23 Aquinas College
24 Arizona State University Main campus
25 Arkansas College (Lyon College)
26 Arkansas Tech University

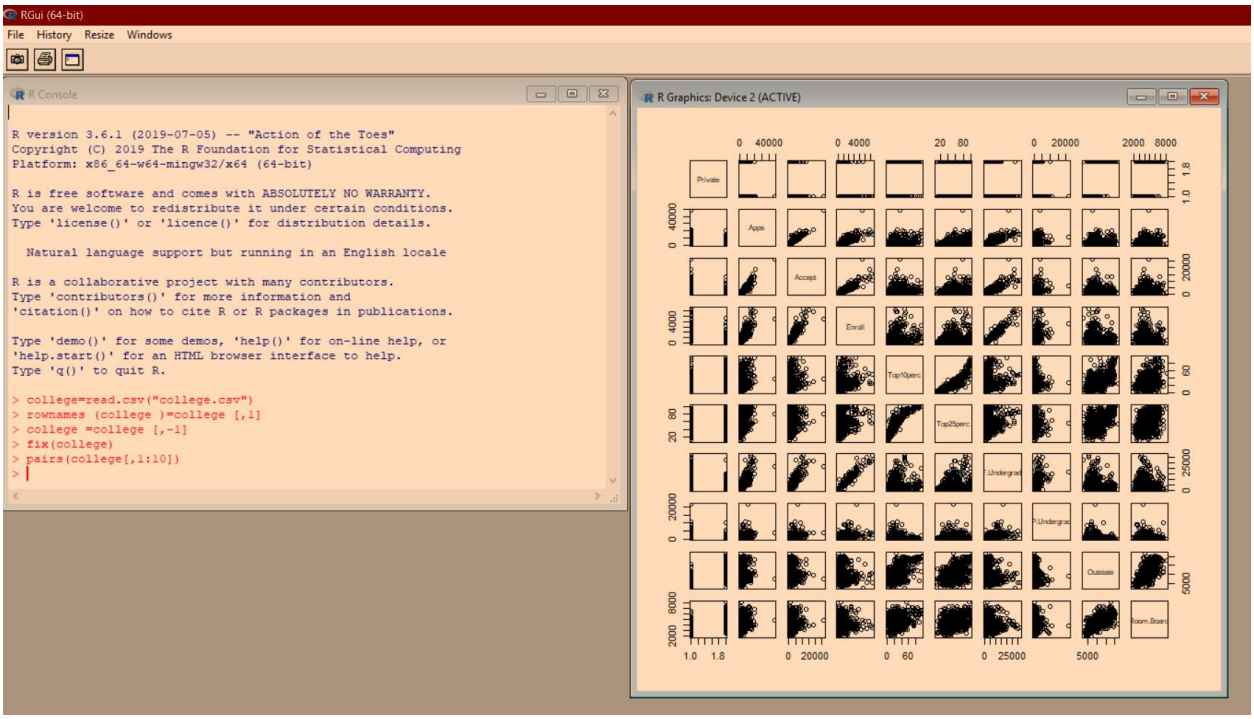

RGui (64-bit)
File Windows Edit Help
R Console
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> college=read.csv("college.csv")
> fix(college)
Data Editor
X Private Apps Accept
1 Abilene Christian University Yes 1660 1232
2 Adelphi University Yes 2186 1924
3 Adrian College Yes 1428 1097
4 Agnes Scott College Yes 417 349
5 Alaska Pacific University Yes 193 146
6 Albertson College Yes 587 479
7 Albertus Magnus College Yes 353 340
8 Albion College Yes 1899 1720
9 Albright College Yes 1038 839
10 Alderson-Broaddus College Yes 582 498
11 Alfred University Yes 1732 1425
12 Allegheny College Yes 2652 1900
13 Allentown Coll. of St. Francis de Sales Yes 1179 780
14 Alma College Yes 1267 1080
15 Alverno College Yes 494 313
16 American International College Yes 1420 1093
17 Amherst College Yes 4302 992
18 Anderson University Yes 1216 908
19 Andrews University Yes 1130 704

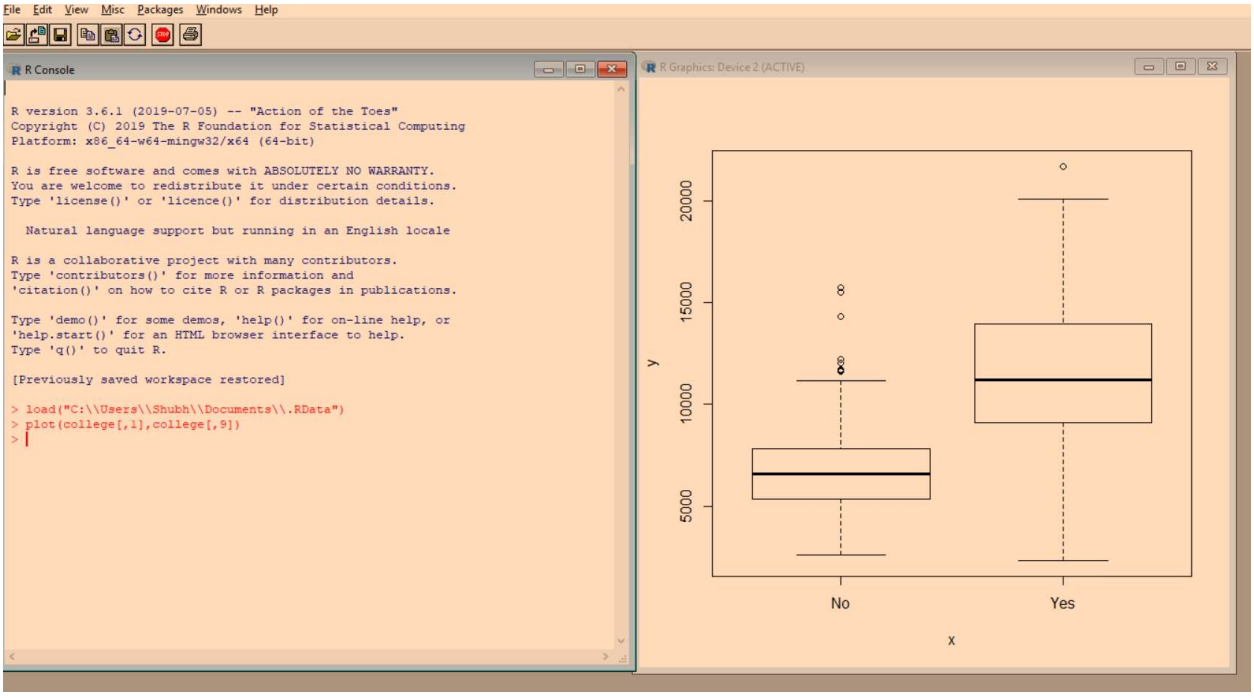


	row.names	Private	Apps	Accept	Enroll
1	Abilene Christian University	Yes	1660	1232	721
2	Adelphi University	Yes	2186	1924	512
3	Adrian College	Yes	1428	1097	336
4	Agnes Scott College	Yes	417	349	137
5	Alaska Pacific University	Yes	193	146	55
6	Albertson College	Yes	587	479	158
7	Albertus Magnus College	Yes	353	340	103
8	Albion College	Yes	1899	1720	489
9	Albright College	Yes	1038	839	227
10	Alderson-Broaddus College	Yes	582	498	172
11	Alfred University	Yes	1732	1425	472
12	Allegheny College	Yes	2652	1900	484
13	Allentown Coll. of St. Francis de Sales	Yes	1179	780	290
14	Alma College	Yes	1267	1080	385
15	Alverno College	Yes	494	313	157
16	American International College	Yes	1420	1093	220
17	Amherst College	Yes	4302	992	418
18	Anderson University	Yes	1216	908	423
19	Andrews University	Yes	1130	704	322
20	Angelo State University	No	3540	2001	1016
21	Antioch University	Yes	713	661	252
22	Appalachian State University	No	7313	4664	1910
23	Aquinas College	Yes	619	516	219
24	Arizona State University Main campus	No	12809	10308	3761
25	Arkansas College (Lyon College)	Yes	708	334	166
26	Arkansas Tech University	No	1734	1729	951

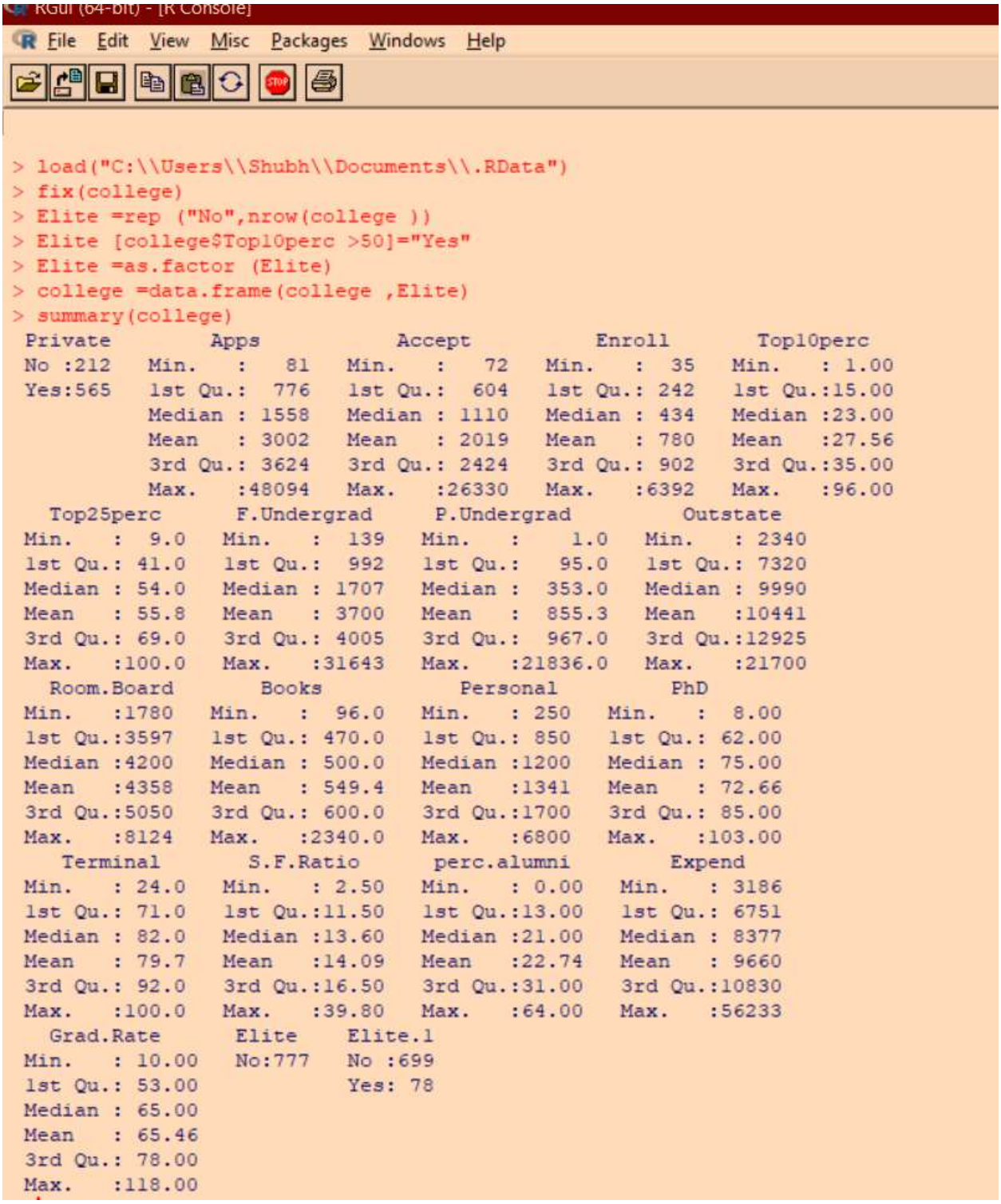
C) i), ii), iii)


RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
> college=read.csv("college.csv")
> fix(college)
> rownames (college)=college [,1]
> fix(college)
> college =college [,-1]
> fix(college)
> 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
>

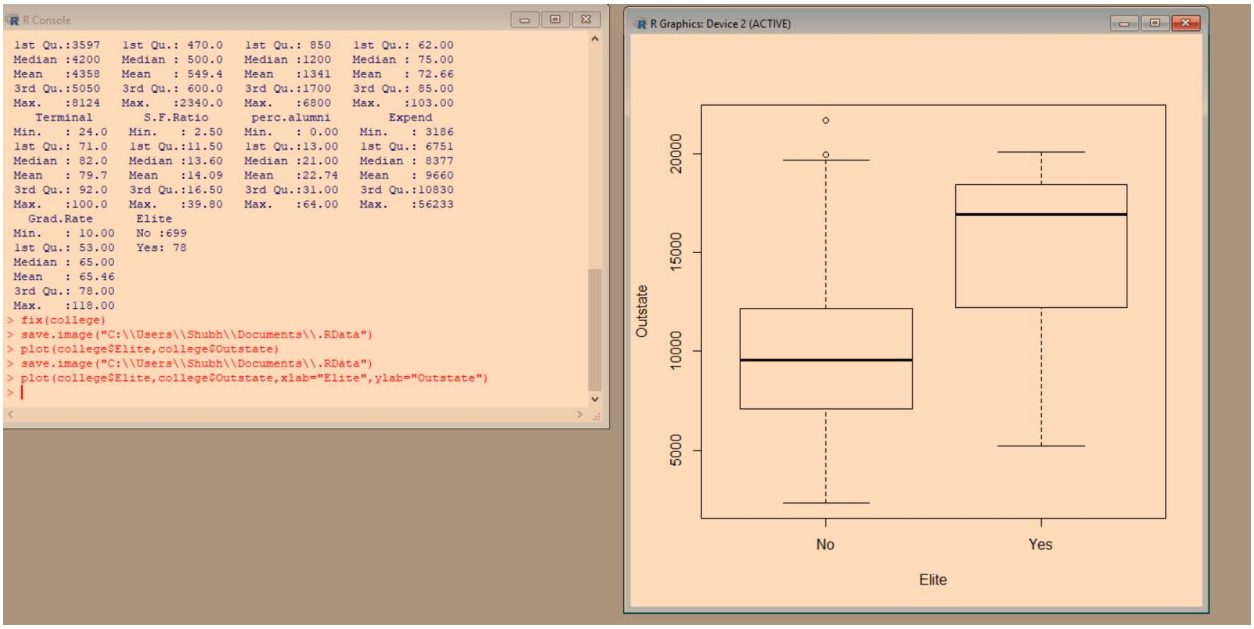

RGui (64-bit)
File History Resize Windows
R Console
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> college=read.csv("college.csv")
> rownames (college)=college [,1]
> college =college [,-1]
> fix(college)
> pairs(college[,1:10])
R Graphics: Device 2 (ACTIVE)
Private
Apps
Accept
Enroll
Top10perc
Top25perc
Undergrad
P.Undergrad
Outstate
Room.Board


File Edit View Misc Packages Windows Help
R Console
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Previously saved workspace restored]
> load("C:\\Users\\Shubh\\Documents\\.RData")
> plot(college[,1],college[,9])
R Graphics: Device 2 (ACTIVE)
y
x
No
Yes

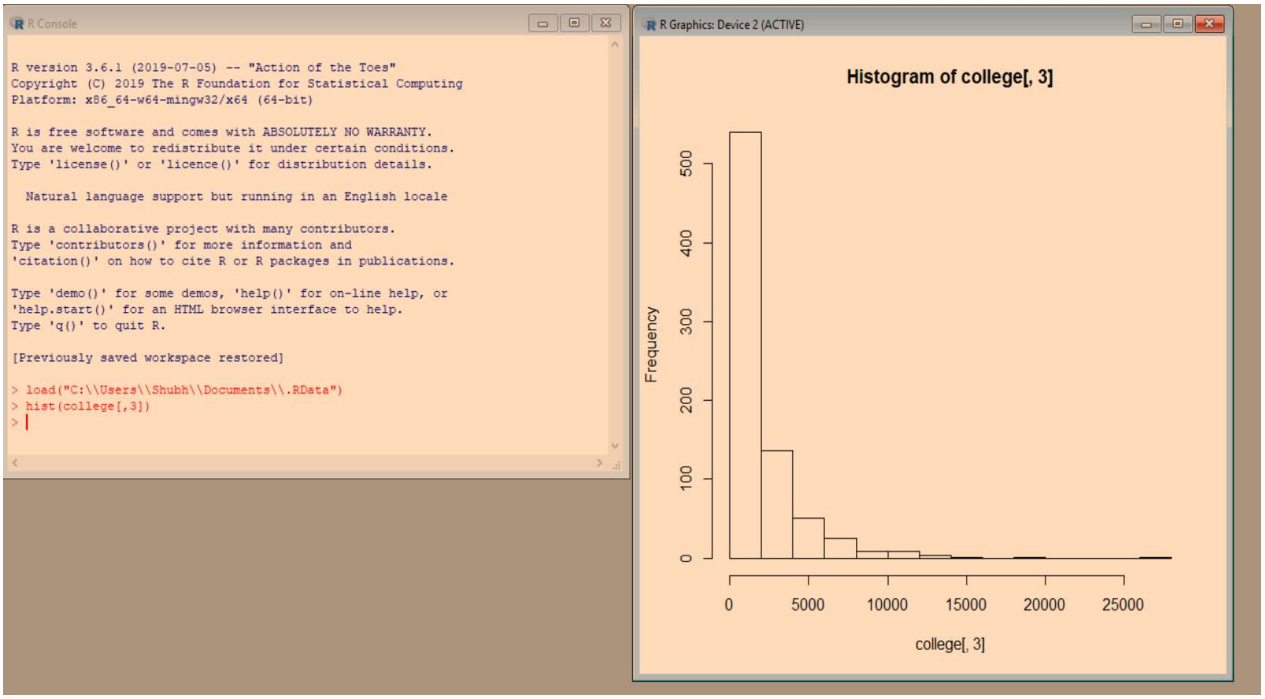
C) iv), v), vi)

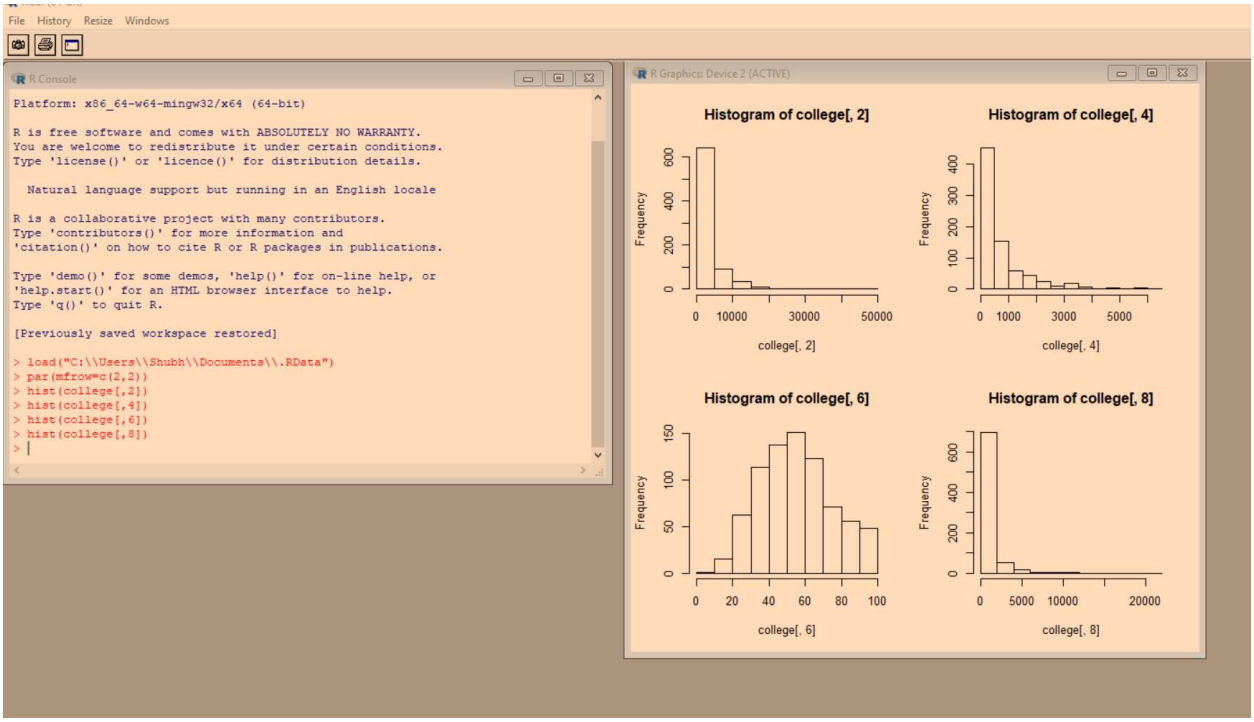


So, there is total 699 Elite universities.

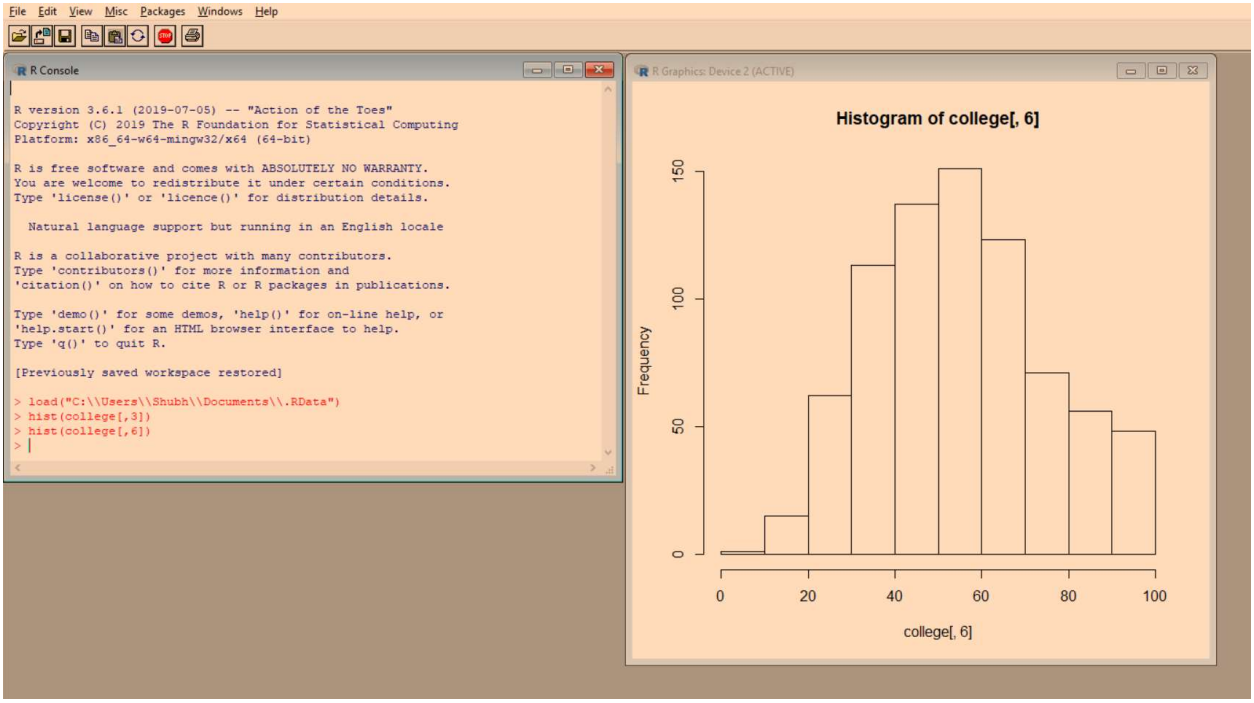

R Console
1st Qu.:3597 1st Qu.: 470.0 1st Qu.: 850 1st Qu.: 62.00
Median :4200 Median : 500.0 Median :1200 Median : 75.00
Mean :4358 Mean : 549.4 Mean :1341 Mean : 72.66
3rd Qu.:5050 3rd Qu.: 600.0 3rd Qu.:1700 3rd Qu.: 85.00
Max. :8124 Max. :2340.0 Max. :6800 Max. :103.00
Terminal S.F.Ratio perc.alumni 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
> fix(college)
> save.image("C:\\Users\\Shubh\\Documents\\.RData")
> plot(college$Elite,college$Outstate)
> save.image("C:\\Users\\Shubh\\Documents\\.RData")
> plot(college$Elite,college$Outstate,xlab="Elite",ylab="Outstate")
R Graphics: Device 2 (ACTIVE)
Outstate
5000
10000
15000
20000
No
Yes
Elite

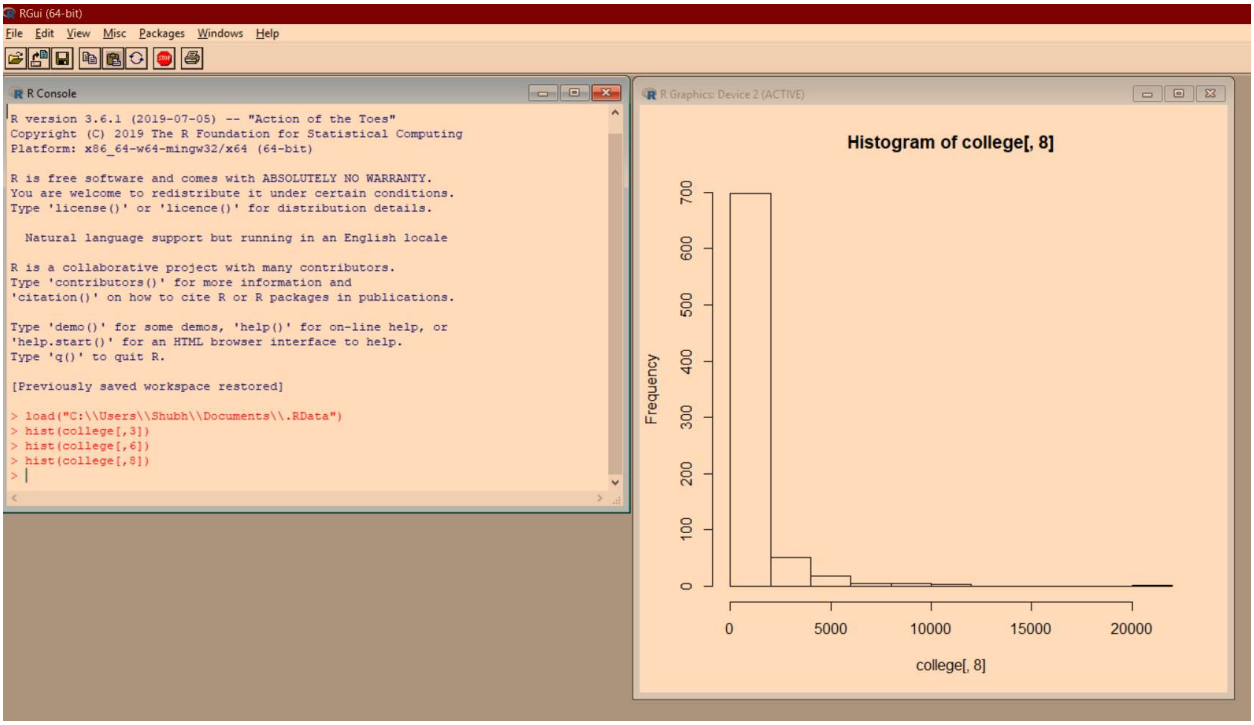
1)c) V)


R Console
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Previously saved workspace restored]
> load("C:\\Users\\Shubh\\Documents\\.RData")
> hist(college[,3])
R Graphics: Device 2 (ACTIVE)
Histogram of college[, 3]
Frequency
0
100
200
300
400
500
0
5000
10000
15000
20000
25000
college[, 3]



vi) Some more exploration:





2) a) The quantitative predictors are mpg, cylinders, displacement, horsepower, weight, acceleration & qualitative predictors are: Name.

```
> summary(auto)
      mpg          cylinders      displacement     horsepower        weight
 Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0   Min.   :1613
 1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0   1st Qu.:2225
 Median :22.75   Median :4.000   Median :151.0   Median : 93.5   Median :2804
 Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5   Mean   :2978
 3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0   3rd Qu.:3615
 Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0   Max.   :5140

  acceleration        year           origin                      name
 Min.   : 8.00   Min.   :70.00   Min.   :1.000   amc matador       :  5
 1st Qu.:13.78   1st Qu.:73.00   1st Qu.:1.000   ford pinto        :  5
 Median :15.50   Median :76.00   Median :1.000   toyota corolla    :  5
 Mean   :15.54   Mean   :75.98   Mean   :1.577   amc gremlin       :  4
 3rd Qu.:17.02   3rd Qu.:79.00   3rd Qu.:2.000   amc hornet        :  4
 Max.   :24.80   Max.   :82.00   Max.   :3.000   chevrolet chevette:  4
                                                 (Other)           :365
> range(auto[,1])
[1]  9.0 46.6
> range(auto[,2])
[1] 3 8
> range(auto[,3])
[1]  68 455
> range(auto[,4])
[1]  46 230
> range(auto[,5])
[1] 1613 5140
> range(auto[,6])
[1]  8.0 24.8
> range(auto[,7])
[1] 70 82
> range(auto[,8])
[1] 1 3
```

b) The range is:

mpg: 9.0 – 46.6

Cylinder: 3 – 8

Displacement: 68 – 455

Horsepower: 46 - 230
Weight: 1613 – 5140
Acceleration: 8.0 – 24.8
Year: 70 – 82
Origin: 1 – 3

c) mean is:

mpg: 23.45
cylinders: 5.472
displacement: 194.4
horsepower: 104.5
weight: 2978
acceleration: 15.54
Year: 75.98
origin: 1.577

Standard deviation is:

mpg: 7.805007
cylinders: 1.705783
displacement: 104.644
horsepower: 38.49116
weight: 849.4026
acceleration: 2.758864
Year: 3.683737

```
> load("C:\\Users\\Shubh\\Documents\\myauto.RData")
> sd(auto[,1])
[1] 7.805007
> sd(auto[,2])
[1] 1.705783
> sd(auto[,3])
[1] 104.644
> sd(auto[,4])
[1] 38.49116
> sd(auto[,5])
[1] 849.4026
> sd(auto[,6])
[1] 2.758864
> sd(auto[,7])
[1] 3.683737
> sd(auto[,8])
[1] 0.8055182
```

origin: 0. 8055182

d) After deleting rows new range is as below:

Range:

mpg: 11.0 - 46.6
cylinders: 3 - 8
displacement: 68 - 455
horsepower: 46 - 230
weight: 1649 - 4997
acceleration: 8.5 – 24.8
Year: 70 – 82
Origin: 1 – 3

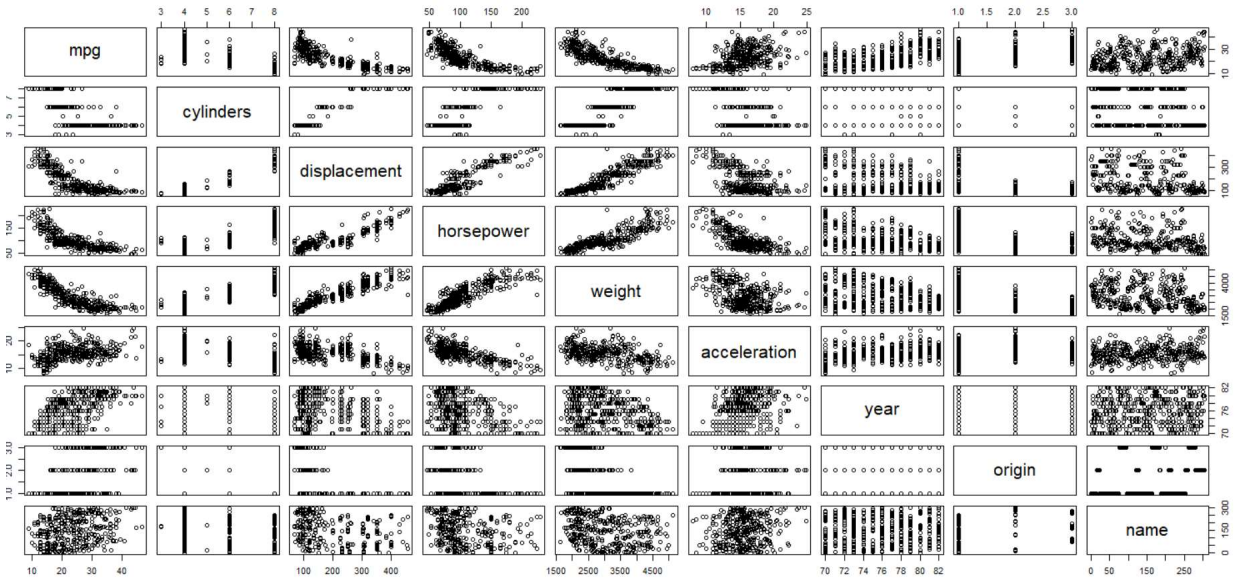
Standard deviation is:

mpg: 7.880898
cylinders: 1.658135
displacement: 99.93949
horsepower: 35.89557
weight 812.6496
acceleration: 2.693813
Year: 3.110026
Origin: 0.8193079

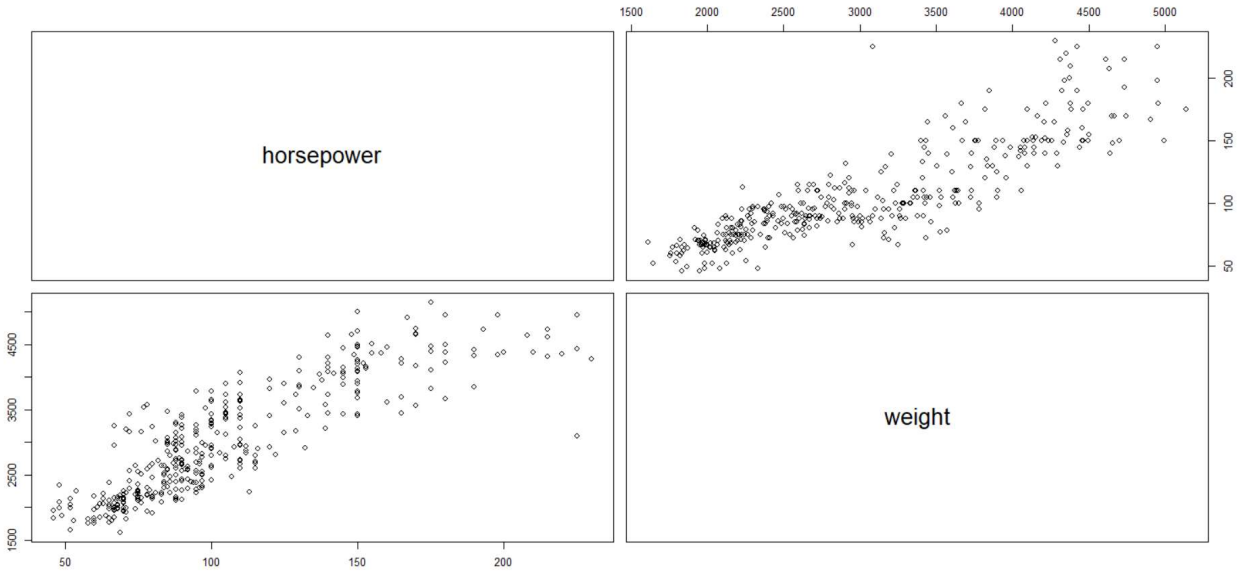
Mean:

mpg: 24.36845
cylinders: 5.381703
displacement: 187.7539
horsepower: 100.9558
weight: 2939.644
acceleration: 15.7183
Year: 77.13249
Origin: 1.599369

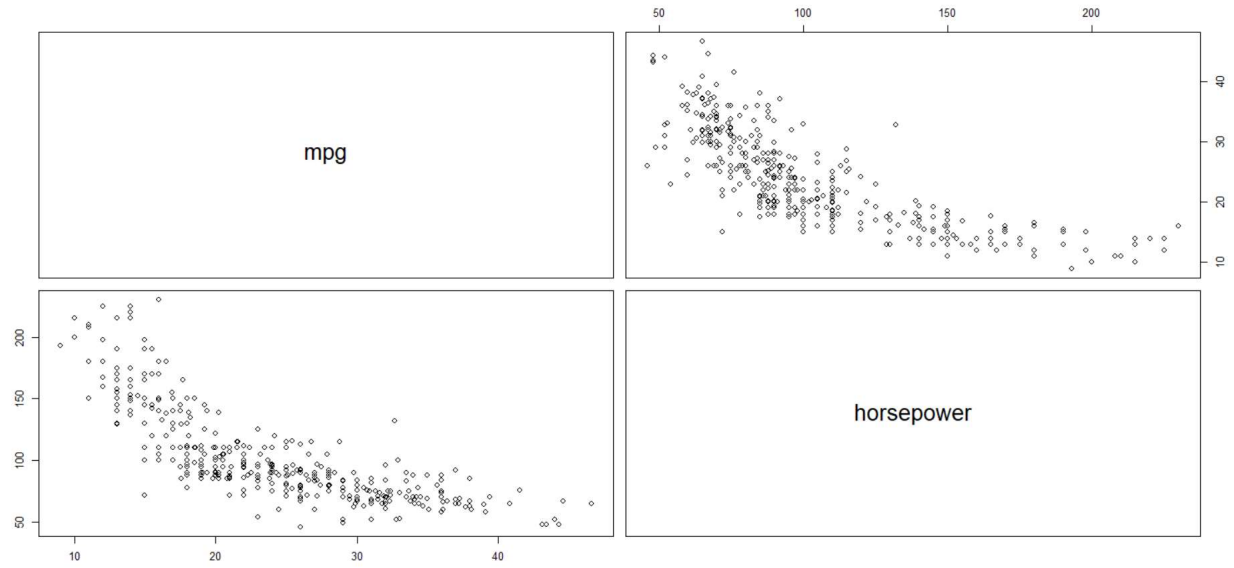
e) The overall scatterplot:



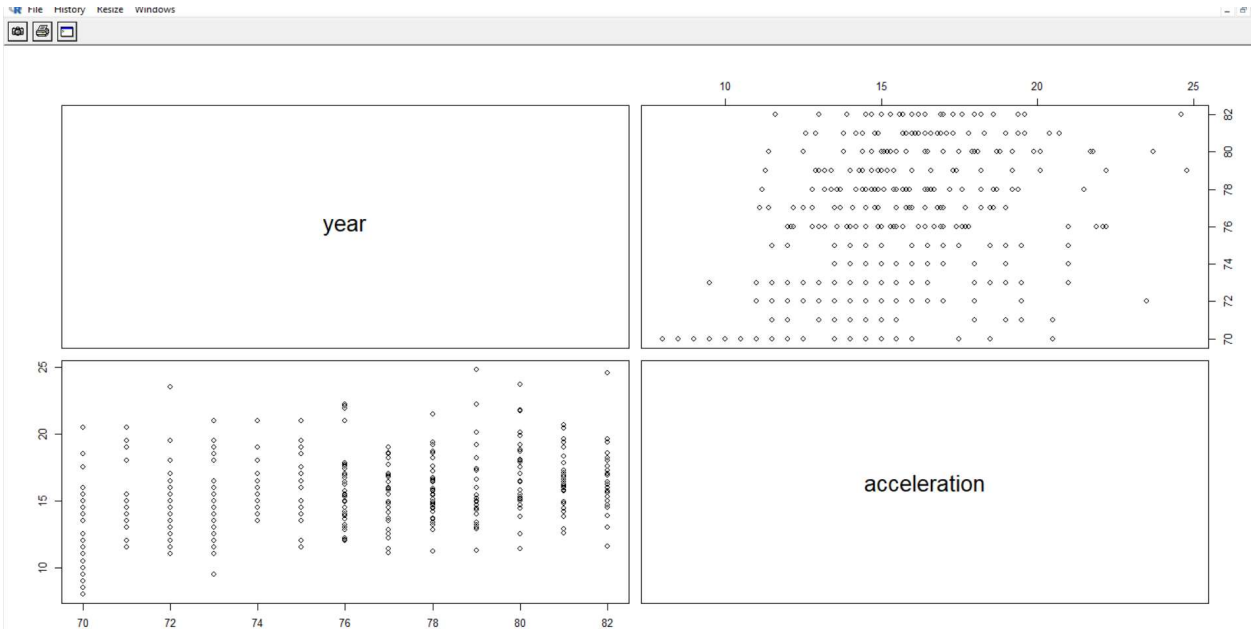
Now, if we plotweight horsepower, they have linear relationship, roughly one increases if other one does.



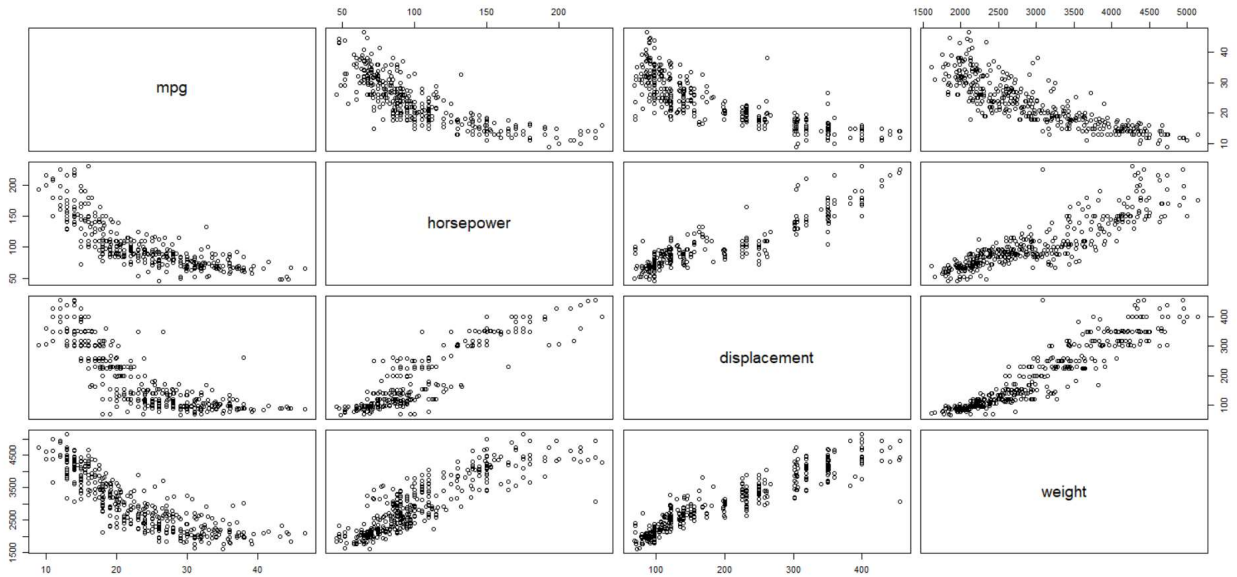
Also, if we plot mpg & horsepower, they are inversely related. i.e. One increases if the other one decreases.

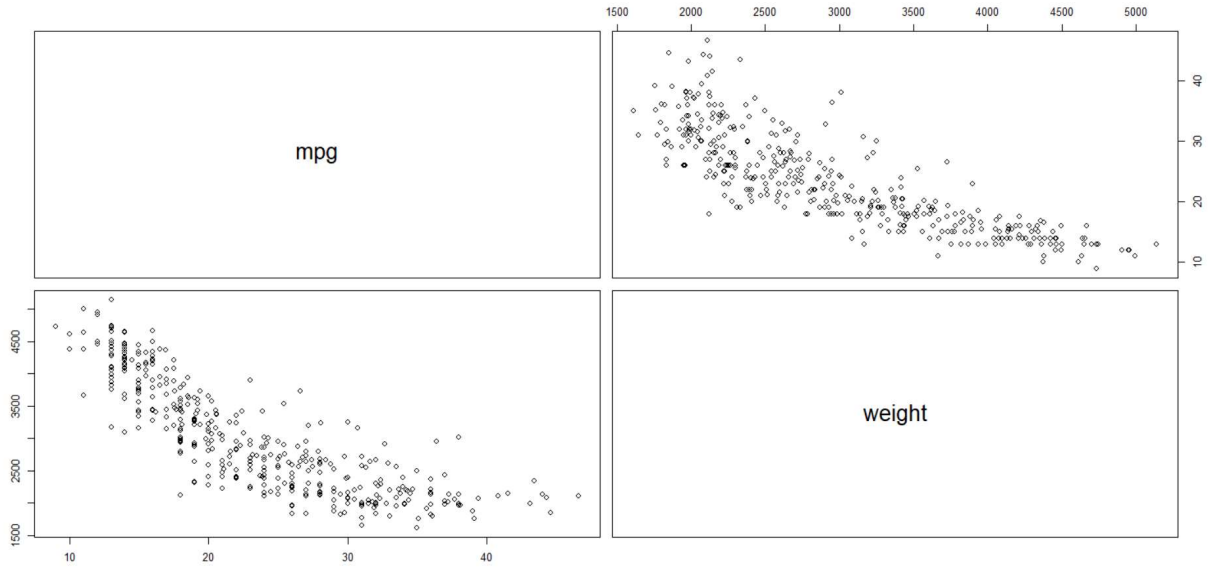


Also, if we consider year and acceleration, they are not related in a distinct manner:



f) If we look at the factors affecting mpg in scatterplot, we see mainly 3 parameters is affecting that, they are, Displacement, horsepower and weight. But weight V mpg has minimum variance among all these three, so we can select weight to predict mpg better. Also, it has to be a quadratic or cubic equation rather than linear , that we can see from the nature of plot.





3) a) The Boston data frame has 506 rows and 14 columns.

Boston {MASS}

Housing Values in Suburbs of Boston

Description

The `Boston` data frame has 506 rows and 14 columns.

Usage

```
Boston
```

Format

This data frame contains the following columns:

`crim`

per capita crime rate by town.

`zn`

proportion of residential land zoned for lots over 25,000 sq.ft.

`indus`

proportion of non-retail business acres per town.

`chas`

Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).

`nox`

nitrogen oxides concentration (parts per 10 million).

`rm`

average number of rooms per dwelling.

`age`

proportion of owner-occupied units built prior to 1940.

`age`

proportion of owner-occupied units built prior to 1940.

`dis`

weighted mean of distances to five Boston employment centres.

`rad`

index of accessibility to radial highways.

`tax`

full-value property-tax rate per \$10,000.

`ptratio`

pupil-teacher ratio by town.

`black`

1000(Bk - 0.63)^2 where *Bk* is the proportion of blacks by town.

`lstat`

lower status of the population (percent).

`medv`

median value of owner-occupied homes in \$1000s.

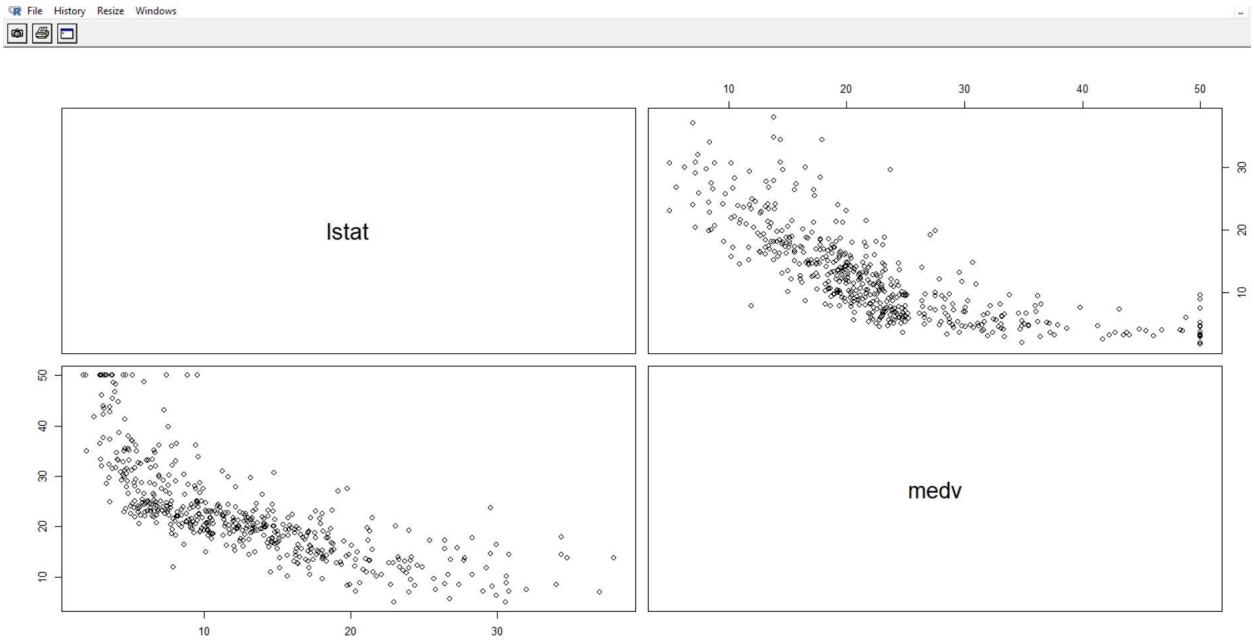
`Source`

Harrison, D. and Rubinfeld, D.L. (1978) Hedonic prices and the demand for clean air. *J. Environ. Economics and Management* **5**, 81–102.

Belsley D.A., Kuh, E. and Welsch, R.E. (1980) *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity.* New York: Wiley.

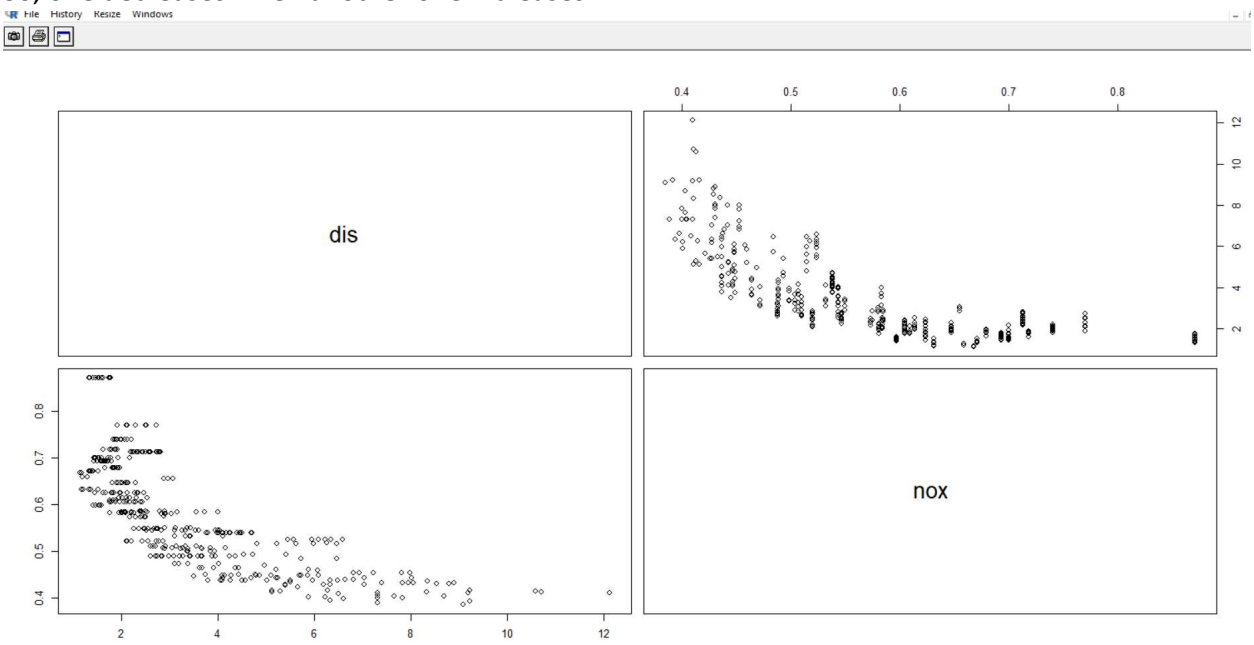
[Package *MASS* version 7.3-51.4 Index]

b) The relation between lstat (lower status of the population (percent) and medv (median value of owner-occupied homes in \$1000s) is inversely proportional. So, one decreases when another one increases.

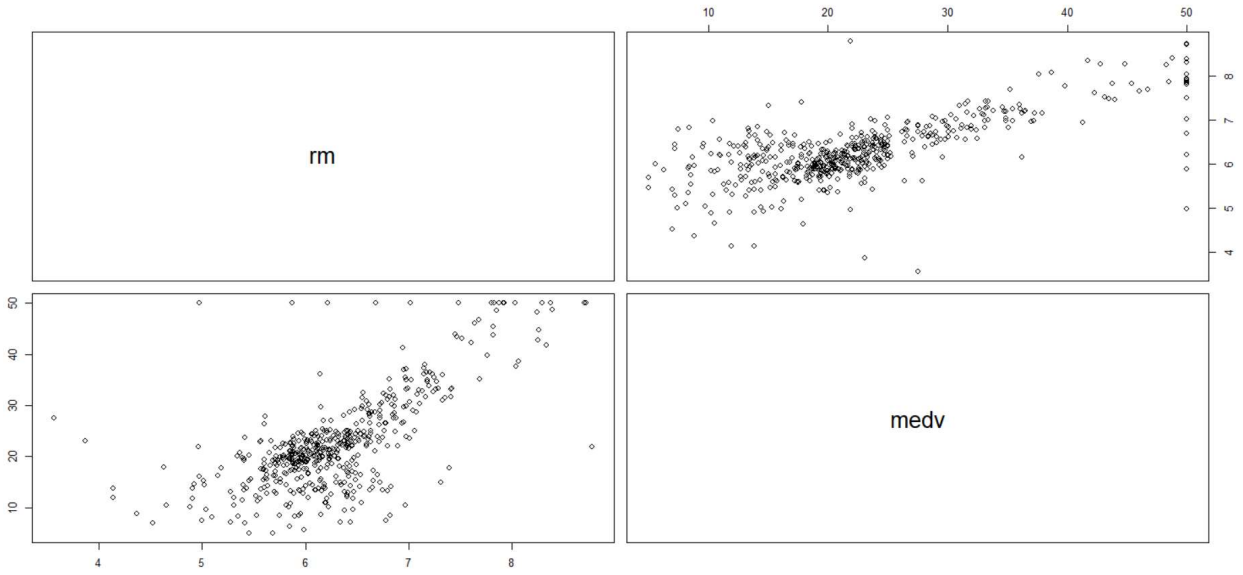


Also, the relation between dis (weighted mean of distances to five Boston employment centers). (percent) and nox (nitrogen oxides concentration (parts per 10 million) is inversely proportional.

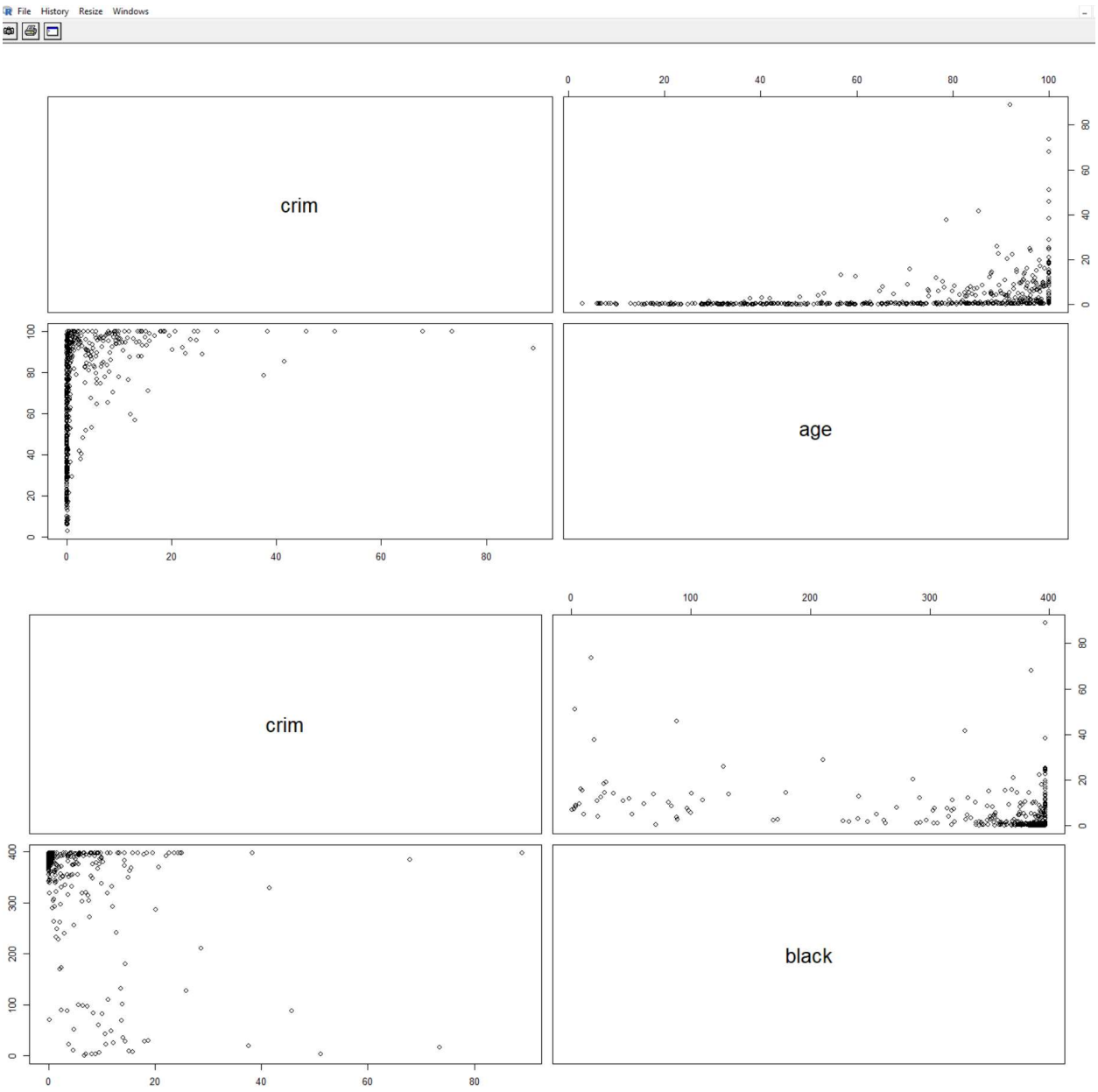
So, one decreases when another one increases.



Also, the relation between rm (average number of rooms per dwelling.) and medv (median value of owner-occupied homes in \$1000s.) is directly proportional. So, one decreases when another one decreases.

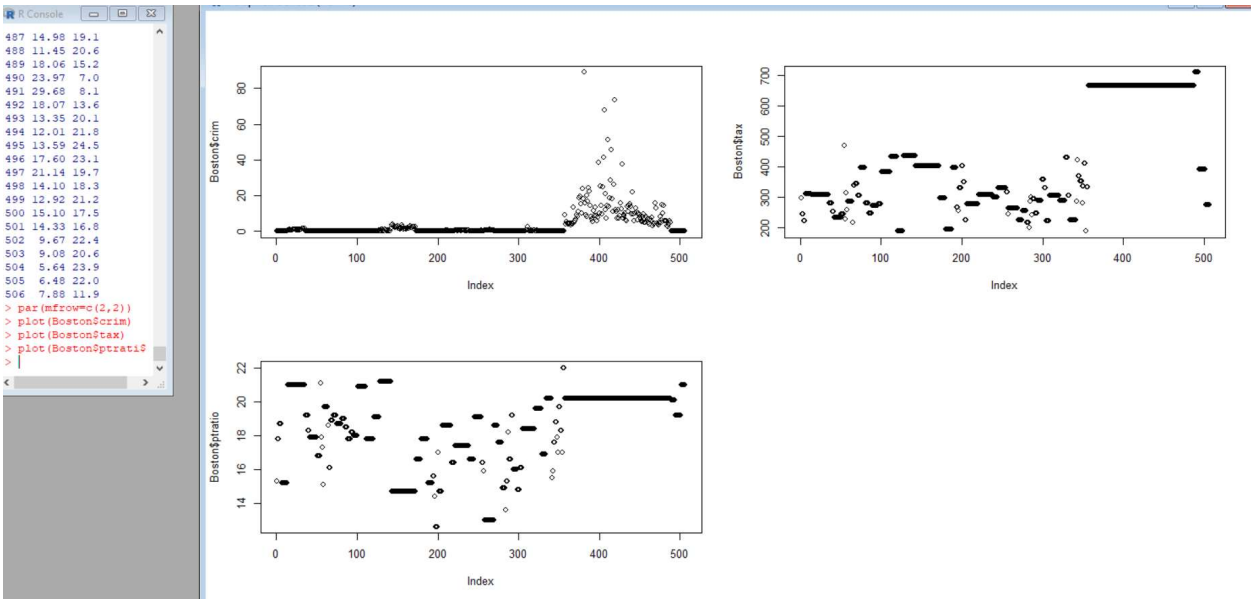


c) We can see if crim (crime) rate is plotted with age (proportion of owner-occupied units built prior to 1940), or 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town or we can infer that, per capita income rate is less than 20 in mostly. overall crime rate is less than 20.



d) Now if we plot crime rates, Tax rates, Pupil-teacher ratios with index. We can observe that:
Crime rate is highest around the sub-urb index number 400.

Crime rates is highest around the sub-urb index number 400.
Tax rate is highest around the sub-urb index number 500.
Pupil-teacher ratios is highest around the sub-urb index number 350 - 400.



e) 35 of the suburbs in this data set bound the Charles river

```
> as.data.frame(table(Boston$chas))
  Var1 Freq
1    0  471
2    1   35
>
```

f) The median pupil-teacher ratio among the towns in this data set is 19.05

g) 399th suburb of Boston has lowest median value of owner occupied homes.

```
> which.min(Boston$medv)
[1] 399
> Boston[399,]
       crim zn indus chas   nox    rm age    dis rad tax ptratio black lstat
399 38.3518  0  18.1    0 0.693 5.453 100 1.4896  24 666    20.2 396.9 30.59
    medv
399    5
>
```

Above are the other values for that sub-urb.

Also we can comment about this suburb that, here proportion of residential land zoned for lots over 25,000 sq.ft is zero.

Here no tract bounds the river.

median value of owner-occupied homes in \$1000s is minimum

h) So, the suburbs total 7 room per dwelling is 51.
So, the suburbs total 7 room per dwelling is 13.

```
> special=rep ("No",nrow(Boston))
> special[Boston$rm >7]=" moreseven"
> special[Boston$rm >8]=" moreeight"
> special=as.factor (special)
> Boston=data.frame(Boston ,special)
> summary(Boston)
      crim                zn              indus            chas              nox               rm             age              dis              rad              tax            ptratio     
 Min.   : 0.00632   Min.   :  0.00   Min.   : 0.46   Min.   :0.00000   Min.   :0.3850   Min.   :3.561   Min.   :  2.90   Min.   : 1.130   Min.   : 1.000   Min.   :187.0   Min.   :12.60  
 1st Qu.: 0.08204   1st Qu.:  0.00   1st Qu.: 5.19   1st Qu.:0.00000   1st Qu.:0.4490   1st Qu.:5.886   1st Qu.: 45.02   1st Qu.: 2.100   1st Qu.: 4.000   1st Qu.:279.0   1st Qu.:17.40  
 Median : 0.25651   Median :  0.00   Median : 9.69   Median :0.00000   Median :0.5380   Median :6.208   Median : 77.50   Median : 3.207   Median : 5.000   Median :330.0   Median :19.05  
 Mean   : 3.61352   Mean   : 11.36   Mean   :11.14   Mean   :0.06917   Mean   :0.5547   Mean   :6.285   Mean   : 68.57   Mean   : 3.795   Mean   : 9.549   Mean   :408.2   Mean   :18.46  
 3rd Qu.: 3.67708   3rd Qu.: 12.50   3rd Qu.:18.10   3rd Qu.:0.00000   3rd Qu.:0.6240   3rd Qu.:6.623   3rd Qu.: 94.08   3rd Qu.: 5.188   3rd Qu.:24.000   3rd Qu.:666.0   3rd Qu.:20.20  
 Max.   :88.97620   Max.   :100.00   Max.   :27.74   Max.   :1.00000   Max.   :0.8710   Max.   :8.780   Max.   :100.00   Max.   :12.127   Max.   :24.000   Max.   :711.0   Max.   :22.00  
     black            lstat            medv            special     
 Min.   :  0.32   Min.   : 1.73   Min.   : 5.00   moreeight: 13  
 1st Qu.:375.38   1st Qu.: 6.95   1st Qu.:17.02   moreseven: 51  
 Median :391.44   Median :11.36   Median :21.20   No       :442  
 Mean   :356.67   Mean   :12.65   Mean   :22.53                  
 3rd Qu.:396.23   3rd Qu.:16.95   3rd Qu.:25.00                  
 Max.   :396.90   Max.   :37.97   Max.   :50.00                  
```