Introduction to Statistical Learning- Lab1

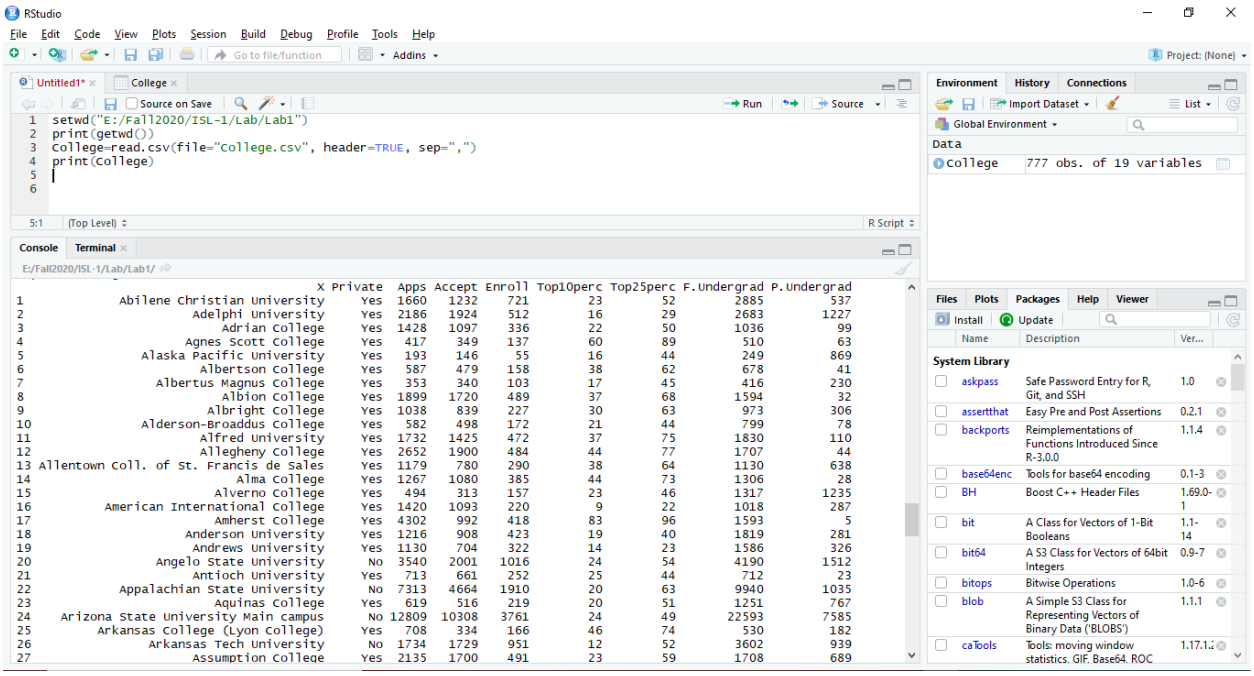
Name: Sandeep Reddy Salkuti

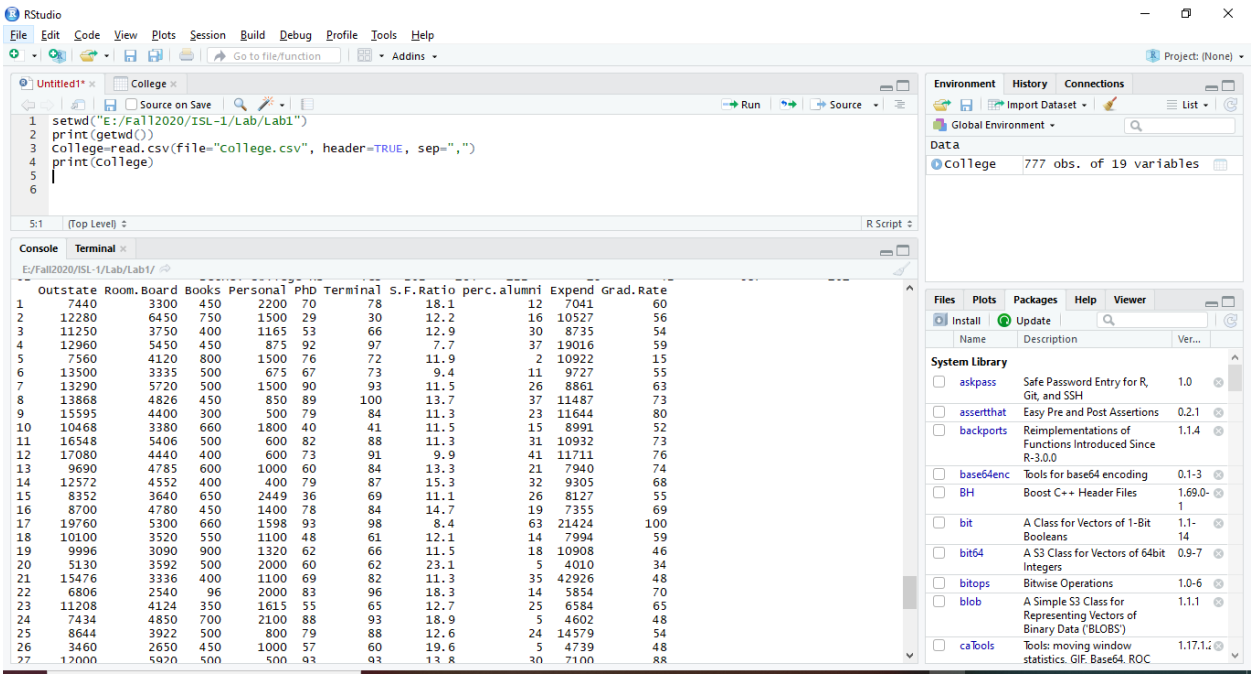
Student id: 16296868

Email: sswf7@umsystem.edu

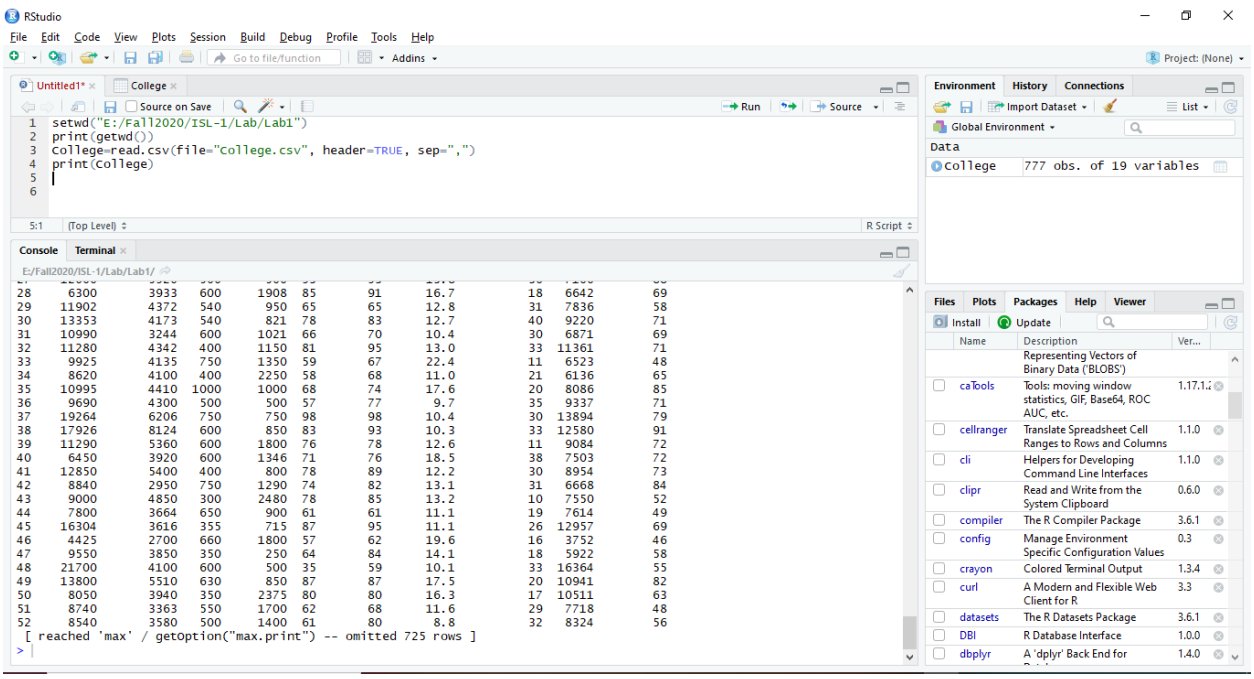
1)

a) Use the read.csv() function to read the data into R. Call the loaded data college. Make sure that you have the directory set to the correct location for the data.





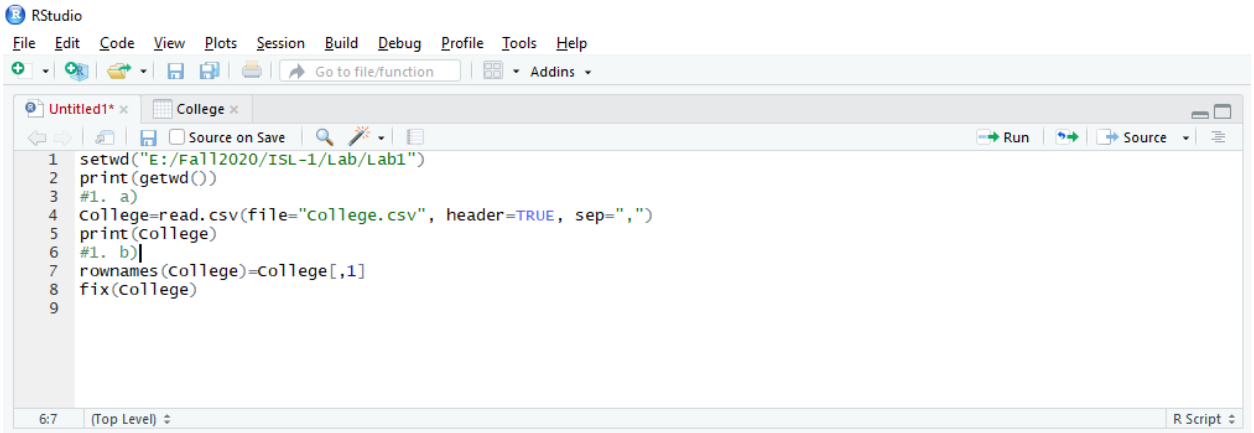

```
   Outstate Room.Board Books Personal Phd Terminal S.F.Ratio perc.alumni Expend Grad.Rate
1      7440       3300   450     2200  70       78      18.1          12   7041        60
2     12280       6450   750     1500  29       30      12.2          16  10527        56
3     11250       3750   400     1165  53       66      12.9          30   8735        54
4     12960       5450   450      875  92       97       7.7          37  19016        59
5      7560       4120   800     1500  76       72      11.9           2  10922        15
6     13500       3335   500      675  67       73       9.4          11   9727        55
7     13290       5720   500     1500  90       93      11.5          26   8861        63
8     13868       4826   450      850  89      100      13.7          37  11487        73
9     15595       4400   300      500  79       84      11.3          23  11644        80
10    10468       3380   660     1800  40       41      11.5          15   8991        52
11    16548       5406   500      600  82       88      11.3          31  10932        73
12    17080       4440   400      600  73       91       9.9          41  11711        76
13     9690       4785   600     1000  60       84      13.3          21   7940        74
14    12572       4552   400      400  79       87      15.3          32   9305        68
15     8352       3640   650     2449  36       69      11.1          26   8127        55
16     8700       4780   450     1400  78       84      14.7          19   7355        69
17    19760       5300   660     1598  93       98       8.4          63  21424       100
18    10100       3520   550     1100  48       61      12.1          14   7994        59
19     9996       3090   900     1320  62       66      11.5          18  10908        46
20     5130       3592   500     2000  60       62      23.1           5   4010        34
21    15476       3336   400     1100  69       82      11.3          35  42926        48
22     6806       2540    96     2000  83       96      18.3          14   5854        70
23    11208       4124   350     1615  55       65      12.7          25   6584        65
24     7434       4850   700     2100  88       93      18.9           5   4602        48
25     8644       3922   500      800  79       88      12.6          24  14579        54
26     3460       2650   450     1000  57       60      19.6           5   4739        48
27    12000       5920   500      500  93       93      13.8          30   7100        88
```

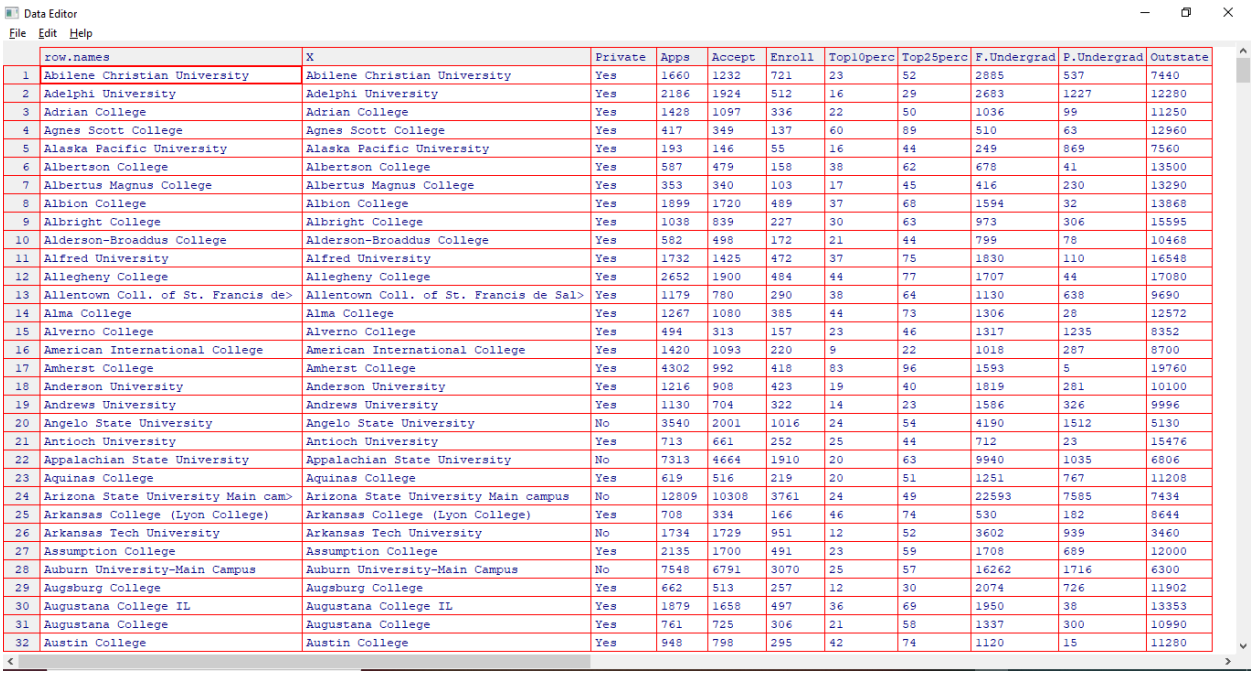



```
28     6300       3933   600     1908  85       91      16.7          18   6642        69
29    11902       4372   540      950  65       65      12.8          31   7836        58
30    13353       4173   540      821  78       83      12.7          40   9220        71
31    10990       3244   600     1021  66       70      10.4          30   6871        69
32    11280       4342   400     1150  81       95      13.0          33  11361        71
33     9925       4135   750     1350  59       67      22.4          11   6523        48
34     8620       4100   400     2250  58       68      11.0          21   6136        65
35    10995       4410  1000     1000  68       74      17.6          20   8086        85
36     9690       4300   500      500  57       77       9.7          35   9337        71
37    19264       6206   750      750  98       98      10.4          30  13894        79
38    17926       8124   600      850  83       93      10.3          33  12580        91
39    11290       5360   600     1800  76       78      12.6          11   9084        72
40     6450       3920   600     1346  71       76      18.5          38   7503        72
41    12850       5400   400      800  78       89      12.2          30   8954        73
42     8840       2950   750     1290  74       82      13.1          31   6668        84
43     9000       4850   300     2480  78       85      13.2          10   7550        52
44     7800       3664   650      900  61       61      11.1          19   7614        49
45    16304       3616   355      715  87       95      11.1          26  12957        69
46     4425       2700   660     1800  57       62      19.6          16   3752        46
47     9550       3850   350      250  64       84      14.1          18   5922        58
48    21700       4100   600      500  35       59      10.1          33  16364        55
49    13800       5510   630      850  87       87      17.5          20  10941        82
50     8050       3940   350     2375  80       80      16.3          17  10511        63
51     8740       3363   550     1700  62       68      11.6          29   7718        48
52     8540       3580   500     1400  61       80       8.8          32   8324        56
 [ reached 'max' / getOption("max.print") -- omitted 725 rows ]
>
```

b) Look at the data using the fix() function. You should notice that the first column is just the name of each university. We don't really want R to treat this as data. However, it may be handy to have these names for later. Try the following commands: > rownames (college)=college [,1] > fix (college) You should see that there is now a row.names column with the name of each university recorded. This means that R has given each row a name corresponding to the appropriate university. R will not try to perform calculations on the row names. However, we still need to eliminate the first column in the data where the names are stored. Try > college =college [,-1] > fix (college) Now you should see that

the first data column is Private. Note that another column labeled row.names now appears before the Private column. However, this is not a data column but rather the name that R is giving to each row.





	row.names	X	Private	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate
1	Abilene Christian University	Abilene Christian University	Yes	1660	1232	721	23	52	2885	537	7440
2	Adelphi University	Adelphi University	Yes	2186	1924	512	16	29	2683	1227	12280
3	Adrian College	Adrian College	Yes	1428	1097	336	22	50	1036	99	11250
4	Agnes Scott College	Agnes Scott College	Yes	417	349	137	60	89	510	63	12960
5	Alaska Pacific University	Alaska Pacific University	Yes	193	146	55	16	44	249	869	7560
6	Albertson College	Albertson College	Yes	587	479	158	38	62	678	41	13500
7	Albertus Magnus College	Albertus Magnus College	Yes	353	340	103	17	45	416	230	13290
8	Albion College	Albion College	Yes	1899	1720	489	37	68	1594	32	13868
9	Albright College	Albright College	Yes	1038	839	227	30	63	973	306	15595
10	Alderson-Broaddus College	Alderson-Broaddus College	Yes	582	498	172	21	44	799	78	10468
11	Alfred University	Alfred University	Yes	1732	1425	472	37	75	1830	110	16548
12	Allegheny College	Allegheny College	Yes	2652	1900	484	44	77	1707	44	17080
13	Allentown Coll. of St. Francis de>	Allentown Coll. of St. Francis de Sal>	Yes	1179	780	290	38	64	1130	638	9690
14	Alma College	Alma College	Yes	1267	1080	385	44	73	1306	28	12572
15	Alverno College	Alverno College	Yes	494	313	157	23	46	1317	1235	8352
16	American International College	American International College	Yes	1420	1093	220	9	22	1018	287	8700
17	Amherst College	Amherst College	Yes	4302	992	418	83	96	1593	5	19760
18	Anderson University	Anderson University	Yes	1216	908	423	19	40	1819	281	10100
19	Andrews University	Andrews University	Yes	1130	704	322	14	23	1586	326	9996
20	Angelo State University	Angelo State University	No	3540	2001	1016	24	54	4190	1512	5130
21	Antioch University	Antioch University	Yes	713	661	252	25	44	712	23	15476
22	Appalachian State University	Appalachian State University	No	7313	4664	1910	20	63	9940	1035	6806
23	Aquinas College	Aquinas College	Yes	619	516	219	20	51	1251	767	11208
24	Arizona State University Main cam>	Arizona State University Main campus	No	12809	10308	3761	24	49	22593	7585	7434
25	Arkansas College (Lyon College)	Arkansas College (Lyon College)	Yes	708	334	166	46	74	530	182	8644
26	Arkansas Tech University	Arkansas Tech University	No	1734	1729	951	12	52	3602	939	3460
27	Assumption College	Assumption College	Yes	2135	1700	491	23	59	1708	689	12000
28	Auburn University-Main Campus	Auburn University-Main Campus	No	7548	6791	3070	25	57	16262	1716	6300
29	Augsburg College	Augsburg College	Yes	662	513	257	12	30	2074	726	11902
30	Augustana College IL	Augustana College IL	Yes	1879	1658	497	36	69	1950	38	13353
31	Augustana College	Augustana College	Yes	761	725	306	21	58	1337	300	10990
32	Austin College	Austin College	Yes	948	798	295	42	74	1120	15	11280

- To eliminate the first column in the data where the names are stored below is the command typed and screenshot attached.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled1*

Source on Save Run Source

```
setwd("E:/Fall2020/ISL-1/Lab/Lab1")
print(getwd())
#1. a)
college=read.csv(file="College.csv", header=TRUE, sep=",")
print(college)
#1. b)
rownames(college)=college[,1]
fix(college)
#Eliminate first column in data
college=college[,-1]
fix(college)
```

10:1 (Top Level) R Script

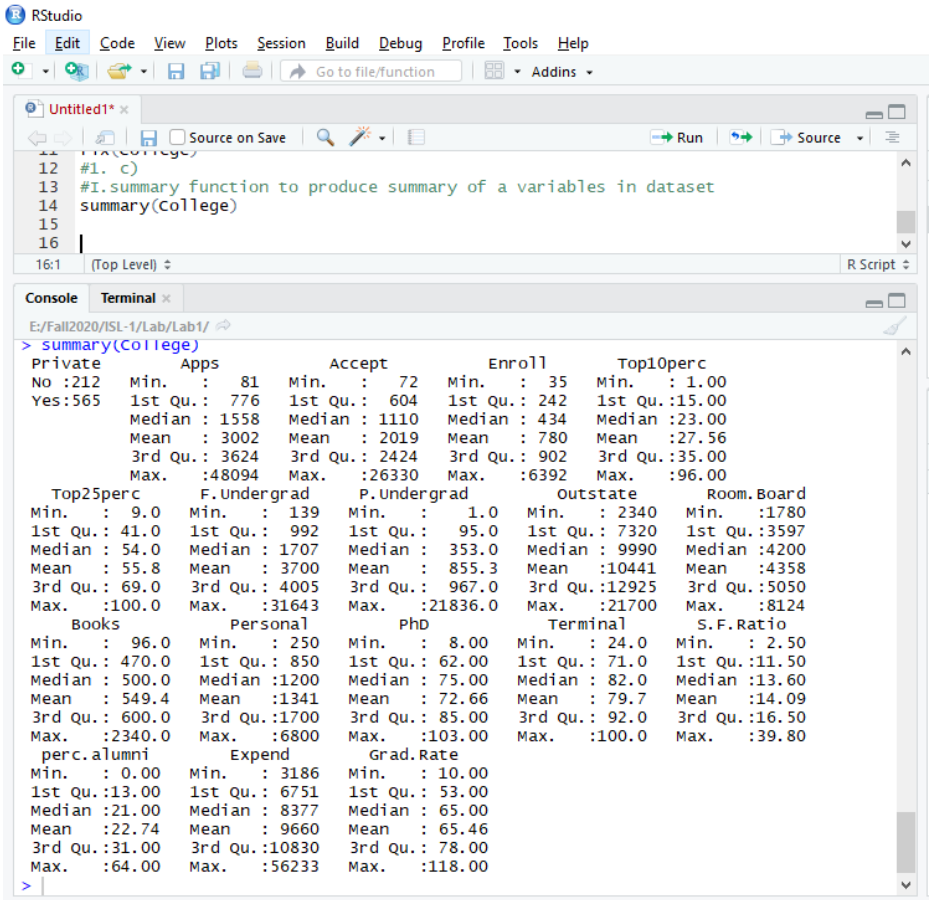
Data Editor

File Edit Help

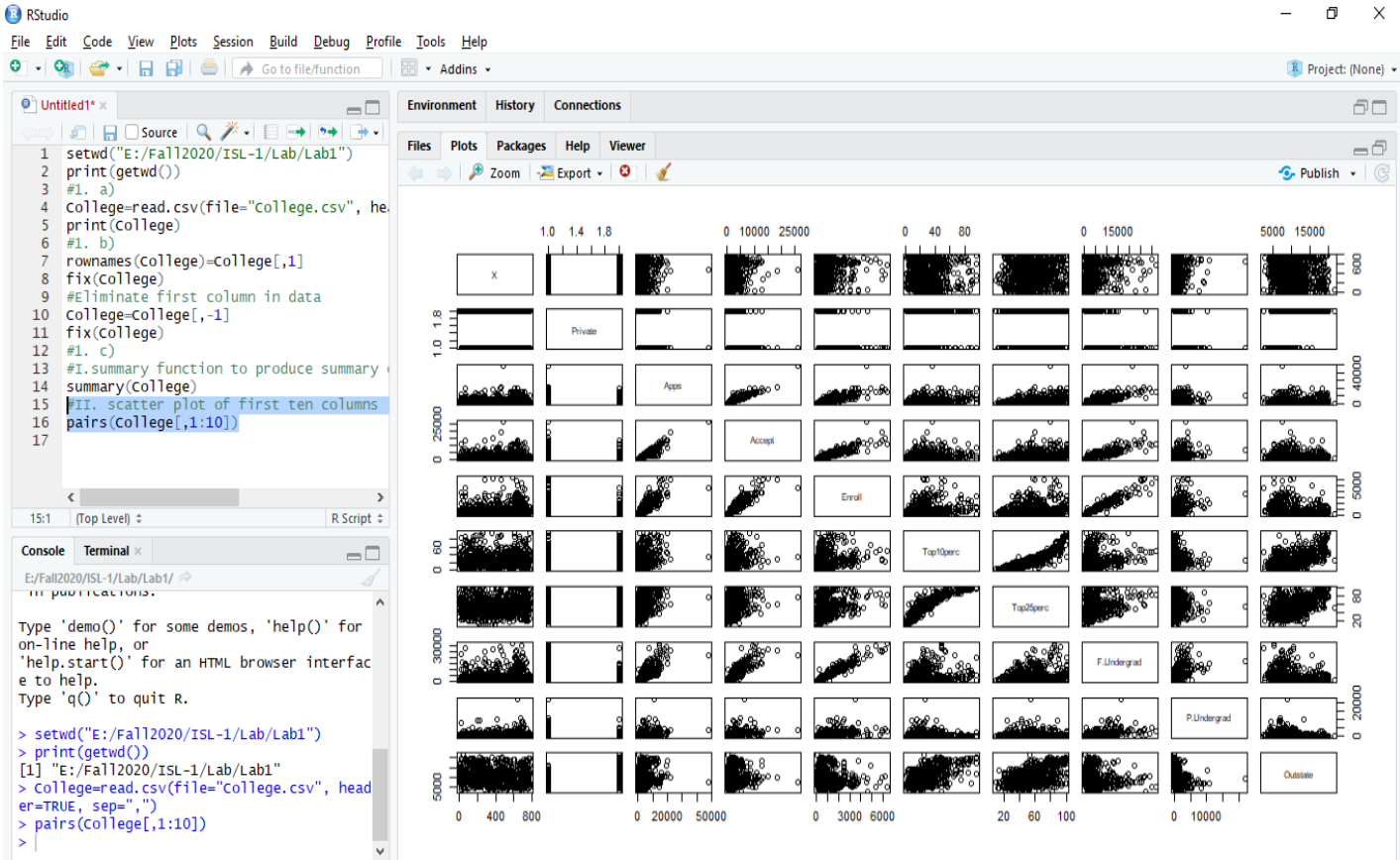
	row.names	Private	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal
1	Abilene Christian University	Yes	1660	1232	721	23	52	2885	537	7440	3300	450	2200
2	Adelphi University	Yes	2186	1924	512	16	29	2683	1227	12280	6450	750	1500
3	Adrian College	Yes	1428	1097	336	22	50	1036	99	11250	3750	400	1165
4	Agnes Scott College	Yes	417	349	137	60	89	510	63	12960	5450	450	875
5	Alaska Pacific University	Yes	193	146	55	16	44	249	869	7560	4120	800	1500
6	Albertson College	Yes	587	479	158	38	62	678	41	13500	3335	500	675
7	Albertus Magnus College	Yes	353	340	103	17	45	416	230	13290	5720	500	1500
8	Albion College	Yes	1899	1720	489	37	68	1594	32	13868	4826	450	850
9	Albright College	Yes	1038	839	227	30	63	973	306	15595	4400	300	500
10	Alderson-Broaddus College	Yes	582	498	172	21	44	799	78	10468	3380	660	1800
11	Alfred University	Yes	1732	1425	472	37	75	1830	110	16548	5406	500	600
12	Allegheny College	Yes	2652	1900	484	44	77	1707	44	17080	4440	400	600
13	Allentown Coll. of St. Francis de Sales	Yes	1179	780	290	38	64	1130	638	9690	4785	600	1000
14	Alma College	Yes	1267	1080	385	44	73	1306	28	12572	4552	400	400
15	Alverno College	Yes	494	313	157	23	46	1317	1235	8352	3640	650	2449
16	American International College	Yes	1420	1093	220	9	22	1018	287	8700	4780	450	1400
17	Amherst College	Yes	4302	992	418	83	96	1593	5	19760	5300	660	1598
18	Anderson University	Yes	1216	908	423	19	40	1819	281	10100	3520	550	1100
19	Andrews University	Yes	1130	704	322	14	23	1586	326	9996	3090	900	1320
20	Angelo State University	No	3540	2001	1016	24	54	4190	1512	5130	3592	500	2000
21	Antioch University	Yes	713	661	252	25	44	712	23	15476	3336	400	1100
22	Appalachian State University	No	7313	4664	1910	20	63	9940	1035	6806	2540	96	2000
23	Aquinas College	Yes	619	516	219	20	51	1251	767	11208	4124	350	1615
24	Arizona State University Main campus	No	12809	10308	3761	24	49	22593	7585	7434	4850	700	2100
25	Arkansas College (Lyon College)	Yes	708	334	166	46	74	530	182	8644	3922	500	800
26	Arkansas Tech University	No	1734	1729	951	12	52	3602	939	3460	2650	450	1000
27	Assumption College	Yes	2135	1700	491	23	59	1708	689	12000	5920	500	500
28	Auburn University-Main Campus	No	7548	6791	3070	25	57	16262	1716	6300	3933	600	1908
29	Augsburg College	Yes	662	513	257	12	30	2074	726	11902	4372	540	950
30	Augustana College IL	Yes	1879	1658	497	36	69	1950	38	13353	4173	540	821
31	Augustana College	Yes	761	725	306	21	58	1337	300	10990	3244	600	1021
32	Austin College	Yes	948	798	295	42	74	1120	15	11280	4342	400	1150

c)

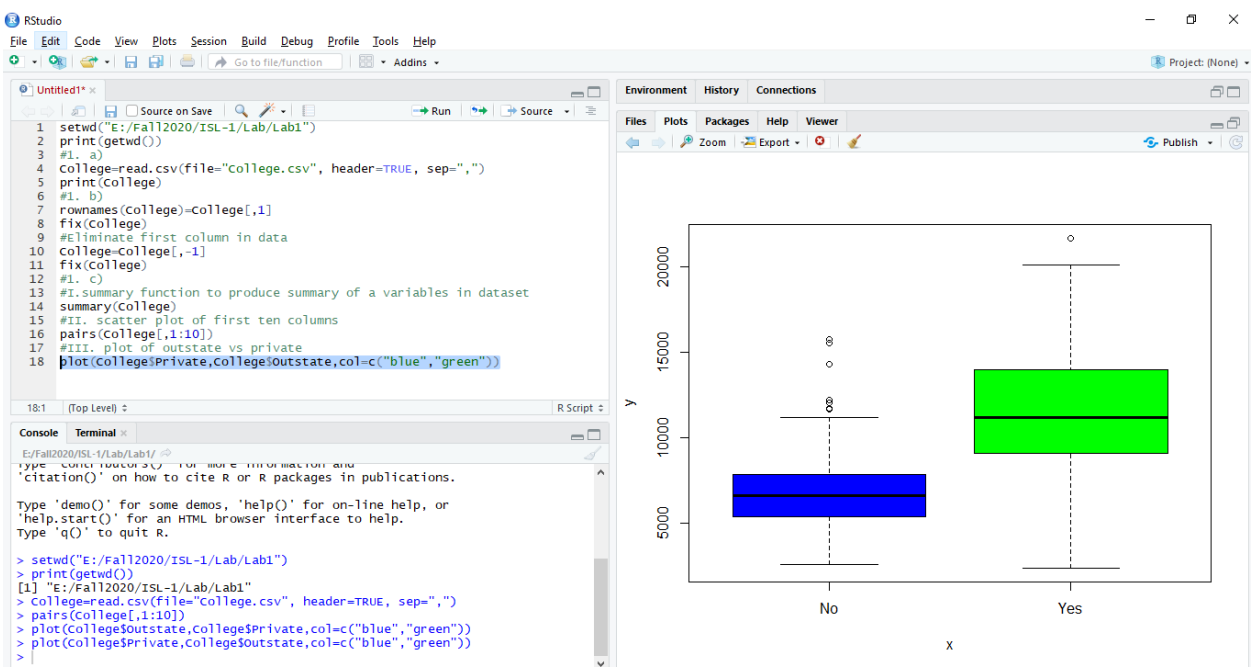
I) Use the summary() function to produce a numerical summary of the variables in the data set



II) Use the pairs() function to produce a scatterplot matrix of the first ten columns or variables of the data. Recall that you can reference the first ten columns of a matrix A using A[,1:10].

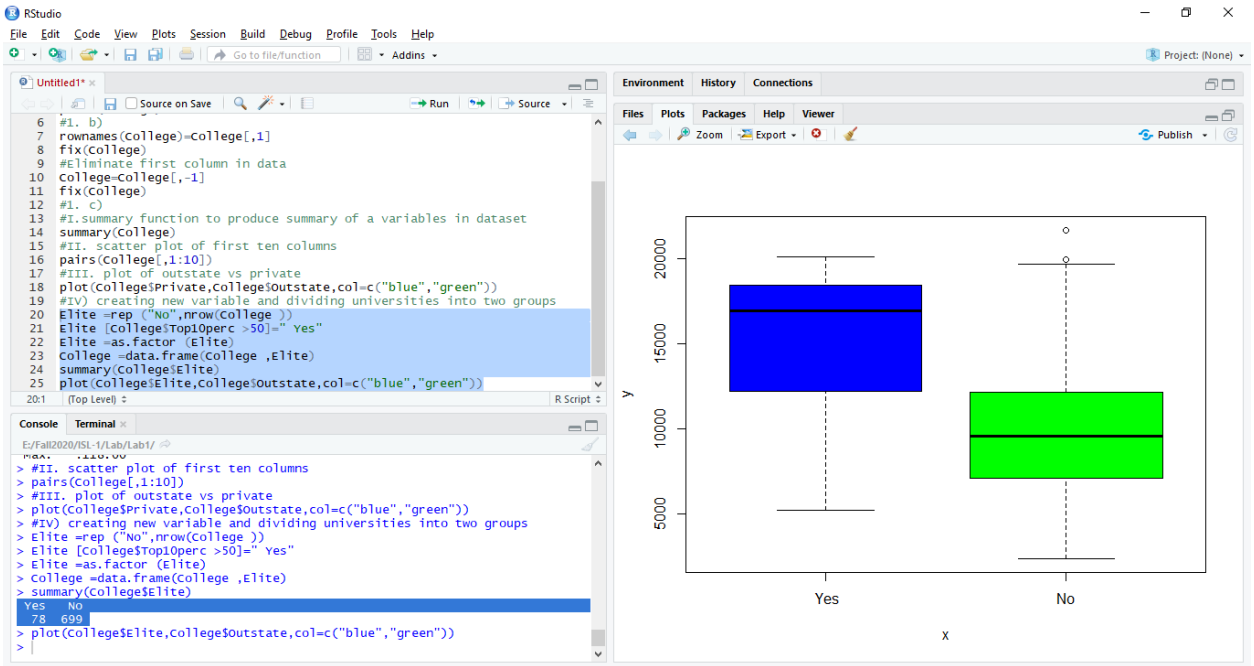


III) Use the plot() function to produce side-by-side boxplots of Outstate versus Private.

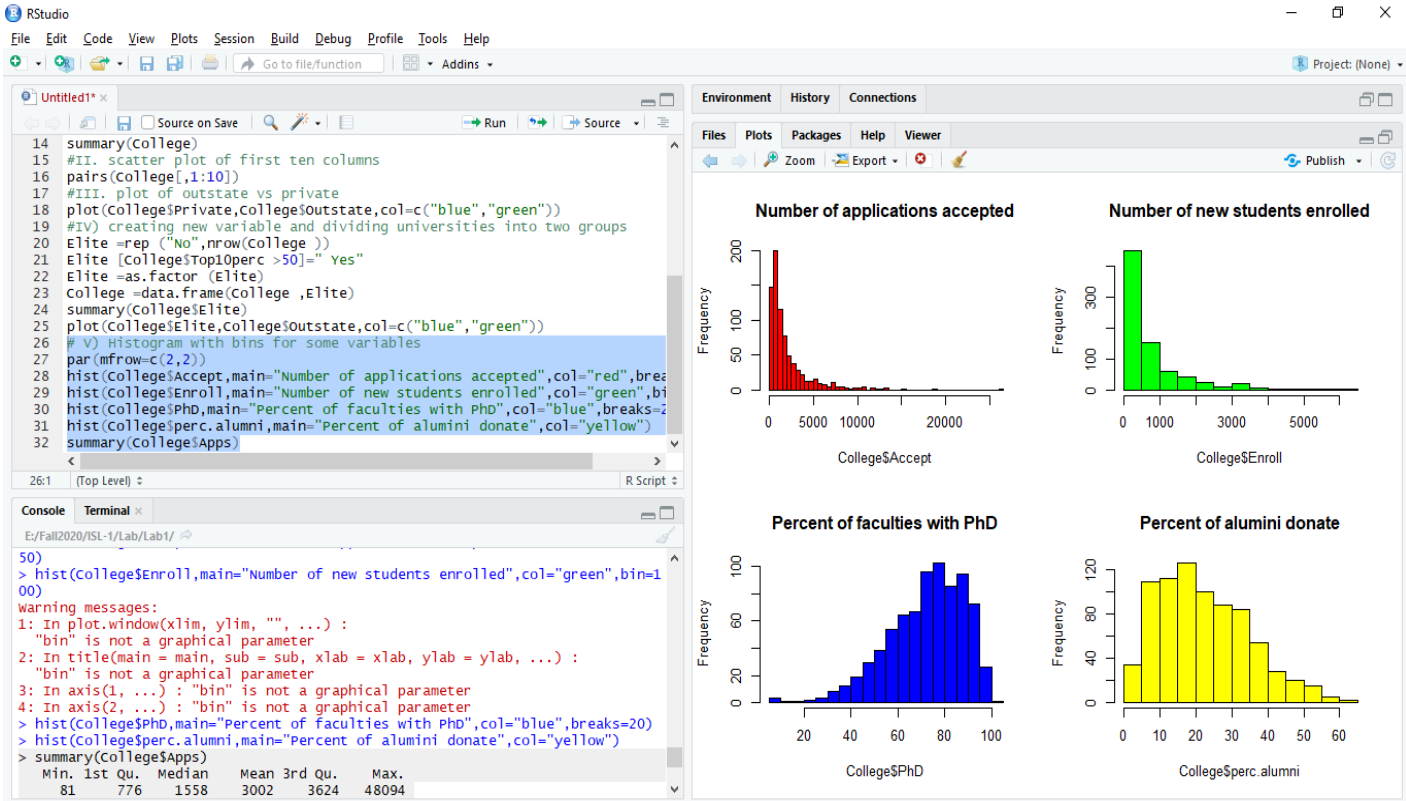


IV) Create a new qualitative variable, called Elite, by binning the Top10perc variable. We are going to divide universities into two groups based on whether or not the proportion of students coming from

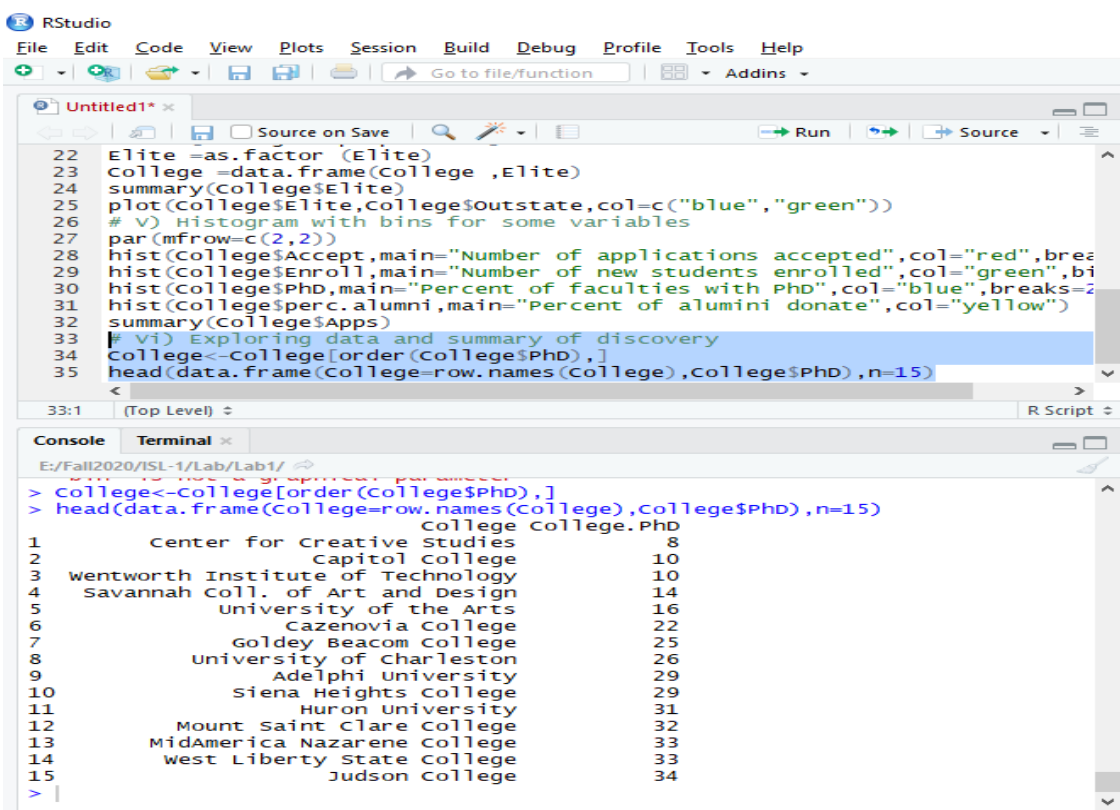
the top 10% of their high school classes exceeds 50%. > Elite =rep ("No",nrow(college)) > Elite [college$Top10perc >50]=" Yes" > Elite =as.factor (Elite) > college =data.frame(college ,Elite) Use the summary() function to see how many elite universities there are. Now use the plot() function to produce side-by-side boxplots of Outstate versus Elite. v. Use the hist() function to produce some histograms with differing



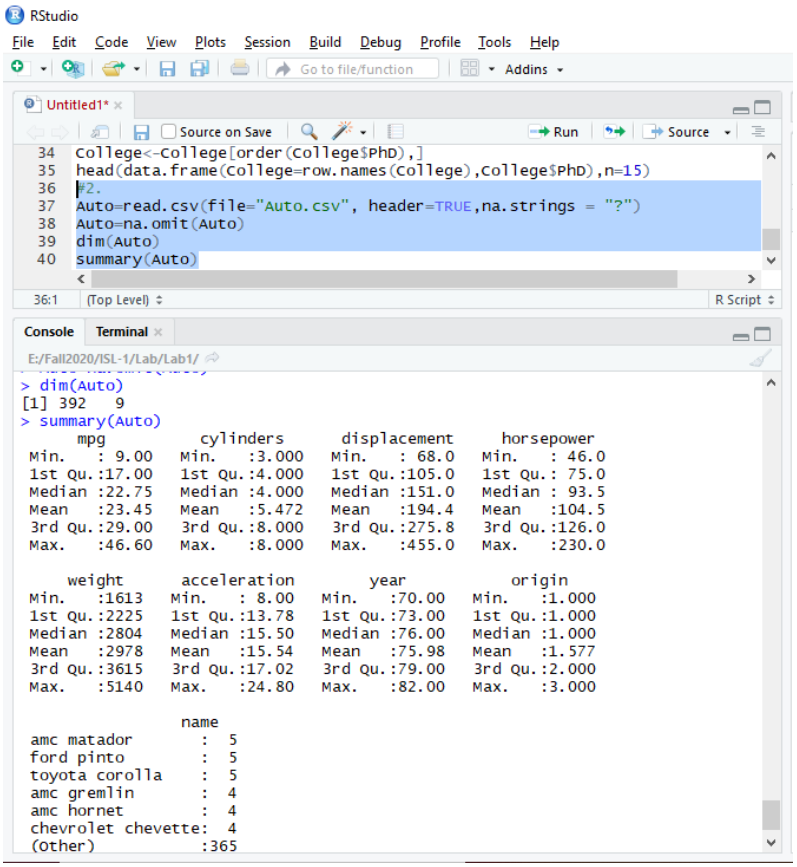
V) Use the hist() function to produce some histograms with differing numbers of bins for a few of the quantitative variables. You may find the command par(mfrow=c(2,2)) useful: it will divide the print window into four regions so that four plots can be made simultaneously. Modifying the arguments to this function will divide the screen in other ways.



VI) Continue exploring the data, and provide a brief summary of what you discover



2)



a) Which of the predictors are quantitative, and which are qualitative?

quantitative variables: mpg, cylinders, displacement, horsepower, weight, acceleration

qualitative variables: year, origin, name

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Untitled1*

Source on Save | Run | Source

```
summary(Auto)
#a.
range(Auto$mpg)
range(Auto$weight)
summary(Auto[,c(7:9),])
summary(Auto[,c(1:6),])
```

41:1 (Top Level) R Script

Console Terminal

E:/Fall2020/ISL-1/Lab/Lab1/

```
> range(Auto$mpg)
[1]  9.0 46.6
> range(Auto$weight)
[1] 1613 5140
> summary(Auto[,c(7:9),])
      year           origin                      name
 Min.   :70.00   Min.   :1.000   amc matador       :  5
 1st Qu.:73.00   1st Qu.:1.000   ford pinto        :  5
 Median :76.00   Median :1.000   toyota corolla    :  5
 Mean   :75.98   Mean   :1.577   amc gremlin       :  4
 3rd Qu.:79.00   3rd Qu.:2.000   amc hornet        :  4
 Max.   :82.00   Max.   :3.000   chevrolet chevette:  4
                                 (other)           :365
> summary(Auto[,c(1:6),])
      mpg          cylinders      displacement     horsepower
 Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0
 1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0
 Median :22.75   Median :4.000   Median :151.0   Median : 93.5
 Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5
 3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0
 Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0
     weight      acceleration
 Min.   :1613   Min.   : 8.00
 1st Qu.:2225   1st Qu.:13.78
 Median :2804   Median :15.50
 Mean   :2978   Mean   :15.54
 3rd Qu.:3615   3rd Qu.:17.02
 Max.   :5140   Max.   :24.80
```

b) What is the range of each quantitative predictor? You can answer this using the range() function.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled1*

Source on Save Run Source

```
#2.
Auto=read.csv(file="Auto.csv", header=TRUE,na.strings = "?")
Auto=na.omit(Auto)
dim(Auto)
summary(Auto)
#a.qualitative and quantitative predictors
range(Auto$mpg)
range(Auto$weight)
summary(Auto[,c(7:9),])
summary(Auto[,c(1:6),])
#b.range of all quantitative predictor
range(Auto$mpg)
range(Auto$cylinders)
range(Auto$displacement)
range(Auto$horsepower)
range(Auto$weight)
range(Auto$acceleration)
```

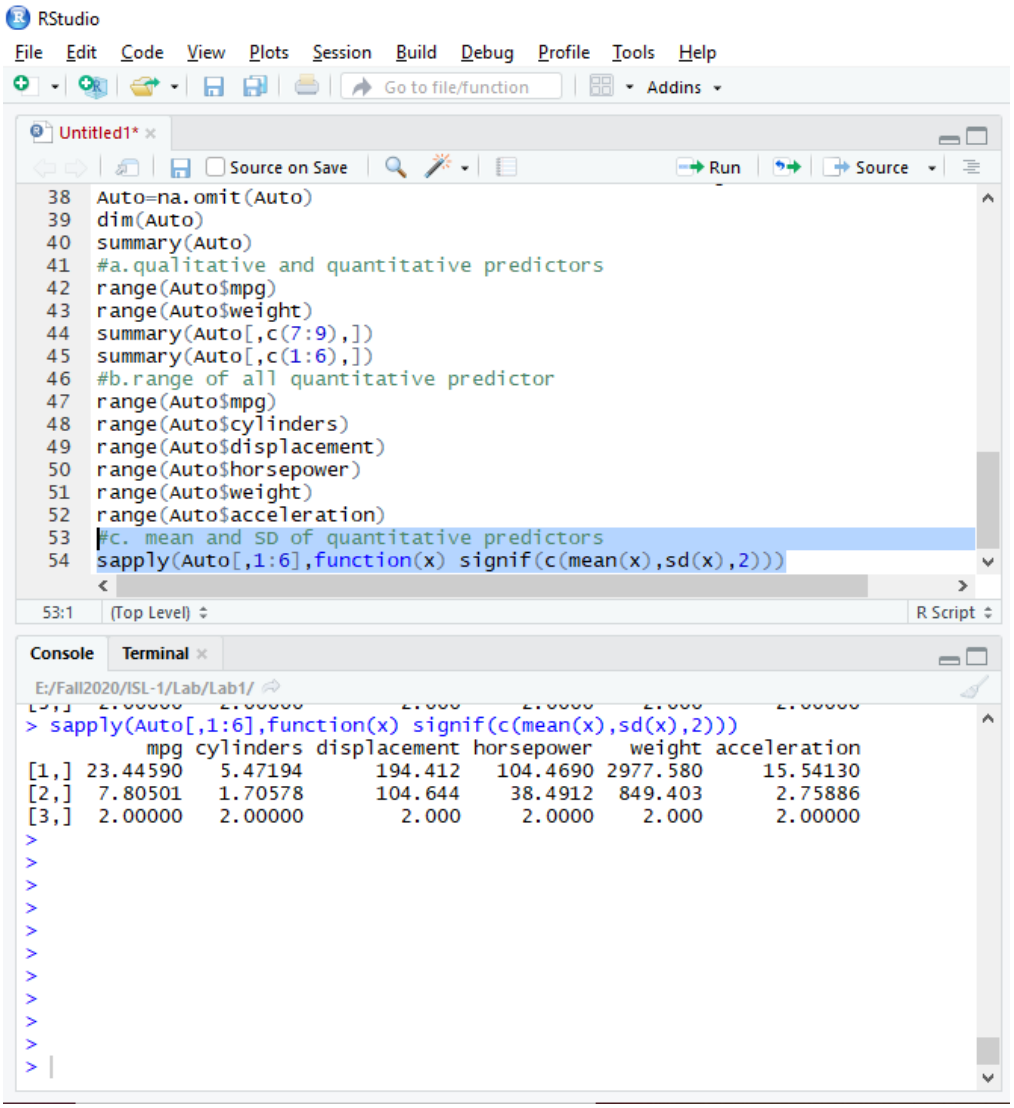
46:1 (Top Level) R Script

Console Terminal

E:/Fall2020/ISL-1/Lab/Lab1/

```
 Mean   :2978   Mean   :15.54
 3rd Qu.:3615   3rd Qu.:17.02
 Max.   :5140   Max.   :24.80
> range(Auto$mpg)
[1]  9.0 46.6
> range(Auto$cylinders)
[1] 3 8
> range(Auto$displacement)
[1]  68 455
> range(Auto$horsepower)
[1]  46 230
> range(Auto$weight)
[1] 1613 5140
> range(Auto$acceleration)
[1]  8.0 24.8
>
```

c) What is the mean and standard deviation of each quantitative predictor?



d) Now remove the 10th through 85th observations. What is the range, mean, and standard deviation of each predictor in the subset of the data that remains?

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Untitled1*

```r
range(Auto$weight)
summary(Auto[,c(7:9),])
summary(Auto[,c(1:6),])
#b.range of all quantitative predictor
range(Auto$mpg)
range(Auto$cylinders)
range(Auto$displacement)
range(Auto$horsepower)
range(Auto$weight)
range(Auto$acceleration)
#c. mean and SD of quantitative predictors
sapply(Auto[,1:6],function(x) signif(c(mean(x),sd(x),2)))
#d. removing 10-85th observations and range,mean,sd of remaining predicto
new.auto=subset(Auto[-c(10:85),])
sapply(new.auto[,-c(9)],range)
sapply(new.auto[,-c(9)],mean)
sapply(new.auto[,-c(9)],sd)
```

55:1 (Top Level) R Script

Console Terminal

E:/Fall2020/ISL-1/Lab/Lab1/

```
> new.auto=subset(Auto[-c(10:85),])
> sapply(new.auto[,-c(9)],range)
      mpg cylinders displacement horsepower weight acceleration year origin
[1,] 11.0         3           68         46   1649          8.5   70      1
[2,] 46.6         8          455        230   4997         24.8   82      3
> sapply(new.auto[,-c(9)],mean)
         mpg    cylinders displacement   horsepower       weight
   24.404430     5.373418   187.240506   100.721519  2935.971519
acceleration         year       origin
   15.726899    77.145570     1.601266
> sapply(new.auto[,-c(9)],sd)
         mpg    cylinders displacement   horsepower       weight
    7.867283     1.654179    99.678367    35.708853   811.300208
acceleration         year       origin
    2.693721     3.106217     0.819910
>
```

e) Using the full data set, investigate the predictors graphically, using scatterplots or other tools of your choice. Create some plots highlighting the relationships among the predictors. Comment on your findings.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

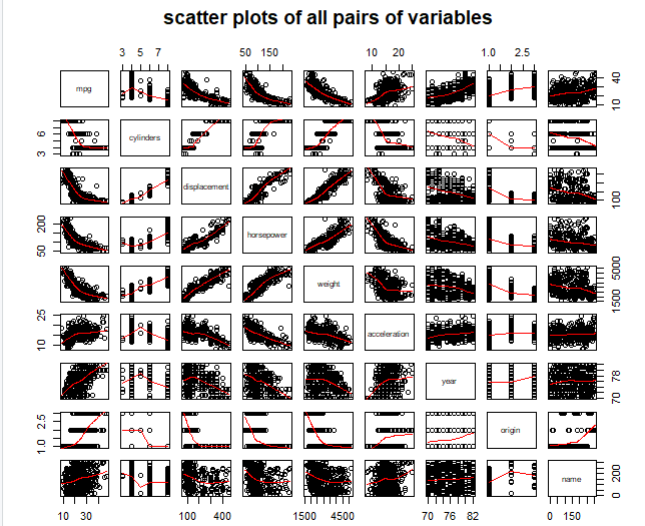
Untitled1*

```r
summary(Auto[,c(1:6),])
#b.range of all quantitative predictor
range(Auto$mpg)
range(Auto$cylinders)
range(Auto$displacement)
range(Auto$horsepower)
range(Auto$weight)
range(Auto$acceleration)
#c. mean and SD of quantitative predictors
sapply(Auto[,1:6],function(x) signif(c(mean(x),sd(x),2)))
#d. removing 10-85th observations and range,mean,sd of remaining predictors
new.auto=subset(Auto[-c(10:85),])
sapply(new.auto[,-c(9)],range)
sapply(new.auto[,-c(9)],mean)
sapply(new.auto[,-c(9)],sd)
#e. Investigating predictors graphically
pairs(Auto,panel=panel.smooth,main="scatter plots of all pairs of variables")
```

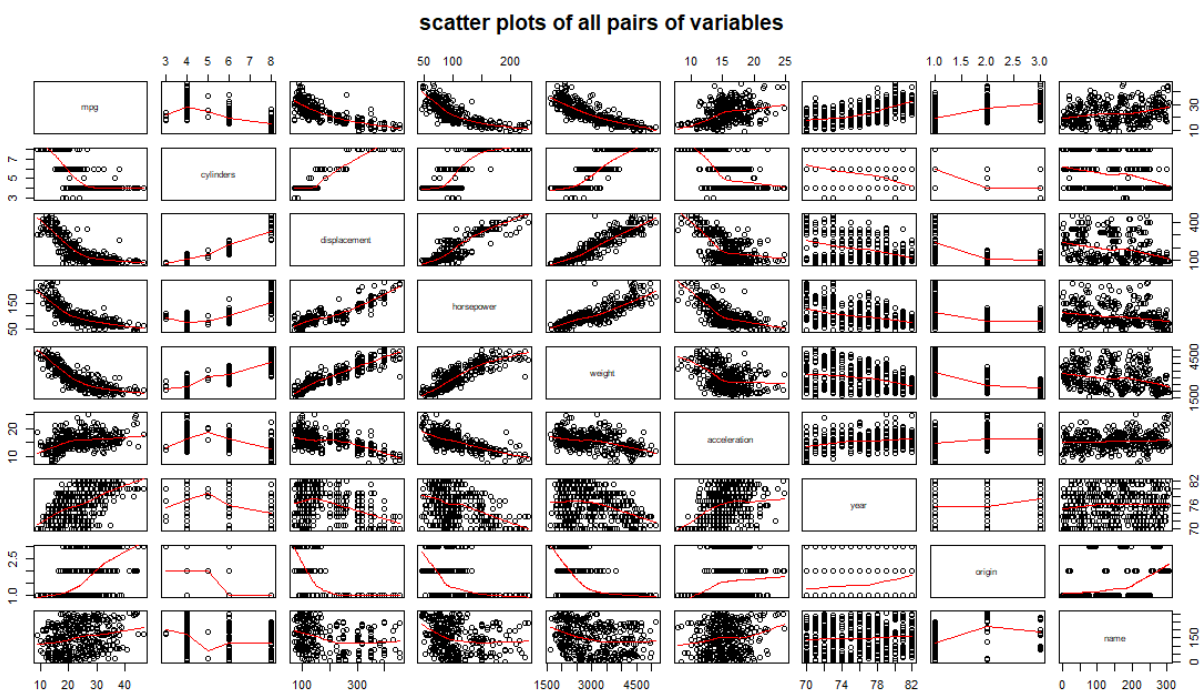
Console Terminal

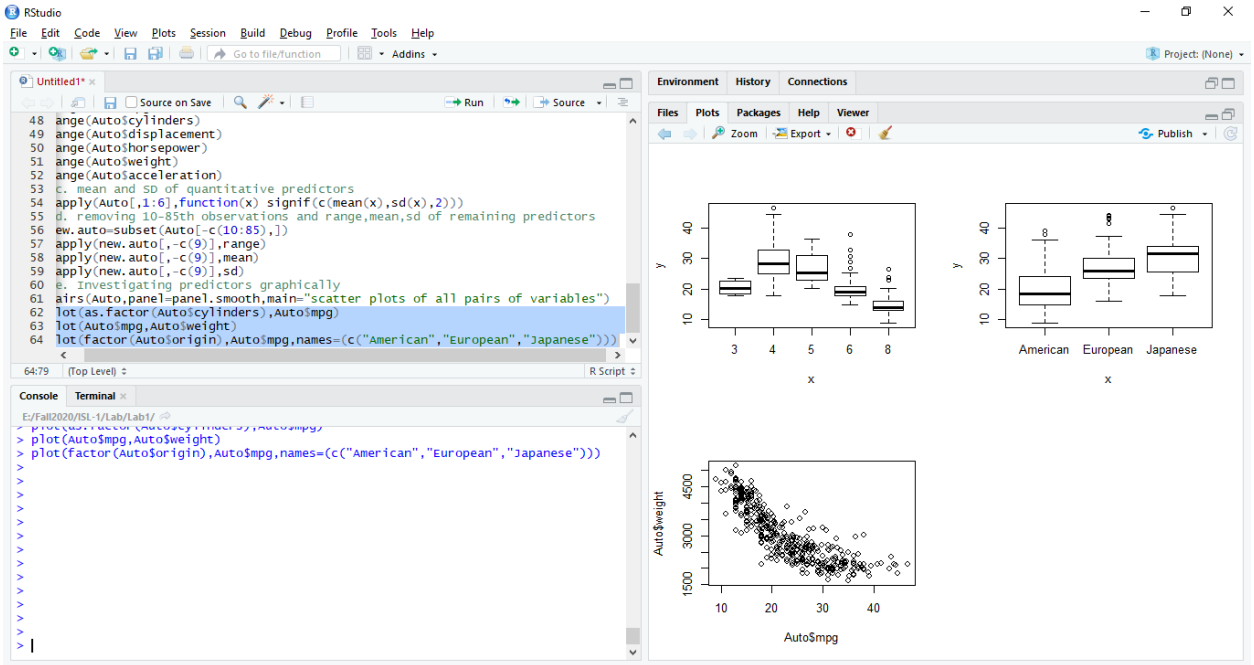
E:/Fall2020/ISL-1/Lab/Lab1/

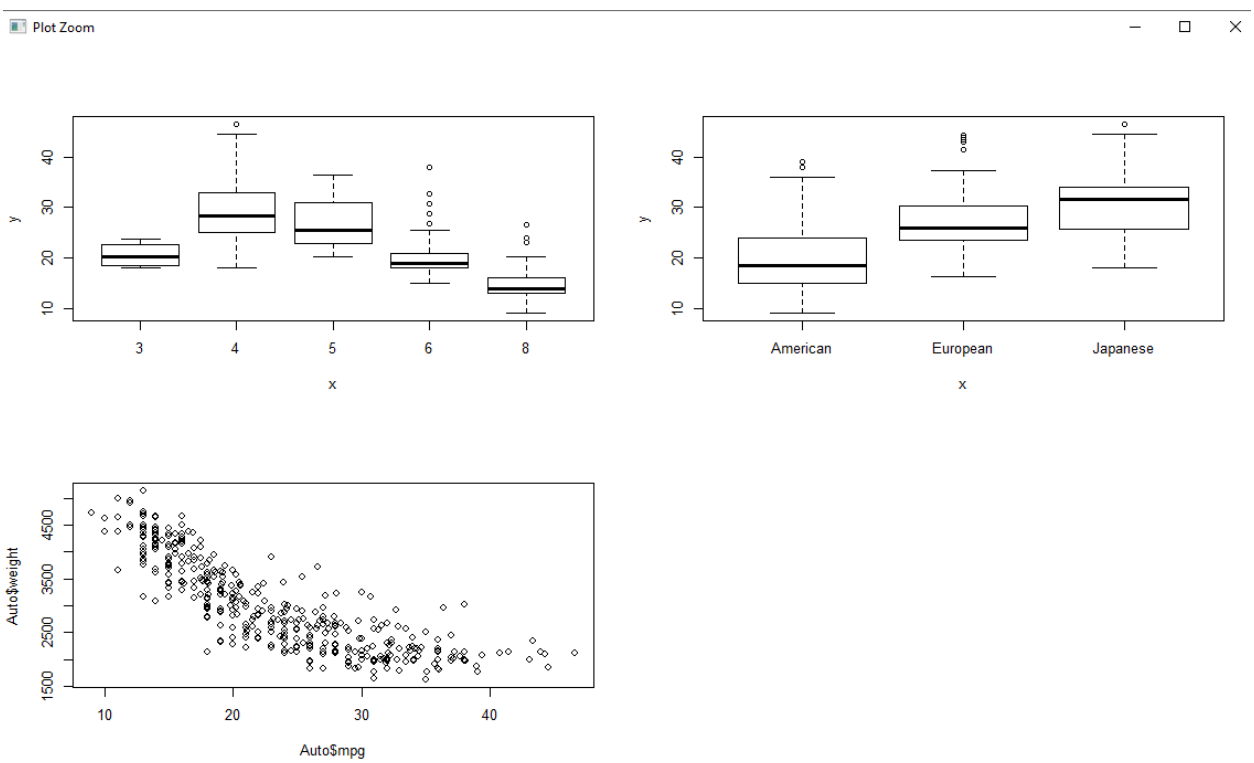
```
> sapply(new.auto[,-c(9)],range)
     mpg cylinders displacement horsepower weight acceleration year origin
[1,] 11.0         3           68         46   1649          8.5   70      1
[2,] 46.6         8          455        230   4997         24.8   82      3
> sapply(new.auto[,-c(9)],mean)
         mpg    cylinders displacement   horsepower       weight
   24.404430     5.373418   187.240506   100.721519  2935.971519
acceleration         year       origin
   15.726899    77.145570     1.601266
> sapply(new.auto[,-c(9)],sd)
         mpg    cylinders displacement   horsepower       weight
    7.867283     1.654179    99.678367    35.708853   811.300208
acceleration         year       origin
    2.693721     3.106217     0.819910
> pairs(Auto,panel=panel.smooth,main="scatter plots of all pairs of variables")
>
```



Plot Zoom







By looking at above graphs Weight, displacement and horse power seems to have an inverse effect with mpg. While displacement with horse power are directly proportional.

f) Suppose that we wish to predict gas mileage (mpg) on the basis of the other variables. Do your plots suggest that any of the other variables might be useful in predicting mpg? Justify your answer.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

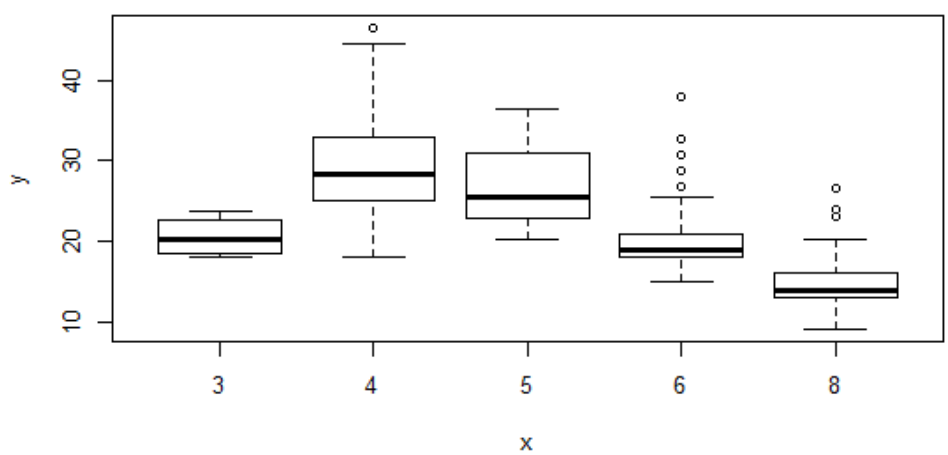
Go to file/function Addins

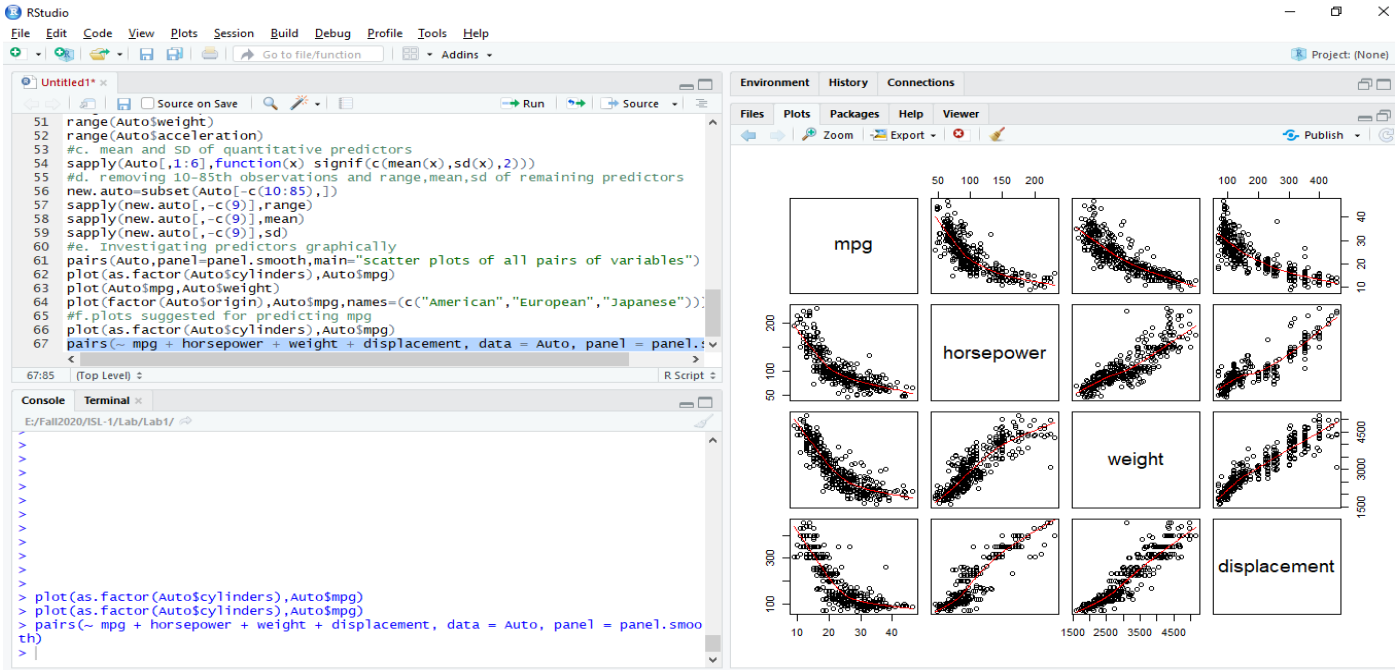
Untitled1*

Source on Save Run Source

```
range(Auto$weight)
range(Auto$acceleration)
#c. mean and SD of quantitative predictors
sapply(Auto[,1:6],function(x) signif(c(mean(x),sd(x),2)))
#d. removing 10-85th observations and range,mean,sd of remaining predictors
new.auto=subset(Auto[-c(10:85),])
sapply(new.auto[,-c(9)],range)
sapply(new.auto[,-c(9)],mean)
sapply(new.auto[,-c(9)],sd)
#e. Investigating predictors graphically
pairs(Auto,panel=panel.smooth,main="scatter plots of all pairs of variables")
plot(as.factor(Auto$cylinders),Auto$mpg)
plot(Auto$mpg,Auto$weight)
plot(factor(Auto$origin),Auto$mpg,names=(c("American","European","Japanese")))
#f.plots suggested for predicting mpg
plot(as.factor(Auto$cylinders),Auto$mpg)
```

Plot Zoom

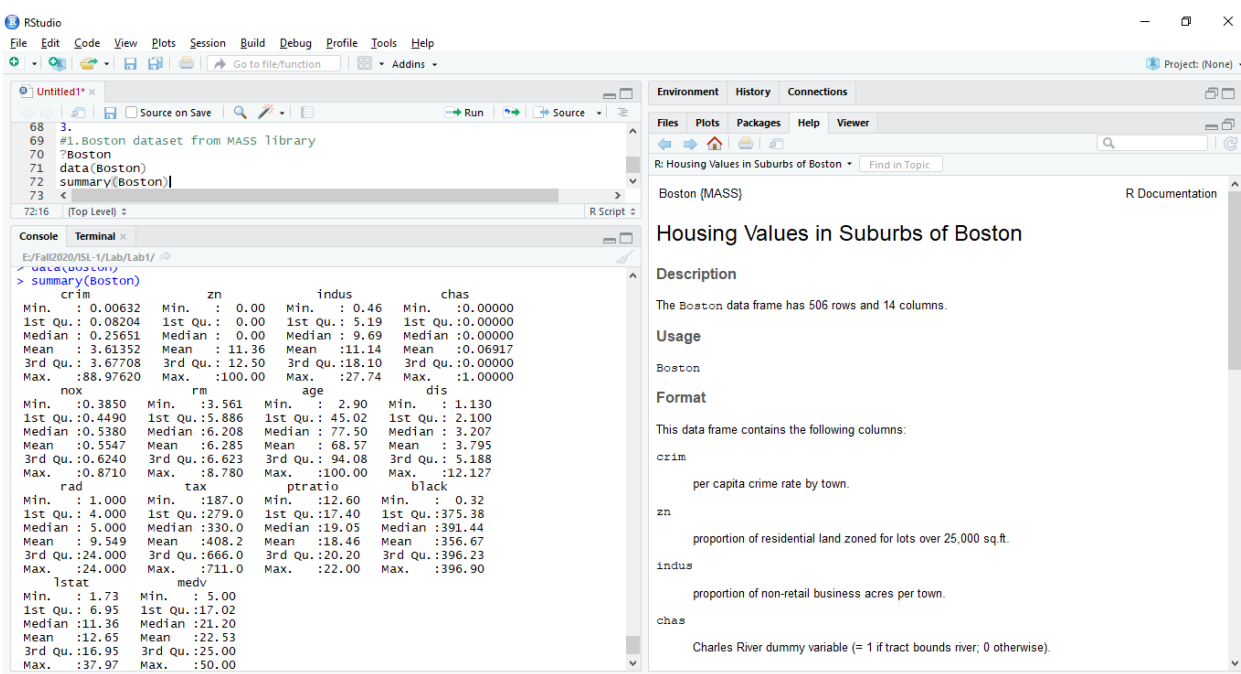


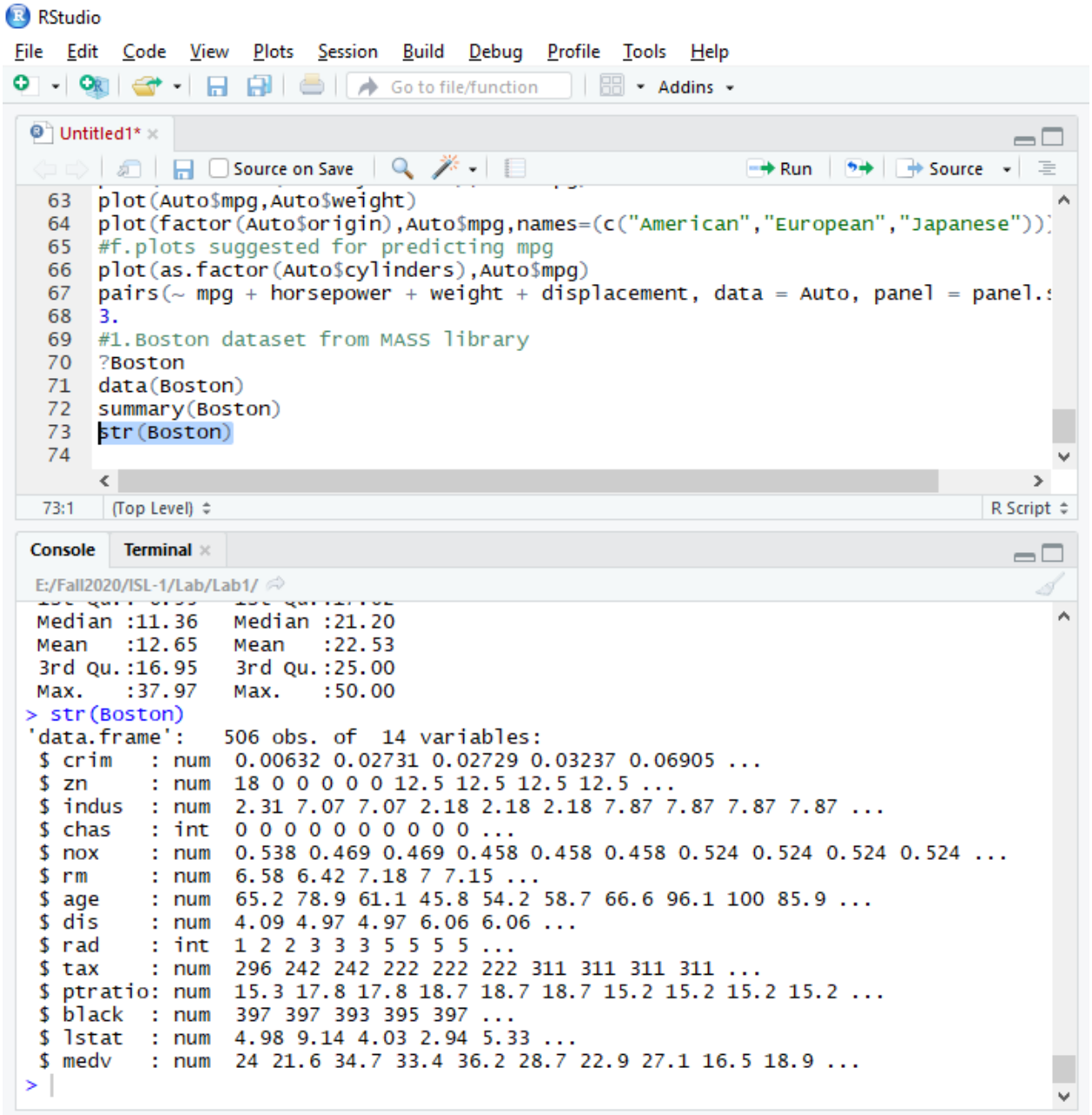


3)

a) To begin, load in the Boston data set. The Boston data set is part of the MASS library in R. > library (MASS) Now the data set is contained in the object Boston. > Boston Read about the data set: > ?Boston How many rows are in this data set? How many columns? What do the rows and columns represent?

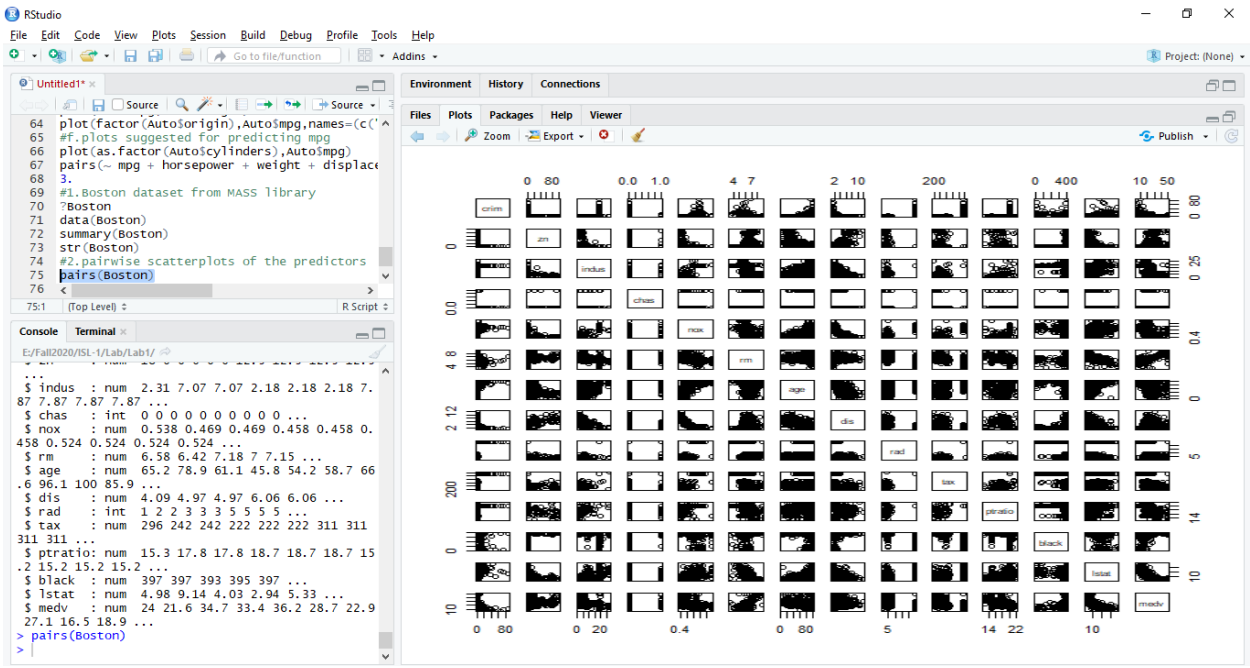
For downloading Boston dataset as it is part of MASS library in R lets install package where dataset exists.

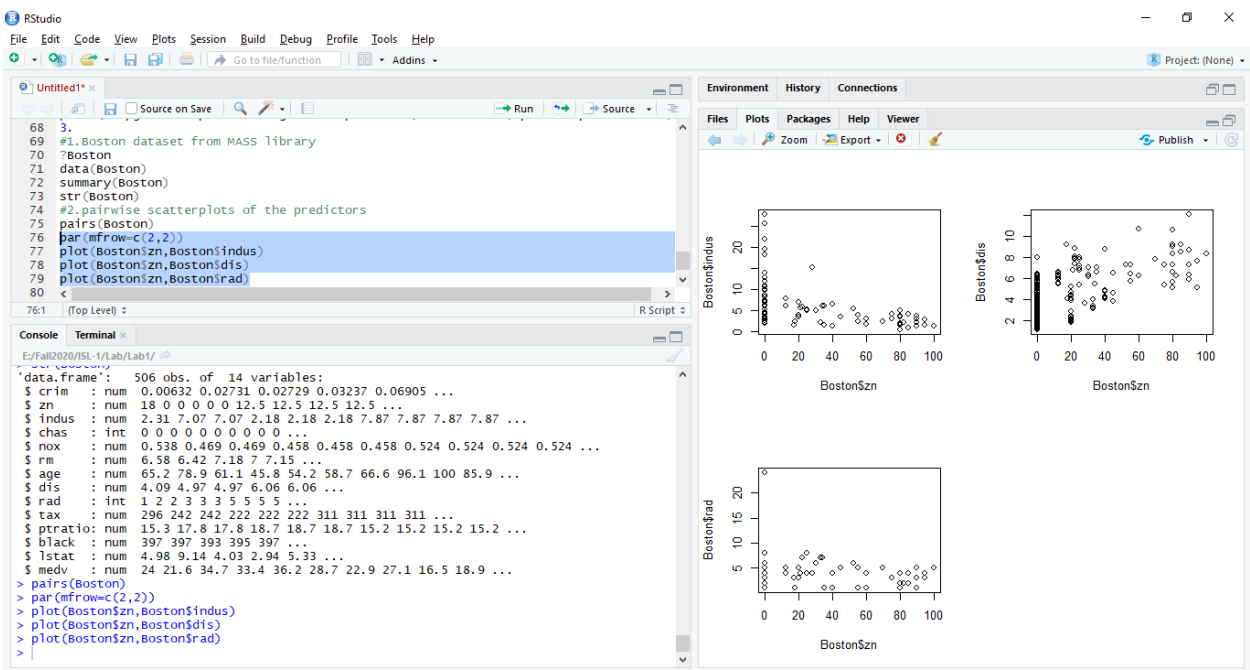




From above there 506 rows and 14 columns

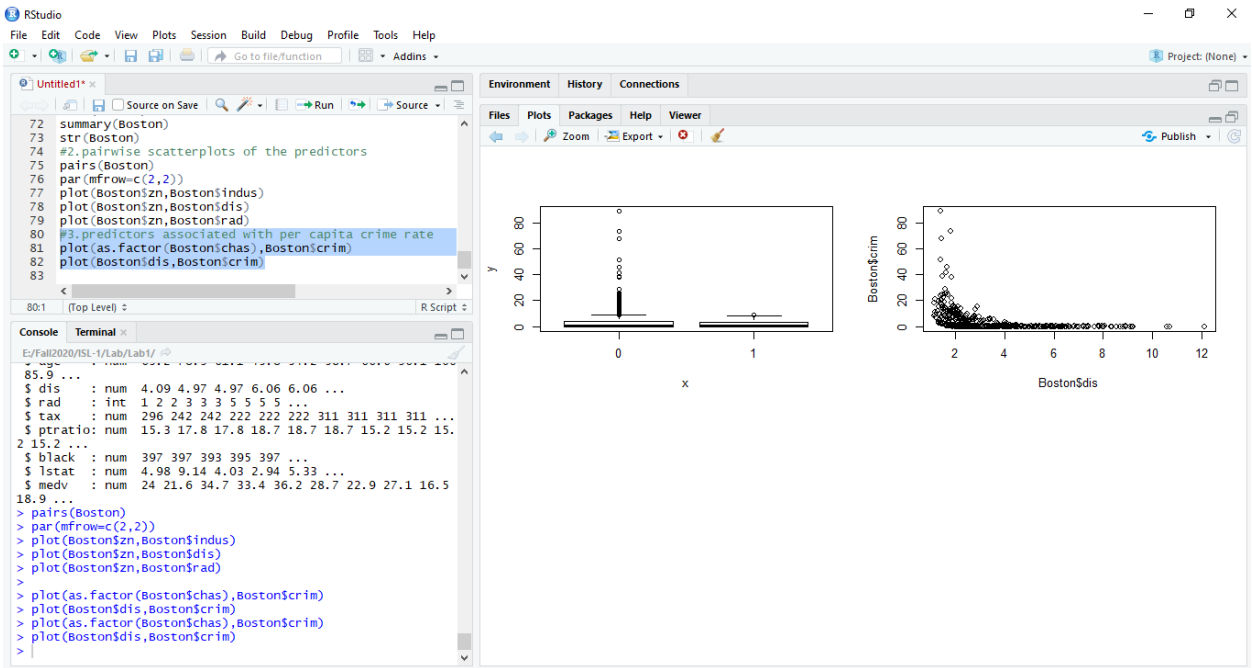
b) Make some pairwise scatterplots of the predictors (columns) in this data set. Describe your findings.



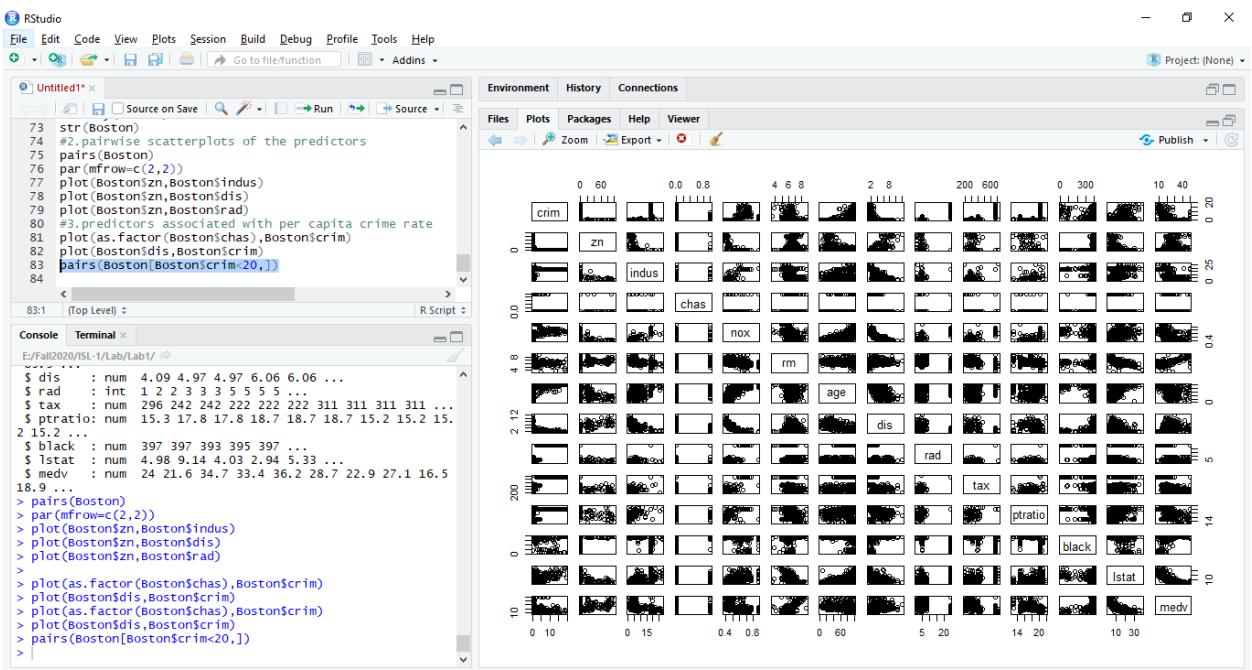


There will be number of scatter plots like above it will become difficult to read all, so maybe a heatmap will be easier to read. Data cleaning is hard.

c) Are any of the predictors associated with per capita crime rate? If so, explain the relationship.

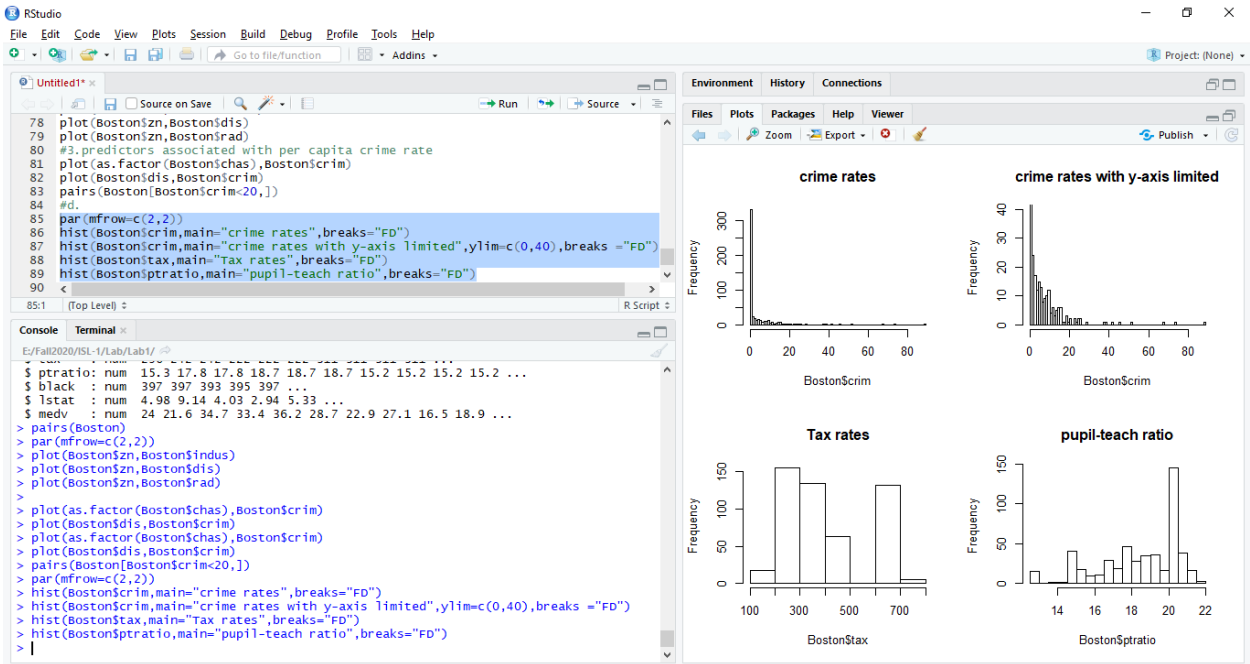


From above diagram crime is large close to five Boston employment centers



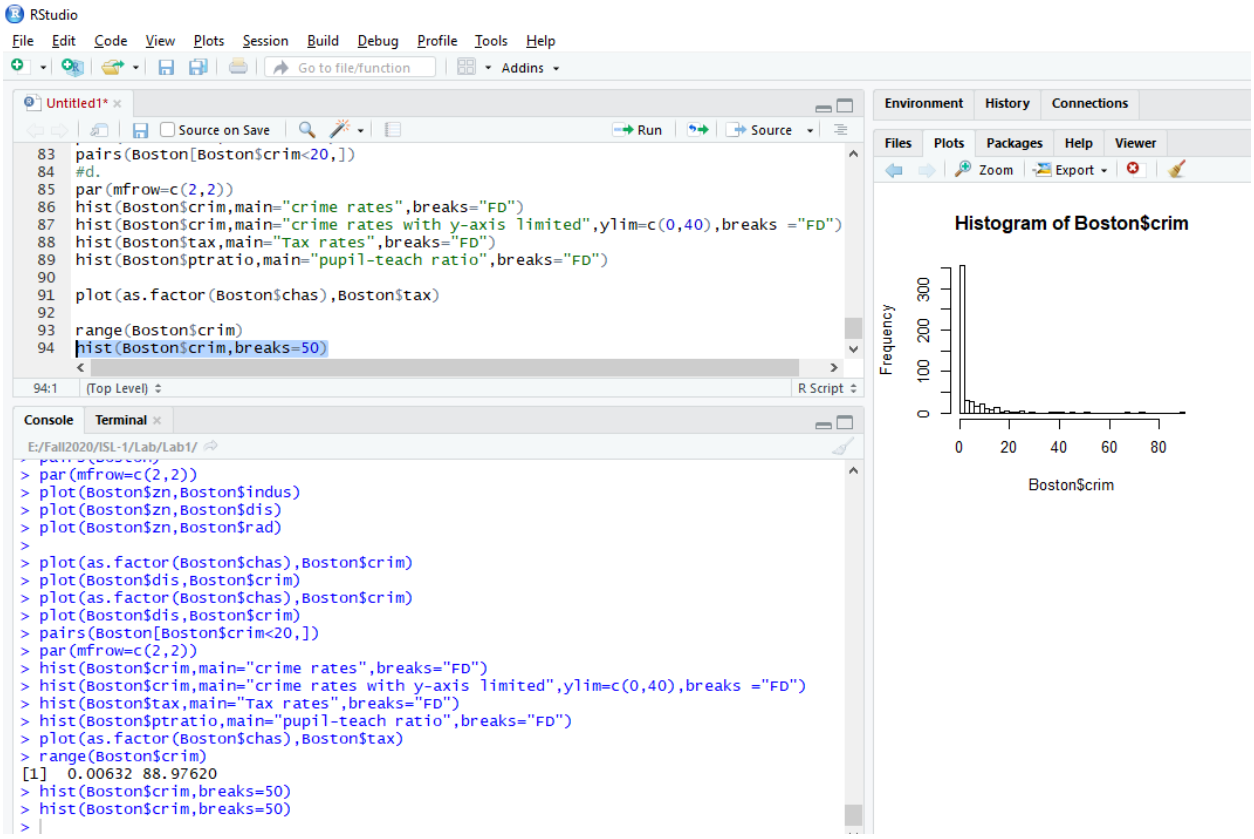
From above Lower status populated areas have the more crime rate.

d) Do any of the suburbs of Boston appear to have particularly high crime rates? Tax rates? Pupil-teacher ratios? Comment on the range of each predictor.



Most of the suburbs do not have any crime rate.

From above surprisingly tax is less near river area



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

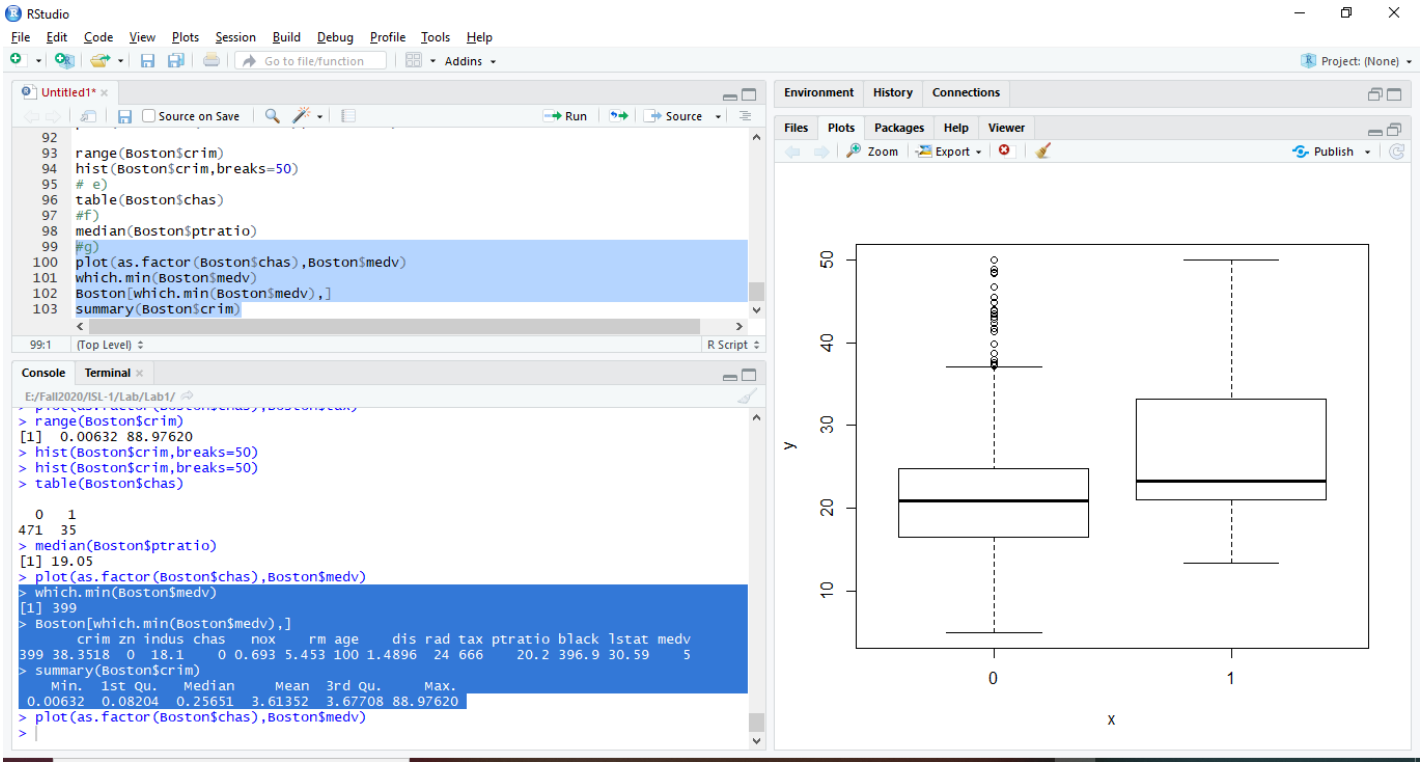
```
> table(Boston$chas)

  0   1
471  35
```

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

```
> median(Boston$ptratio)
[1] 19.05
```

g) Which suburb of Boston has lowest median value of owner occupied homes? 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



Median 0.25 , Maximum is 88.7 and the crime in the median value of owner occupied homes is 38.3518 and we can see that the crime is larger in this area.. It is far from radial highways and Charles river area.

h) In this data set, how many of the suburbs average more than seven rooms per dwelling? More than eight rooms per dwelling? Comment on the suburbs that average more than eight rooms per dwelling.

```
> summary(Boston$rm)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.561   5.886   6.208   6.285   6.623   8.780
```

Average is around 6.285 rooms

```
> table(Boston$rm > 7)

FALSE  TRUE
  442    64
```

From above more than 7 rooms is 64 houses

```
> table(Boston$rm >8)

FALSE  TRUE
  493    13
```

More than 8 rooms – 13 houses and these have the lesser crime rate.

```
> rooms8 = Boston[Boston$rm > 8, ]
> summary(rooms8)

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             rad
 Min.   :0.4161   Min.   :8.034   Min.   : 8.40   Min.   :1.801   Min.   : 2.
000
 1st Qu.:0.5040   1st Qu.:8.247   1st Qu.:70.40   1st Qu.:2.288   1st Qu.: 5.
000
 Median :0.5070   Median :8.297   Median :78.30   Median :2.894   Median : 7.
000
 Mean   :0.5392   Mean   :8.349   Mean   :71.54   Mean   :3.430   Mean   : 7.
462
 3rd Qu.:0.6050   3rd Qu.:8.398   3rd Qu.:86.50   3rd Qu.:3.652   3rd Qu.: 8.
000
 Max.   :0.7180   Max.   :8.780   Max.   :93.90   Max.   :8.907   Max.   :24.
000
      tax           ptratio          black           lstat           medv
 Min.   :224.0   Min.   :13.00   Min.   :354.6   Min.   :2.47   Min.   :21.9
 1st Qu.:264.0   1st Qu.:14.70   1st Qu.:384.5   1st Qu.:3.32   1st Qu.:41.7
 Median :307.0   Median :17.40   Median :386.9   Median :4.14   Median :48.3
 Mean   :325.1   Mean   :16.36   Mean   :385.2   Mean   :4.31   Mean   :44.2
 3rd Qu.:307.0   3rd Qu.:17.40   3rd Qu.:389.7   3rd Qu.:5.12   3rd Qu.:50.0
 Max.   :666.0   Max.   :20.20   Max.   :396.9   Max.   :7.44   Max.   :50.0


> table(rooms8$chas)

 0  1
11  2
```

Crime seems to be less in the houses which have 8 rooms

```
> summary(rooms8$black)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  354.6   384.5   386.9   385.2   389.7   396.9
```

11 of the houses with 8 rooms are not near Charles river (only 2 are near Charles river)

```
> summary(Boston$black)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.32  375.38  391.44  356.67  396.23  396.90
```

All the rooms8 houses blacks population