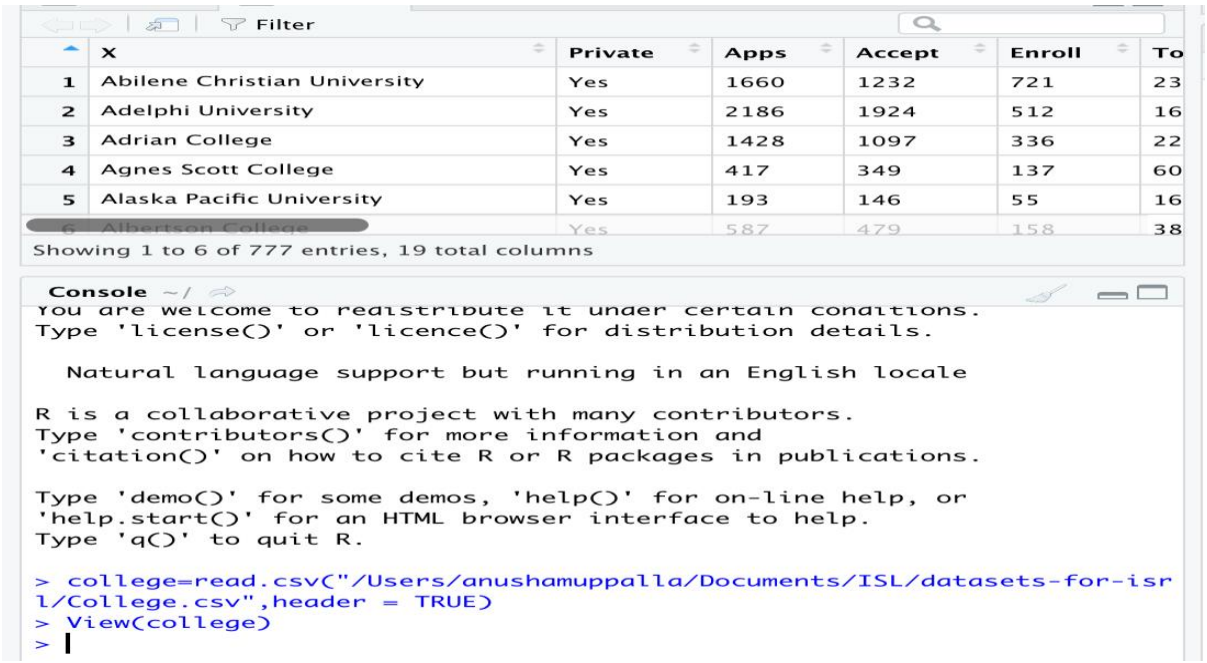
LAB ASSIGNMENT-1

1.This exercise relates to the College data set, which can be found in the file College.csv. It contains a number of variables for 777 different universities and colleges in the US. The variables are

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.



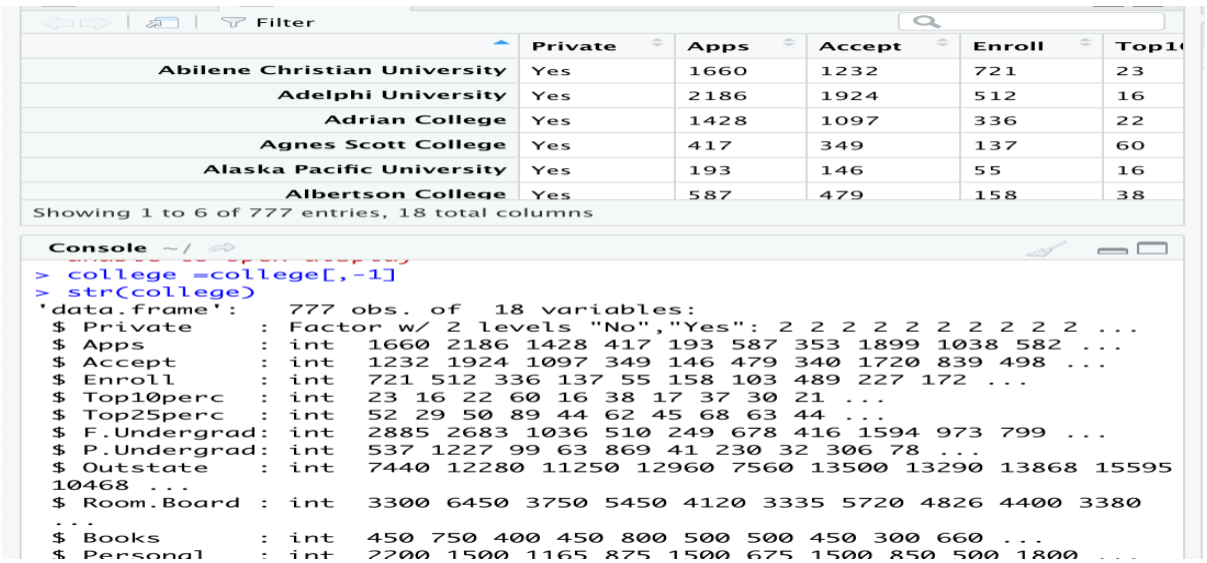
Filter

	X	Private	Apps	Accept	Enroll	To
1	Abilene Christian University	Yes	1660	1232	721	23
2	Adelphi University	Yes	2186	1924	512	16
3	Adrian College	Yes	1428	1097	336	22
4	Agnes Scott College	Yes	417	349	137	60
5	Alaska Pacific University	Yes	193	146	55	16
6	Albertson College	Yes	587	479	158	38

Showing 1 to 6 of 777 entries, 19 total columns

```
Console ~/
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("/Users/anushamuppalla/Documents/ISL/datasets-for-isr
l/College.csv",header = TRUE)
> View(college)
>
```

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?

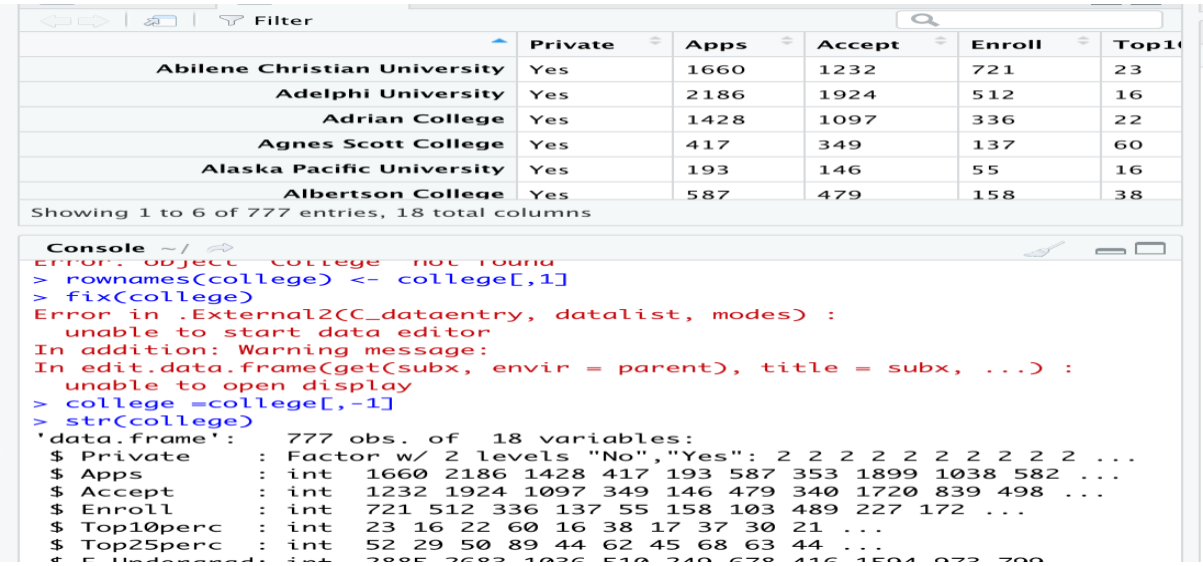


Filter

	Private	Apps	Accept	Enroll	Top1
Abilene Christian University	Yes	1660	1232	721	23
Adelphi University	Yes	2186	1924	512	16
Adrian College	Yes	1428	1097	336	22
Agnes Scott College	Yes	417	349	137	60
Alaska Pacific University	Yes	193	146	55	16
Albertson College	Yes	587	479	158	38

Showing 1 to 6 of 777 entries, 18 total columns

```
Console ~/
> college =college[,-1]
> str(college)
'data.frame':    777 obs. of  18 variables:
 $ Private    : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...
 $ Apps       : int  1660 2186 1428 417 193 587 353 1899 1038 582 ...
 $ Accept     : int  1232 1924 1097 349 146 479 340 1720 839 498 ...
 $ Enroll     : int  721 512 336 137 55 158 103 489 227 172 ...
 $ Top10perc  : int  23 16 22 60 16 38 17 37 30 21 ...
 $ Top25perc  : int  52 29 50 89 44 62 45 68 63 44 ...
 $ F.Undergrad: int  2885 2683 1036 510 249 678 416 1594 973 799 ...
 $ P.Undergrad: int  537 1227 99 63 869 41 230 32 306 78 ...
 $ Outstate   : int  7440 12280 11250 12960 7560 13500 13290 13868 15595
 10468 ...
 $ Room.Board : int  3300 6450 3750 5450 4120 3335 5720 4826 4400 3380
 ...
 $ Books      : int  450 750 400 450 800 500 500 450 300 660 ...
 $ Personal   : int  2200 1500 1165 875 1500 675 1500 850 500 1800 ...
```



	Private	Apps	Accept	Enroll	Top1
Abilene Christian University	Yes	1660	1232	721	23
Adelphi University	Yes	2186	1924	512	16
Adrian College	Yes	1428	1097	336	22
Agnes Scott College	Yes	417	349	137	60
Alaska Pacific University	Yes	193	146	55	16
Albertson College	Yes	587	479	158	38

Showing 1 to 6 of 777 entries, 18 total columns

```
> rownames(college) <- college[,1]
> fix(college)
Error in .External2(C_dataentry, datalist, modes) :
  unable to start data editor
In addition: Warning message:
In edit.data.frame(get(subx, envir = parent), title = subx, ...) :
  unable to open display
> college =college[,-1]
> str(college)
'data.frame':    777 obs. of  18 variables:
 $ Private     : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...
 $ Apps        : int  1660 2186 1428 417 193 587 353 1899 1038 582 ...
 $ Accept      : int  1232 1924 1097 349 146 479 340 1720 839 498 ...
 $ Enroll      : int  721 512 336 137 55 158 103 489 227 172 ...
 $ Top10perc   : int  23 16 22 60 16 38 17 37 30 21 ...
 $ Top25perc   : int  52 29 50 89 44 62 45 68 63 44 ...
```

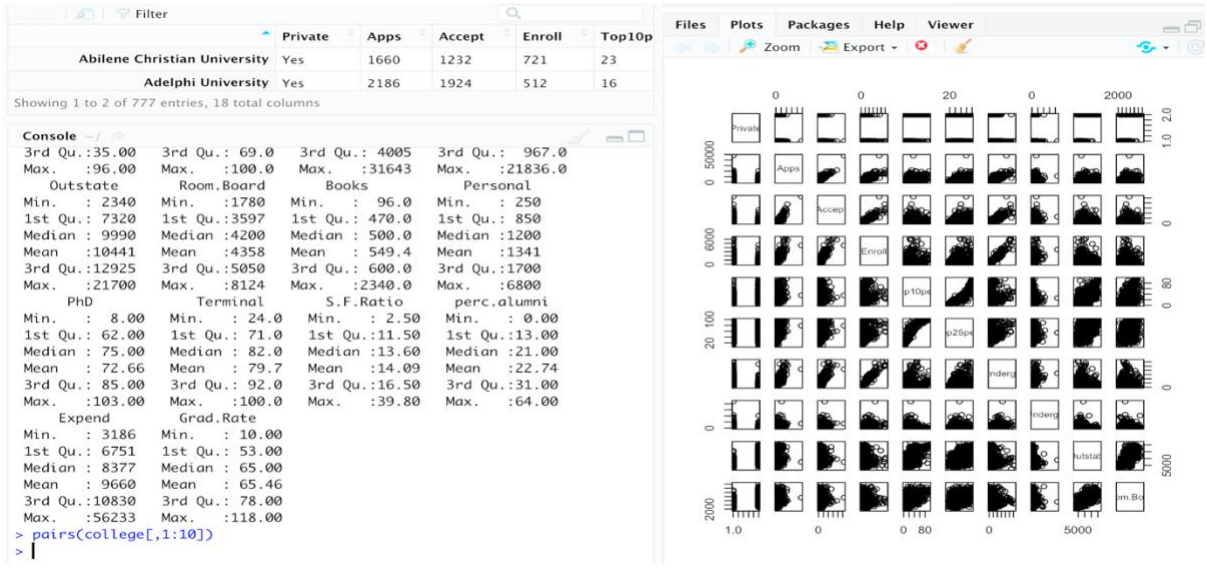
c) i. Use the summary() function to produce a numerical summary of the variables in the data set.

```
> summary(college)
 Private        Apps           Accept          Enroll
 No :212   Min.   :   81   Min.   :   72   Min.   :  35
 Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242
           Median : 1558   Median : 1110   Median : 434
           Mean   : 3002   Mean   : 2019   Mean   : 780
           3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902
           Max.   :48094   Max.   :26330   Max.   :6392
   Top10perc       Top25perc      F.Undergrad     P.Undergrad
 Min.   : 1.00   Min.   :  9.0   Min.   :  139   Min.   :    1.0
 1st Qu.:15.00   1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0
 Median :23.00   Median : 54.0   Median : 1707   Median :  353.0
 Mean   :27.56   Mean   : 55.8   Mean   : 3700   Mean   :  855.3
 3rd Qu.:35.00   3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0
 Max.   :96.00   Max.   :100.0   Max.   :31643   Max.   :21836.0
    Outstate       Room.Board       Books           Personal
 Min.   : 2340   Min.   :1780   Min.   :  96.0   Min.   : 250
```

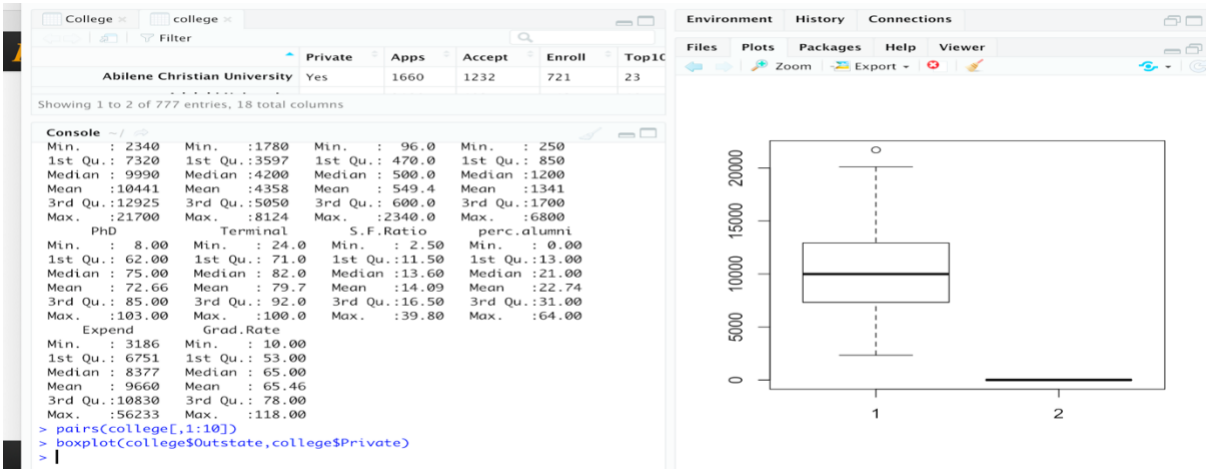
ii. Use the pairs() function to produce a scatterplot matrix of the first ten columns or variables of the data. Recall that you can reference the first ten columns of a matrix A using A[,1:10].


```
> summary(college)
 Private        Apps           Accept          Enroll
 No :212   Min.   :   81   Min.   :   72   Min.   :  35
 Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242
           Median : 1558   Median : 1110   Median : 434
           Mean   : 3002   Mean   : 2019   Mean   : 780
           3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902
           Max.   :48094   Max.   :26330   Max.   :6392
   Top10perc       Top25perc      F.Undergrad     P.Undergrad
 Min.   : 1.00   Min.   :  9.0   Min.   :  139   Min.   :    1.0
 1st Qu.:15.00   1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0
 Median :23.00   Median : 54.0   Median : 1707   Median :  353.0
 Mean   :27.56   Mean   : 55.8   Mean   : 3700   Mean   :  855.3
 3rd Qu.:35.00   3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0
 Max.   :96.00   Max.   :100.0   Max.   :31643   Max.   :21836.0
    Outstate       Room.Board       Books           Personal
 Min.   : 2340   Min.   :1780   Min.   :  96.0   Min.   : 250
```

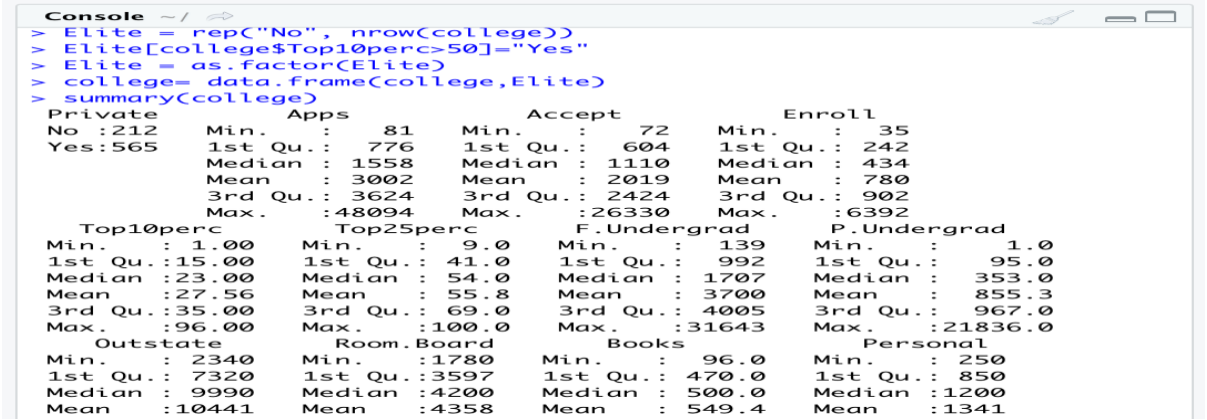
iii. Use the plot() function to produce side-by-side boxplots of Outstate versus Private.



```
> pairs(college[,1:10])
> boxplot(college$Outstate,college$Private)
```

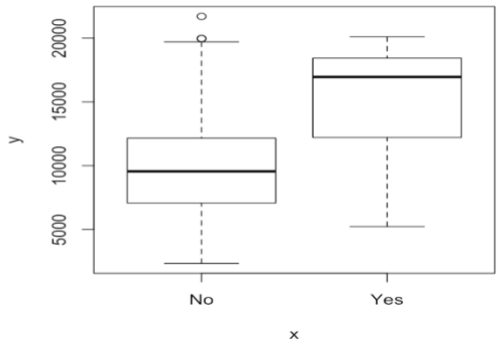


iv. Create a new qualitative variable, called Elite, by *binning* the Top10perc variable. We are going to divide universities into two groups based on whether or not the proportion of students coming from the top 10% of their high school classes exceeds 50%.
Use the summary() function to see how many elite universities there are. Now use the plot() function to produce side-by-side boxplots of Outstate versus Elite.



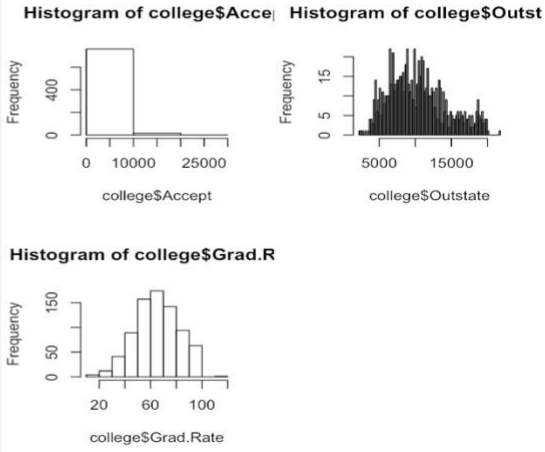
```
> Elite = rep("No", nrow(college))
> Elite[college$Top10perc>50]="Yes"
> Elite = as.factor(Elite)
> college= data.frame(college,Elite)
> summary(college)
 Private        Apps           Accept          Enroll
 No :212    Min.   :   81   Min.   :   72   Min.   :  35
 Yes:565    1st Qu.:  776   1st Qu.:  604   1st Qu.: 242
            Median : 1558   Median : 1110   Median : 434
            Mean   : 3002   Mean   : 2019   Mean   : 780
            3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902
            Max.   :48094   Max.   :26330   Max.   :6392
   Top10perc       Top25perc      F.Undergrad     P.Undergrad
 Min.   : 1.00   Min.   :  9.0   Min.   :  139   Min.   :    1.0
 1st Qu.:15.00   1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0
 Median :23.00   Median : 54.0   Median : 1707   Median :  353.0
 Mean   :27.56   Mean   : 55.8   Mean   : 3700   Mean   :  855.3
 3rd Qu.:35.00   3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0
 Max.   :96.00   Max.   :100.0   Max.   :31643   Max.   :21836.0
    Outstate       Room.Board       Books           Personal
 Min.   : 2340   Min.   :1780   Min.   :  96.0   Min.   : 250
 1st Qu.: 7320   1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850
 Median : 9990   Median :4200   Median : 500.0   Median :1200
 Mean   :10441   Mean   :4358   Mean   : 549.4   Mean   :1341
```

```
Console
    Outstate       Room.Board       Books           Personal
 Min.   : 2340   Min.   :1780   Min.   :  96.0   Min.   : 250
 1st Qu.: 7320   1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850
 Median : 9990   Median :4200   Median : 500.0   Median :1200
 Mean   :10441   Mean   :4358   Mean   : 549.4   Mean   :1341
 3rd Qu.:12925   3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700
 Max.   :21700   Max.   :8124   Max.   :2340.0   Max.   :6800
      PhD            Terminal       S.F.Ratio      perc.alumni
 Min.   :  8.00   Min.   : 24.0   Min.   : 2.50   Min.   : 0.00
 1st Qu.: 62.00   1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00
 Median : 75.00   Median : 82.0   Median :13.60   Median :21.00
 Mean   : 72.66   Mean   : 79.7   Mean   :14.09   Mean   :22.74
 3rd Qu.: 85.00   3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00
 Max.   :103.00   Max.   :100.0   Max.   :39.80   Max.   :64.00
     Expend         Grad.Rate      Elite
 Min.   : 3186   Min.   : 10.00   No :699
 1st Qu.: 6751   1st Qu.: 53.00   Yes: 78
 Median : 8377   Median : 65.00
 Mean   : 9660   Mean   : 65.46
 3rd Qu.:10830   3rd Qu.: 78.00
 Max.   :56233   Max.   :118.00
> plot(college$Elite,college$Outstate)
>
```



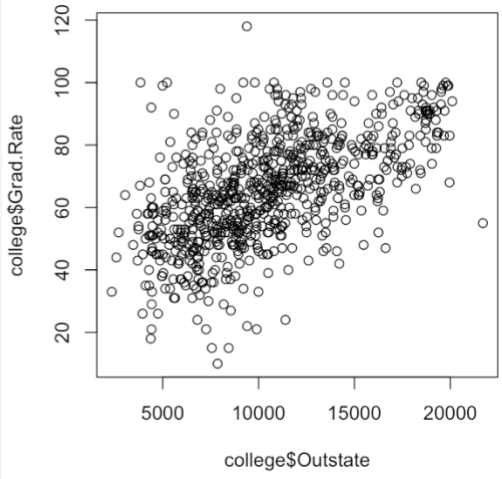
v.Use the hist() function to produce some histograms with differing numbers of bins for a few of the quantitative variables. You may find the command par(mfrow=c(2,2)) useful: it will divide the print window into four regions so that four plots can be made simultaneously. Modifying the arguments to this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

```
Showing 1 to 1 of 777 entries, 19 total columns
Console
 3rd Qu.: 85.00   3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00
 Max.   :103.00   Max.   :100.0   Max.   :39.80   Max.   :64.00
     Expend         Grad.Rate      Elite
 Min.   : 3186   Min.   : 10.00   No :699
 1st Qu.: 6751   1st Qu.: 53.00   Yes: 78
 Median : 8377   Median : 65.00
 Mean   : 9660   Mean   : 65.46
 3rd Qu.:10830   3rd Qu.: 78.00
 Max.   :56233   Max.   :118.00
> plot(college$Elite,college$Outstate)
> par(mfrow=c(2,2))
> hist(college$Accept,breaks = 3)
> hist(college$Outstate, breaks = 100)
> hist(college$Grad.Rate
+
+ )
> hist(college$Outstate, breaks = 100)
> hist(college$Outstate, breaks = 100)
> par(mfrow=c(2,2))
> hist(college$Accept,breaks = 3)
> hist(college$Outstate, breaks = 100)
> hist(college$Grad.Rate)
>
```



vi. Continue exploring the data, and provide a brief summary of what you discover.

```
Console
     Expend         Grad.Rate      Elite
 Min.   : 3186   Min.   : 10.00   No :699
 1st Qu.: 6751   1st Qu.: 53.00   Yes: 78
 Median : 8377   Median : 65.00
 Mean   : 9660   Mean   : 65.46
 3rd Qu.:10830   3rd Qu.: 78.00
 Max.   :56233   Max.   :118.00
> plot(college$Elite,college$Outstate)
> par(mfrow=c(2,2))
> hist(college$Accept,breaks = 3)
> hist(college$Outstate, breaks = 100)
> hist(college$Grad.Rate
+
+ )
> hist(college$Outstate, breaks = 100)
> hist(college$Outstate, breaks = 100)
> par(mfrow=c(2,2))
> hist(college$Accept,breaks = 3)
> hist(college$Outstate, breaks = 100)
> hist(college$Grad.Rate)
> par(mfrow=c(1,1))
> plot(college$Outstate, college$Grad.Rate)
>
```

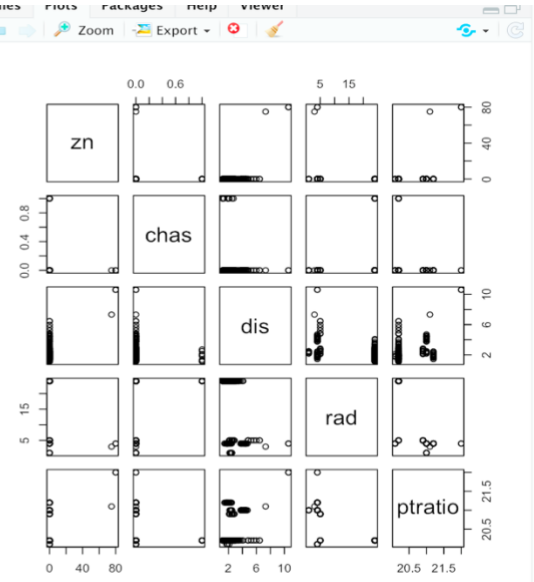


2.This exercise involves the Auto data set studied in the lab. Make sure that the missing values have been removed from the data.

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

```
1  install.packages("ISLR")
2  library(ISLR)
3  data("Auto")
4  summary(Auto)
5  summary(complete.cases(Auto))
6  sapply(Auto, class)
7
```

```
Console ~/
Median :2804   Median :15.50   Median :76.00   Median :1.000
Mean   :2978   Mean   :15.54   Mean   :75.98   Mean   :1.577
3rd Qu.:3615   3rd Qu.:17.02   3rd Qu.:79.00   3rd Qu.:2.000
Max.   :5140   Max.   :24.80   Max.   :82.00   Max.   :3.000

                name
amc matador       :  5
ford pinto        :  5
toyota corolla    :  5
amc gremlin       :  4
amc hornet        :  4
chevrolet chevette:  4
(Other)           :365
> summary(complete.cases(Auto))
   Mode    TRUE
logical     392
> sapply(Auto, class)
         mpg    cylinders displacement   horsepower       weight
   "numeric"    "numeric"    "numeric"    "numeric"    "numeric"
acceleration         year       origin         name
   "numeric"    "numeric"    "numeric"     "factor"
>
```



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

```
> summary(complete.cases(Auto))
   Mode    TRUE
logical     392
> sapply(Auto, class)
         mpg    cylinders displacement   horsepower       weight
   "numeric"    "numeric"    "numeric"    "numeric"    "numeric"
acceleration         year       origin         name
   "numeric"    "numeric"    "numeric"     "factor"
> sapply(Auto[ ,(c(1,3,4,5,6))], range)
      mpg displacement horsepower weight acceleration
[1,]  9.0           68         46   1613          8.0
[2,] 46.6          455        230   5140         24.8
>
```

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

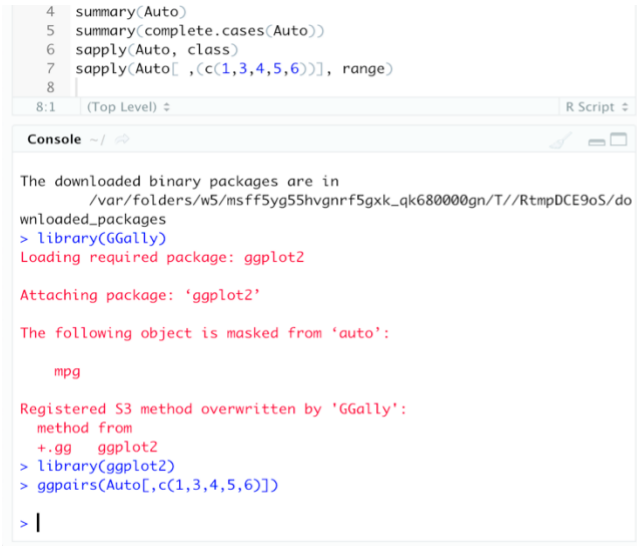
```
> sapply(Auto[ ,(c(1,3,4,5,6))], range)
      mpg displacement horsepower weight acceleration
[1,]  9.0           68         46   1613          8.0
[2,] 46.6          455        230   5140         24.8
>
> sapply(Auto[ ,(c(1,3,4,5,6))], mean)
         mpg displacement   horsepower       weight acceleration
    23.44592    194.41199    104.46939   2977.58418     15.54133
>
```

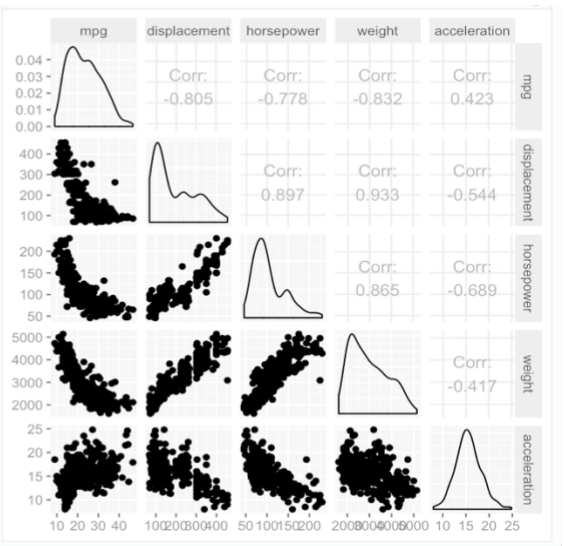
```
> sapply(Auto[ ,(c(1,3,4,5,6))], sd)
         mpg displacement   horsepower       weight acceleration
    7.805007   104.644004    38.491160   849.402560     2.758864
>
```

(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?

```
> Auto1 <- Auto[-(10:85),]
> sapply(Auto1[ ,(c(1,3,4,5,6))], range)
      mpg displacement horsepower weight acceleration
[1,] 11.0           68         46   1649          8.5
[2,] 46.6          455        230   4997         24.8
> sapply(Auto1[ ,(c(1,3,4,5,6))], mean)
         mpg displacement   horsepower       weight acceleration
    24.40443    187.24051    100.72152   2935.97152     15.72690
> sapply(Auto1[ ,(c(1,3,4,5,6))], sd)
         mpg displacement   horsepower       weight acceleration
    7.867283    99.678367    35.708853   811.300208     2.693721
>
```

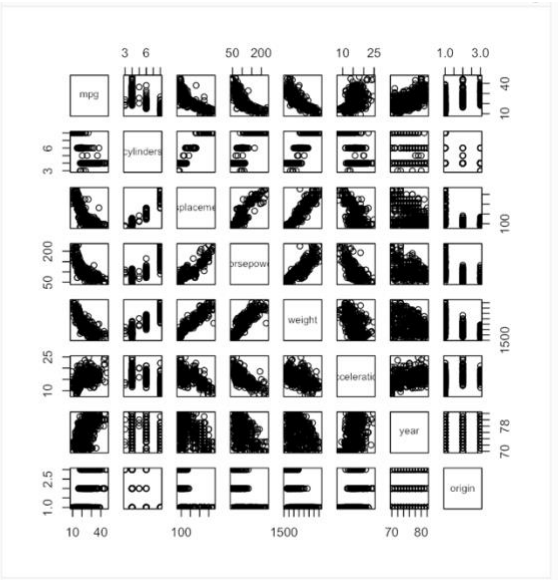
(e) Using the full data set, investigate the predictors graphically, using scatterplots or other tools of your choice. Create some plots highlighting the relationships among the predictors. Comment on your findings.





```
summary(Auto)
summary(complete.cases(Auto))
sapply(Auto, class)
sapply(Auto[ ,(c(1,3,4,5,6))], range)
```

```
Console ~/
The downloaded binary packages are in
        /var/folders/w5/msff5yg55hvgnrf5gxk_qk680000gn/T//RtmpDCE9oS/do
wnloaded_packages
> library(GGally)
Loading required package: ggplot2

Attaching package: 'ggplot2'

The following object is masked from 'auto':

    mpg

Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2
> library(ggplot2)
> ggpairs(Auto[,c(1,3,4,5,6)])

> pairs(Auto[,1:8])
>
```



(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.

```
8:1  (Top Level)  R Script
Console ~/
> pairs(Auto[,1:8])
> model <- lm(mpg ~ cylinders+year+weight+horsepower+displacement+accel
eration+origin, data= Auto)
>
> t(coefficients(model))
     (Intercept)  cylinders      year       weight  horsepower
[1,]   -17.21843 -0.4933763 0.7507727 -0.006474043 -0.01695114
     displacement acceleration  origin
[1,]   0.01989564   0.08057584 1.42614
> anova(model)
Analysis of Variance Table

Response: mpg
              Df  Sum Sq Mean Sq   F value    Pr(>F)
cylinders      1 14403.1 14403.1 1300.6838 < 2.2e-16 ***
year           1  2629.2  2629.2  237.4301 < 2.2e-16 ***
weight         1  2222.7  2222.7  200.7267 < 2.2e-16 ***
horsepower     1     1.6     1.6    0.1446    0.7039
displacement   1    10.8    10.8    0.9756    0.3239
acceleration   1     8.2     8.2    0.7443    0.3888
origin         1   291.1   291.1   26.2912 4.666e-07 ***
Residuals    384  4252.2    11.1
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

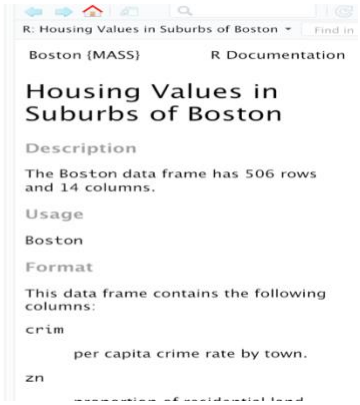
3.This exercise involves the Boston housing data set.

(a) To begin, load in the Boston data set. The Boston data set is part of the MASS *library* in R. Now the data set is contained in the object Boston.

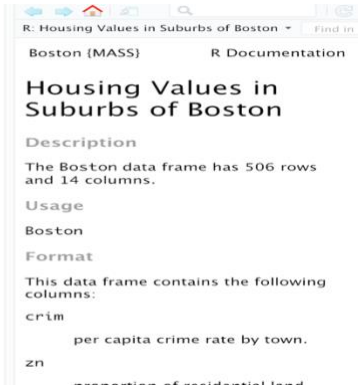
Read about the data set:

How many rows are in this data set? How many columns? What do the rows and columns represent?

```
> library(MASS)
> Boston
      crim    zn indus chas    nox    rm   age    dis rad tax ptratio  black lstat medv
1  0.00632  18.0  2.31    0 0.5380 6.575  65.2 4.0900   1 296    15.3 396.90  4.98 24.0
2  0.02731   0.0  7.07    0 0.4690 6.421  78.9 4.9671   2 242    17.8 396.90  9.14 21.6
3  0.02729   0.0  7.07    0 0.4690 7.185  61.1 4.9671   2 242    17.8 392.83  4.03 34.7
4  0.03237   0.0  2.18    0 0.4580 6.998  45.8 6.0622   3 222    18.7 394.63  2.94 33.4
5  0.06905   0.0  2.18    0 0.4580 7.147  54.2 6.0622   3 222    18.7 396.90  5.33 36.2
6  0.02985   0.0  2.18    0 0.4580 6.430  58.7 6.0622   3 222    18.7 394.12  5.21 28.7
7  0.08829  12.5  7.87    0 0.5240 6.012  66.6 5.5605   5 311    15.2 395.60 12.43 22.9
8  0.14455  12.5  7.87    0 0.5240 6.172  96.1 5.9505   5 311    15.2 396.90 19.15 27.1
9  0.21124  12.5  7.87    0 0.5240 5.631 100.0 6.0821   5 311    15.2 386.63 29.93 16.5
10 0.17004  12.5  7.87    0 0.5240 6.004  85.9 6.5921   5 311    15.2 386.71 17.10 18.9
11 0.22489  12.5  7.87    0 0.5240 6.377  94.3 6.3467   5 311    15.2 392.52 20.45 15.0
12 0.11747  12.5  7.87    0 0.5240 6.009  82.9 6.2267   5 311    15.2 396.90 13.27 18.9
13 0.09378  12.5  7.87    0 0.5240 5.889  39.0 5.4509   5 311    15.2 390.50 15.71 21.7
14 0.62976   0.0  8.14    0 0.5380 5.949  61.8 4.7075   4 307    21.0 396.90  8.26 20.4
15 0.63796   0.0  8.14    0 0.5380 6.096  84.5 4.4619   4 307    21.0 380.02 10.26 18.2
16 0.62739   0.0  8.14    0 0.5380 5.834  56.5 4.4986   4 307    21.0 395.62  8.47 19.9
17 1.05393   0.0  8.14    0 0.5380 5.935  29.3 4.4986   4 307    21.0 386.85  6.58 23.1
18 0.78420   0.0  8.14    0 0.5380 5.990  81.7 4.2579   4 307    21.0 386.75 14.67 17.5
19 0.80271   0.0  8.14    0 0.5380 5.456  36.6 3.7965   4 307    21.0 288.99 11.69 20.2
20 0.72580   0.0  8.14    0 0.5380 5.727  69.5 3.7965   4 307    21.0 390.95 11.28 18.2
21 1.25179   0.0  8.14    0 0.5380 5.570  98.1 3.7979   4 307    21.0 376.57 21.02 13.6
22 0.85204   0.0  8.14    0 0.5380 5.965  89.2 4.0123   4 307    21.0 392.53 13.83 19.6
23 1.23247   0.0  8.14    0 0.5380 6.142  91.7 3.9769   4 307    21.0 396.90 18.72 15.2
24 0.98843   0.0  8.14    0 0.5380 5.813 100.0 4.0952   4 307    21.0 394.54 19.88 14.5
25 0.75026   0.0  8.14    0 0.5380 5.924  94.1 4.3996   4 307    21.0 394.33 16.30 15.6
26 0.84054   0.0  8.14    0 0.5380 5.599  85.7 4.4546   4 307    21.0 303.42 16.51 13.9
```

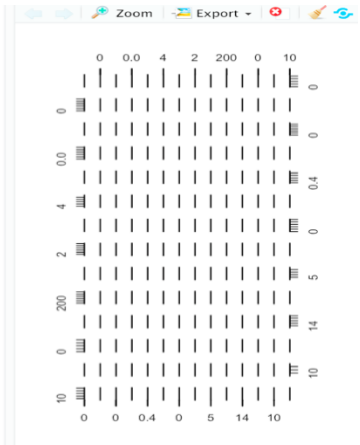



```
48 0.22927   0.0  6.91    0 0.4480 6.030  85.5 5.6894   3 233    17.9 392.74 18.80 16.6
49 0.25387   0.0  6.91    0 0.4480 5.399  95.3 5.8700   3 233    17.9 396.90 30.81 14.4
50 0.21977   0.0  6.91    0 0.4480 5.602  62.0 6.0877   3 233    17.9 396.90 16.20 19.4
51 0.08873  21.0  5.64    0 0.4390 5.963  45.7 6.8147   4 243    16.8 395.56 13.45 19.7
52 0.04337  21.0  5.64    0 0.4390 6.115  63.0 6.8147   4 243    16.8 393.97  9.43 20.5
53 0.05360  21.0  5.64    0 0.4390 6.511  21.1 6.8147   4 243    16.8 396.90  5.28 25.0
54 0.04981  21.0  5.64    0 0.4390 5.998  21.4 6.8147   4 243    16.8 396.90  8.43 23.4
55 0.01360  75.0  4.00    0 0.4100 5.888  47.6 7.3197   3 469    21.1 396.90 14.80 18.9
56 0.01311  90.0  1.22    0 0.4030 7.249  21.9 8.6966   5 226    17.9 395.93  4.81 35.4
57 0.02055  85.0  0.74    0 0.4100 6.383  35.7 9.1876   2 313    17.3 396.90  5.77 24.7
58 0.01432 100.0  1.32    0 0.4110 6.816  40.5 8.3248   5 256    15.1 392.90  3.95 31.6
59 0.15445  25.0  5.13    0 0.4530 6.145  29.2 7.8148   8 284    19.7 390.68  6.86 23.3
60 0.10328  25.0  5.13    0 0.4530 5.927  47.2 6.9320   8 284    19.7 396.90  9.22 19.6
61 0.14932  25.0  5.13    0 0.4530 5.741  66.2 7.2254   8 284    19.7 395.11 13.15 18.7
62 0.17171  25.0  5.13    0 0.4530 5.966  93.4 6.8185   8 284    19.7 378.08 14.44 16.0
63 0.11027  25.0  5.13    0 0.4530 6.456  67.8 7.2255   8 284    19.7 396.90  6.73 22.2
64 0.12650  25.0  5.13    0 0.4530 6.762  43.4 7.9809   8 284    19.7 395.58  9.50 25.0
65 0.01951  17.5  1.38    0 0.4161 7.104  59.5 9.2229   3 216    18.6 393.24  8.05 33.0
66 0.03584  80.0  3.37    0 0.3980 6.290  17.8 6.6115   4 337    16.1 396.90  4.67 23.5
67 0.04379  80.0  3.37    0 0.3980 5.787  31.1 6.6115   4 337    16.1 396.90 10.24 19.4
68 0.05789  12.5  6.07    0 0.4090 5.878  21.4 6.4980   4 345    18.9 396.21  8.10 22.0
69 0.13554  12.5  6.07    0 0.4090 5.594  36.8 6.4980   4 345    18.9 396.90 13.09 17.4
70 0.12816  12.5  6.07    0 0.4090 5.885  33.0 6.4980   4 345    18.9 396.90  8.79 20.9
71 0.08826   0.0 10.81    0 0.4130 6.417   6.6 5.2873   4 305    19.2 383.73  6.72 24.2
 [ reached 'max' / getOption("max.print") -- omitted 435 rows ]
> ?Boston
> dim(Boston)
[1] 506  14
>
```



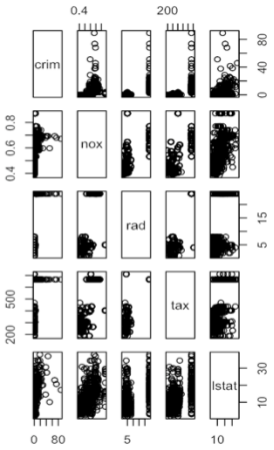
(b) Make some pairwise scatterplots of the predictors (columns) in this data set. Describe your findings.

```
65 0.01951  17.5  1.38    0 0.4161 7.104  59.5 9.2229   3 216    18.6 393.24  8.05 33.0
66 0.03584  80.0  3.37    0 0.3980 6.290  17.8 6.6115   4 337    16.1 396.90  4.67 23.5
67 0.04379  80.0  3.37    0 0.3980 5.787  31.1 6.6115   4 337    16.1 396.90 10.24 19.4
68 0.05789  12.5  6.07    0 0.4090 5.878  21.4 6.4980   4 345    18.9 396.21  8.10 22.0
69 0.13554  12.5  6.07    0 0.4090 5.594  36.8 6.4980   4 345    18.9 396.90 13.09 17.4
70 0.12816  12.5  6.07    0 0.4090 5.885  33.0 6.4980   4 345    18.9 396.90  8.79 20.9
71 0.08826   0.0 10.81    0 0.4130 6.417   6.6 5.2873   4 305    19.2 383.73  6.72 24.2
 [ reached 'max' / getOption("max.print") -- omitted 435 rows ]
> ?Boston
> dim(Boston)
[1] 506  14
> pairs(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 ...
>
```

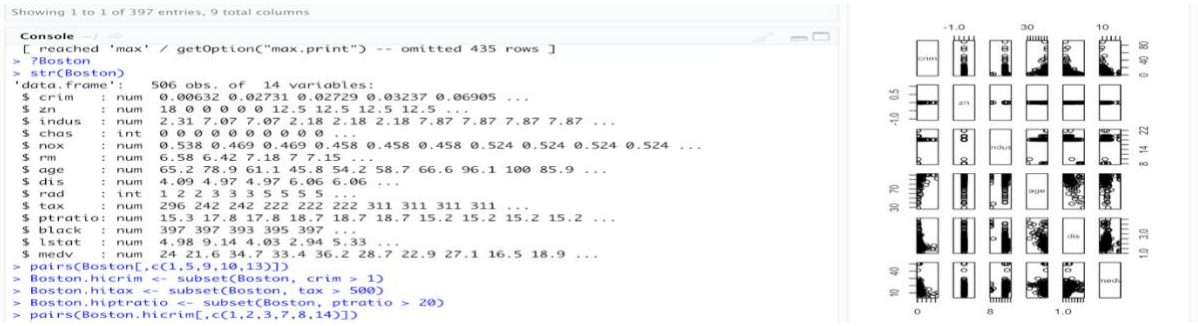


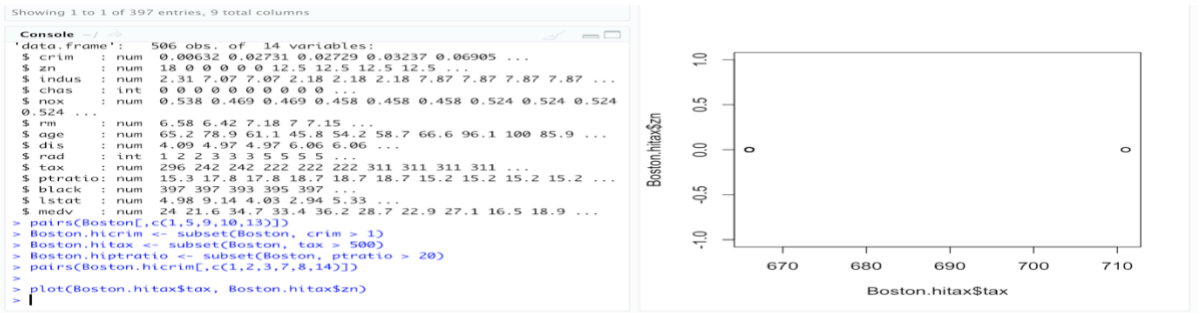
(c) Are any of the predictors associated with per capita crime rate?
If so, explain the relationship.

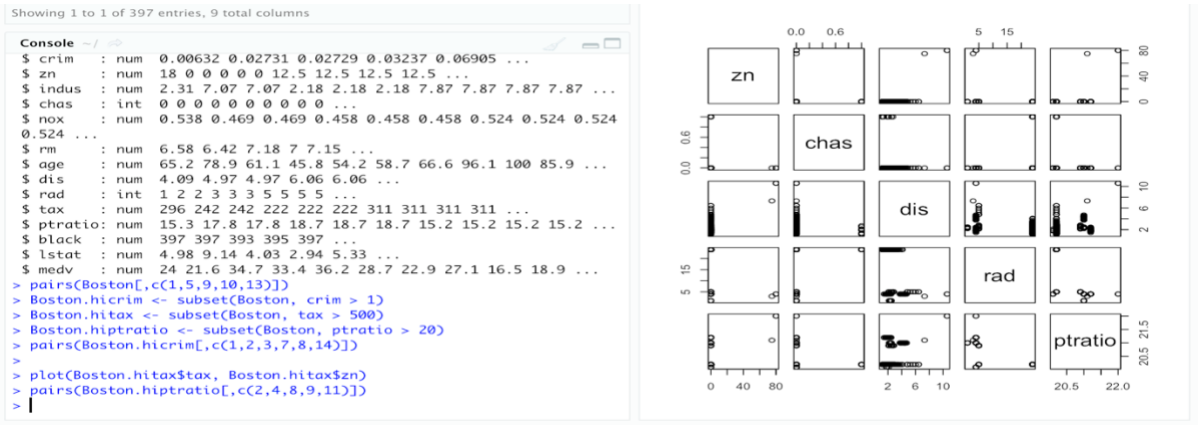
```
69 0.13554  12.5  6.07    0 0.4090 5.594  36.8 6.4980   4 345    18.9 396.90 13.09 17.4
70 0.12816  12.5  6.07    0 0.4090 5.885  33.0 6.4980   4 345    18.9 396.90  8.79 20.9
71 0.08826   0.0 10.81    0 0.4130 6.417   6.6 5.2873   4 305    19.2 383.73  6.72 24.2
 [ reached 'max' / getOption("max.print") -- omitted 435 rows ]
> ?Boston
> dim(Boston)
[1] 506  14
> pairs(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 ...
> pairs(Boston[,c(1,5,9,10,13)])
>
```



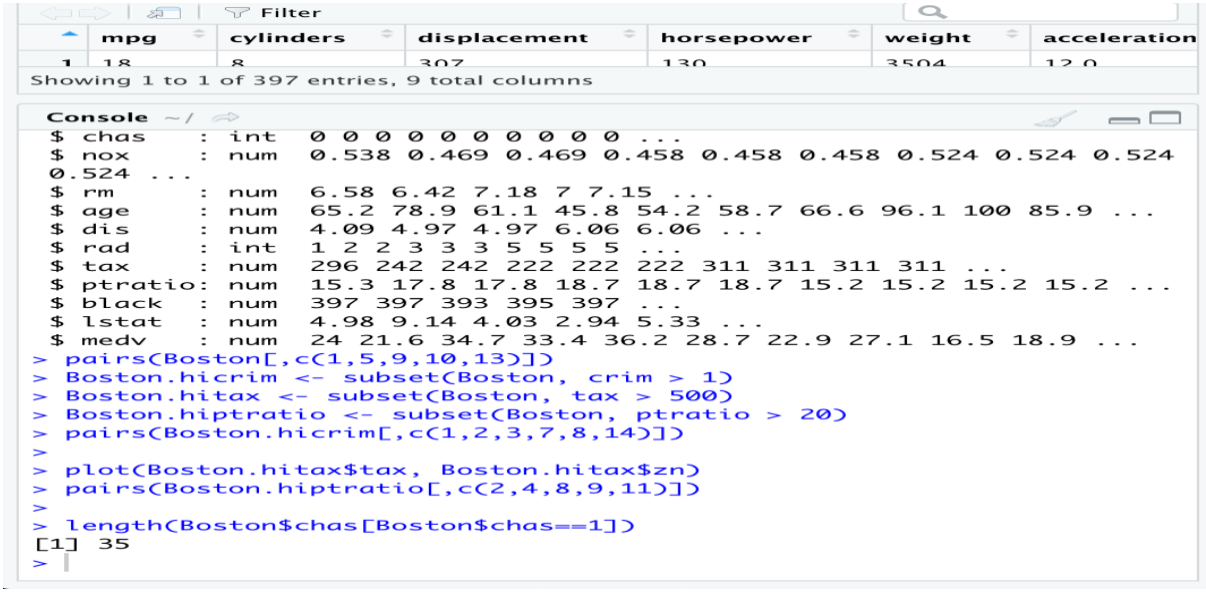
(d) Do any of the suburbs of Boston appear to have particularly high crime rates? Tax rates? Pupil-teacher ratios? Comment on the range of each predictor.







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



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

```
> length(Boston$chas[Boston$chas==1])
[1] 35
> 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.

```
> g10 <- rbind(Boston[Boston$medv==min(Boston$medv),],sapply(Boston, ra
nge))
> rownames(g10) <- c("Lowest medv 1", "Lowest medv 2", "Min", "Max")
> g10
                  crim  zn indus chas   nox    rm   age     dis
Lowest medv 1 38.35180   0 18.10    0 0.693 5.453 100.0  1.4896
Lowest medv 2 67.92080   0 18.10    0 0.693 5.683 100.0  1.4254
Min            0.00632   0  0.46    0 0.385 3.561   2.9  1.1296
Max           88.97620 100 27.74    1 0.871 8.780 100.0 12.1265
              rad tax ptratio  black lstat medv
Lowest medv 1  24 666    20.2 396.90 30.59    5
Lowest medv 2  24 666    20.2 384.97 22.98    5
Min             1 187    12.6   0.32  1.73    5
Max            24 711    22.0 396.90 37.97   50
>
```

(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.

```
> h10 <- rbind(sapply(Boston[Boston$rm>8,], range), sapply(Boston, rang
e))
> rownames(h10) <- c("Min rm>8", "Max rm>8", "Boston min", "Boston ma
x")
> h10
               crim  zn indus chas    nox    rm   age     dis rad
Min rm>8    0.02009   0  2.68    0 0.4161 8.034   8.4  1.8010   2
Max rm>8    3.47428  95 19.58    1 0.7180 8.780  93.9  8.9067  24
Boston min  0.00632   0  0.46    0 0.3850 3.561   2.9  1.1296   1
Boston max 88.97620 100 27.74    1 0.8710 8.780 100.0 12.1265  24
           tax ptratio  black lstat medv
Min rm>8   224    13.0 354.55  2.47 21.9
Max rm>8   666    20.2 396.90  7.44 50.0
Boston min 187    12.6   0.32  1.73  5.0
Boston max 711    22.0 396.90 37.97 50.0
>
```