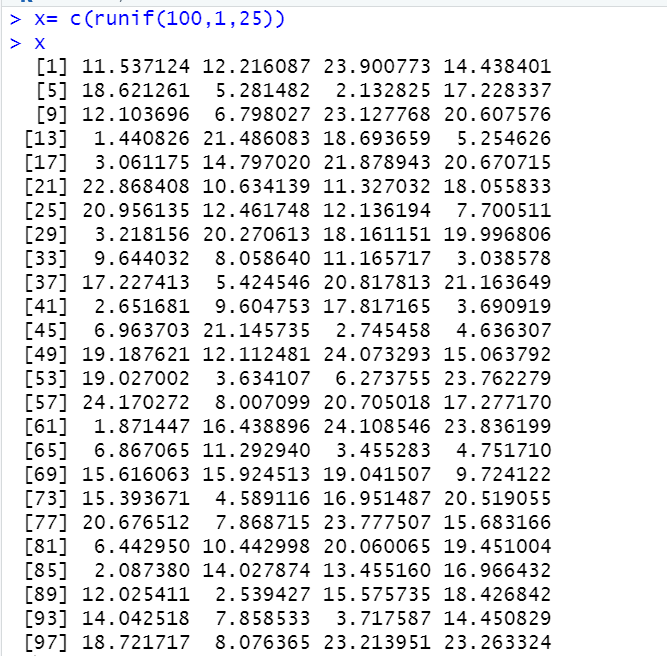
1. Vectors & Matrices

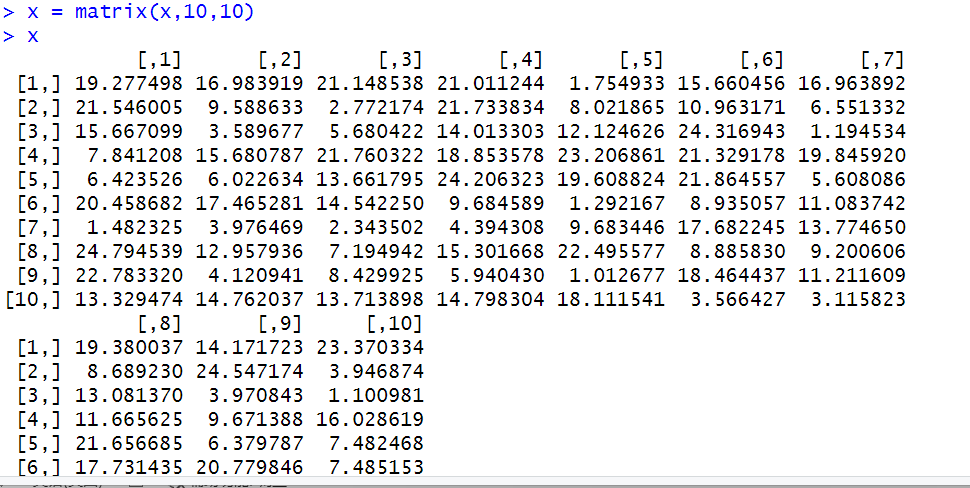
a. Create a vector with 100 elements with randomly generated real numbers ranging

from 1 to 25

(Hint: use runif() function).



b. Reshape this vector into a 10 by 10 matrix reading by column.

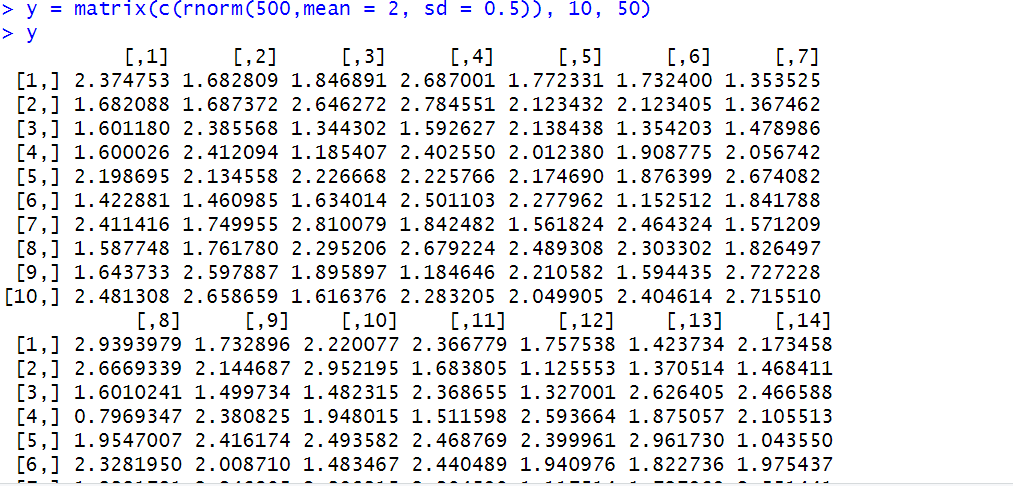


c. Similarly, generate another vector with 500 elements of standard normal distribution

(with mean = 2 and sd =0.5) and plot its histogram. Reshape this into a 10 by 50

matrix reading by row.

(Hint: use matrix() and rnorm() functions)



2. This exercise relates to the College data set, which can be found in the file

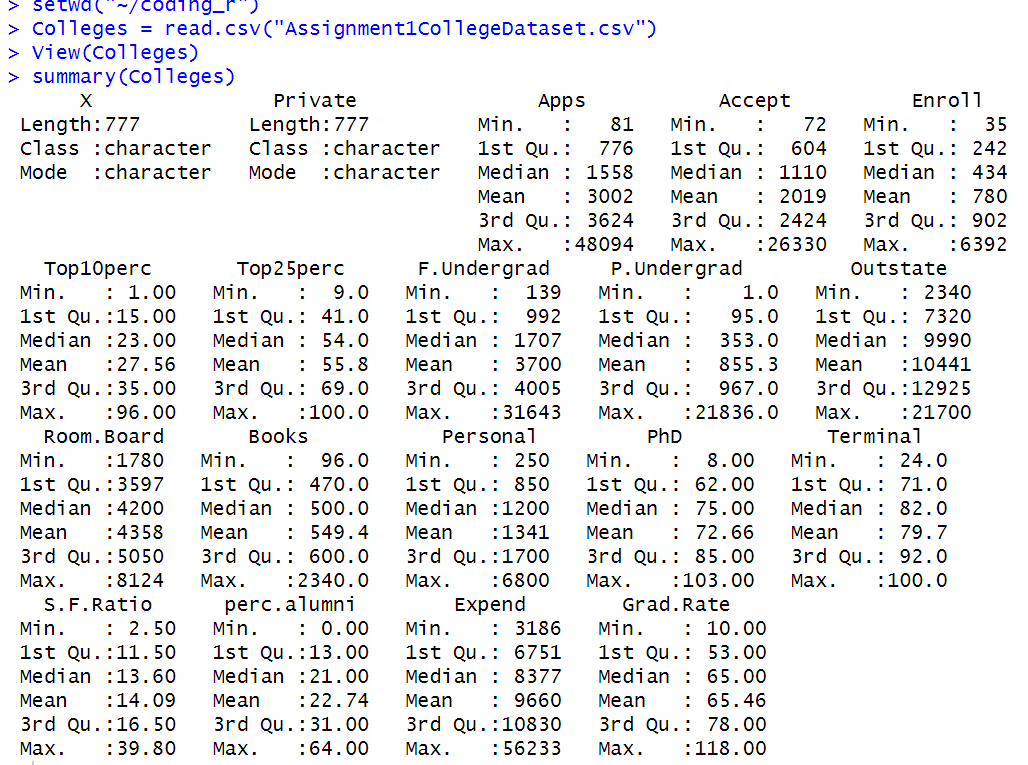
College.csv. It contains a number of variables for 777 different universities and colleges

in the US. 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.

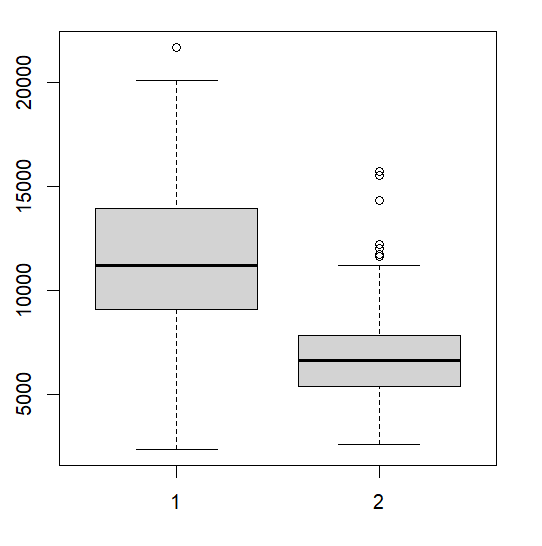
a. Use the summary() function to produce a numerical summary of the variables in

the data set.



b. Create a boxplot of variable "Outstate" as a function of "Private" (Yes for Private; No

for Public). Explain your interpretation of the boxplots.



So, in the plot, number 1 is outstate-student number vs. private school. The average of outstate-student number in private school is about 12000. And in most private school, the outstate students’ number is around 9000 to 14000. The limit of that is 2000 to 21000. There is only 1 outlier(school) which outstate student is over 20000.

However, number 2 is outstate-student number vs. public school. The average of outstate-student number in public school is about 6500. And in most public school, the outstate students’ number is around 9000 to 14000. The limit of that is 2000 to 12000. There are some outliers (schools) which outstate-student is over 12000.

In general, the average number of private outstate students is larger than the number of public school outstate students.

3. Load the Auto data set, which is in the ISLR library. Understand information about this

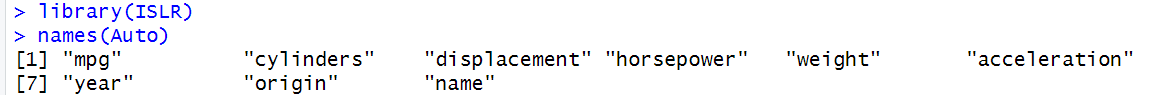
data set by either ways we introduced in class (like “?Auto” and names(Auto))

a. Make a scatterplot between cylinders and mpg. Draw pairwise scatterplot between

“mpg”, “displacement”, “horsepower”, “weight”, “acceleration” (try to plot all

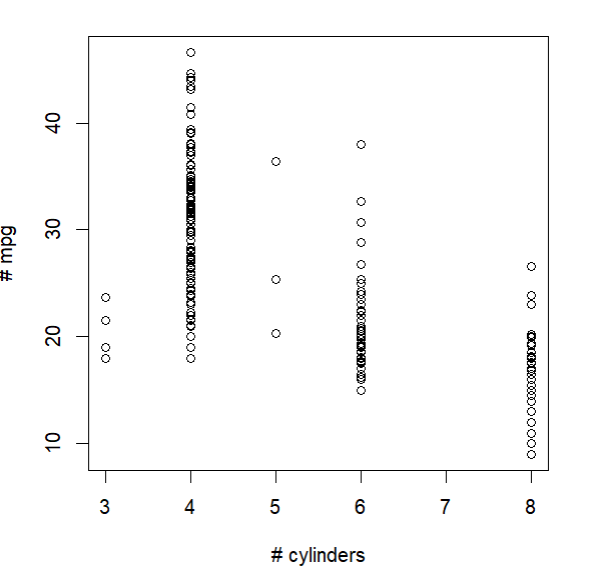
scatterplots in one figure; hint: use pairs() command). By observing the plots, do

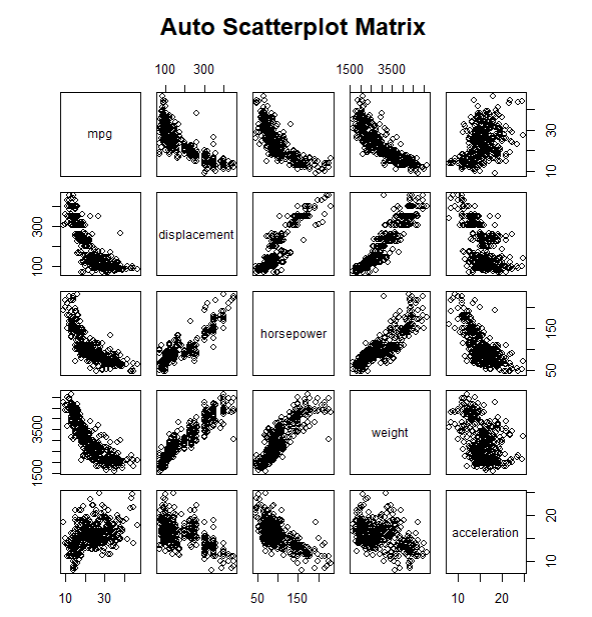
you think the two variables in each scatterplot are correlated? If so, how?











So, for mgp in this plot, it has negative trend relationship with displacement, horsepower and weight. But mgp does not have a very obvious relationship with acceleration.

For displacement, it has positive trend relationship with horsepower and weight. But displacement has low negative trend relationship with acceleration.

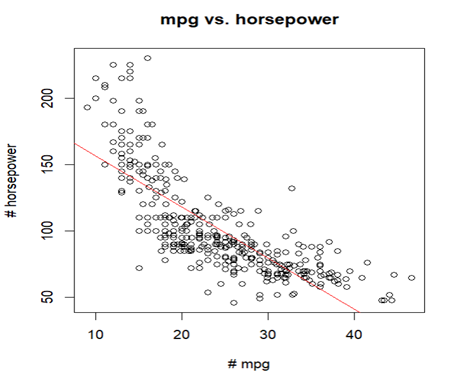
For horsepower, it has strong positive trend relationship with weight and a strong negative relationship with acceleration.

For weight, it dose not have a very obvious relationship with acceleration.

b. Create a scatterplot between mpg and horsepower. Draw a straight line on the

scatterplot to represent relationship between the two variables.

(Hint: Search for “R Plot Line in Scatterplot”)



c. Is there a better way to represent their relationship rather than the linear model you

just drew? (No need to use mathematical formula or plotting techniques. Just draw by

hand on the scatterplot. We will learn non-linear fitting approaches later in the

course.)