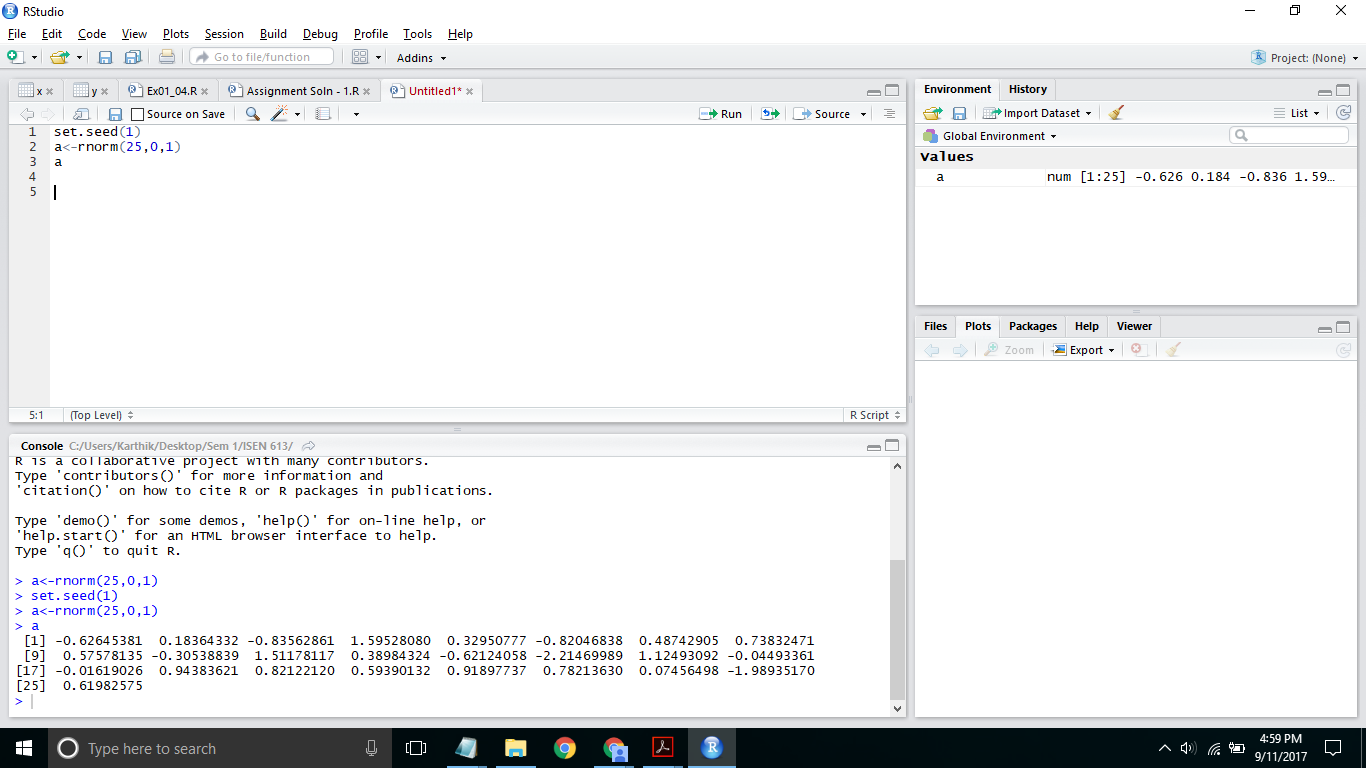
1)

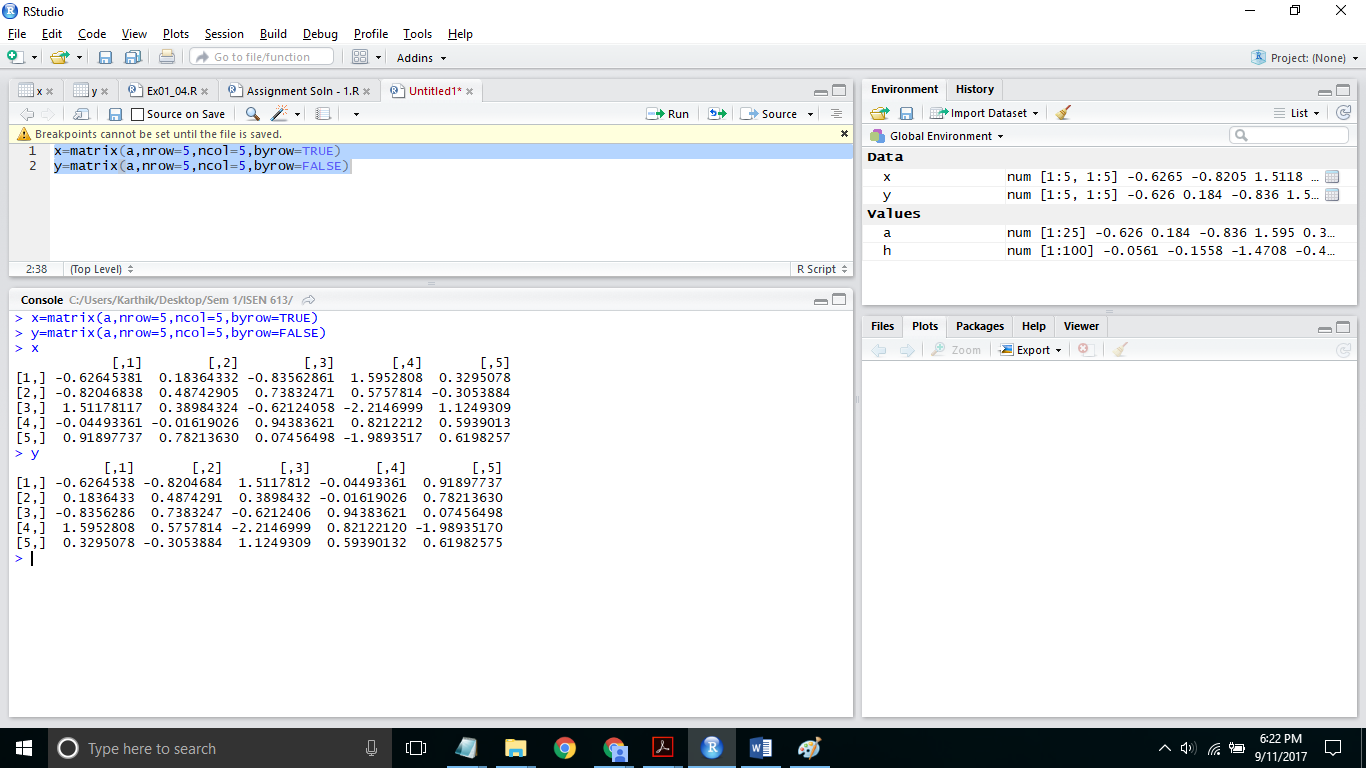
(1)

Generate a vector with 25 elements and each element independently follows a normal distribution (with mean =0 and sd=1);

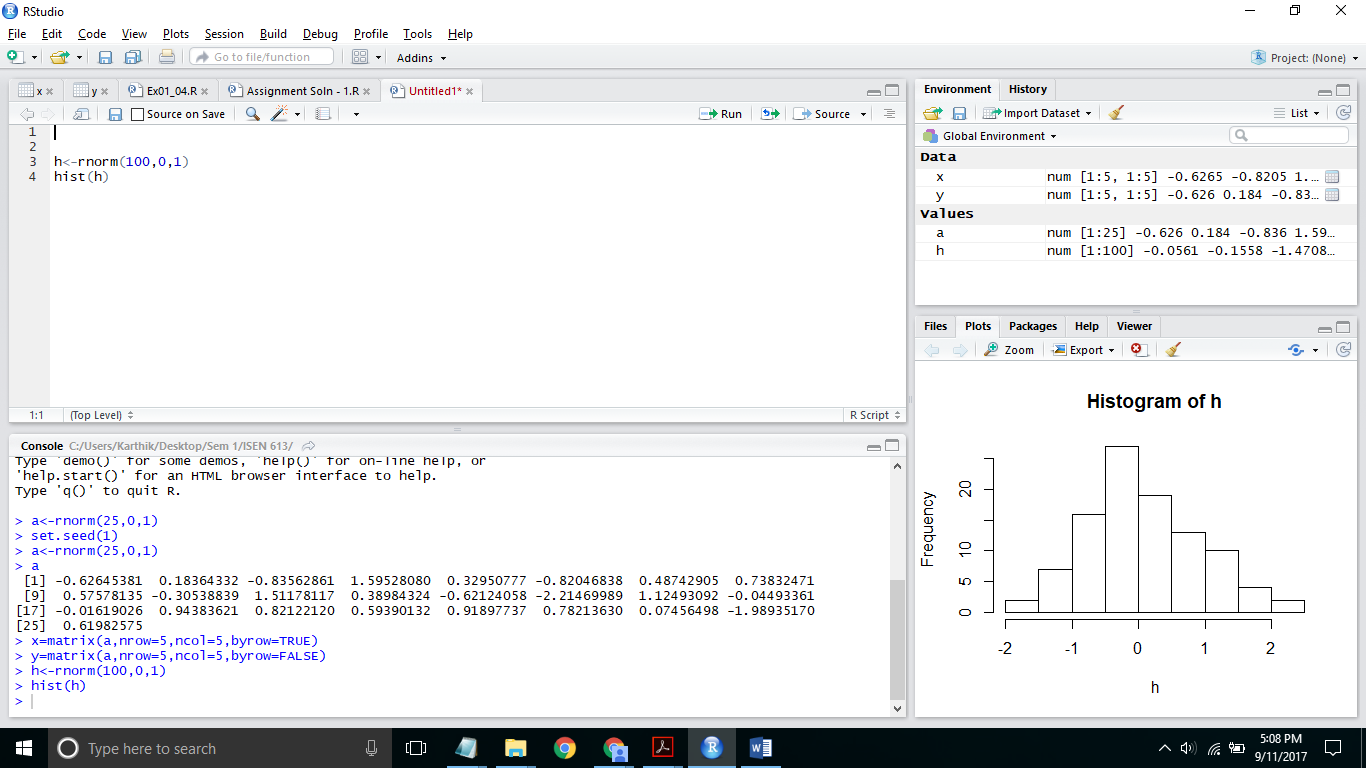


(2)

Reshape this vector into a 5 by 5 matrix in two ways (arranged by row and column);

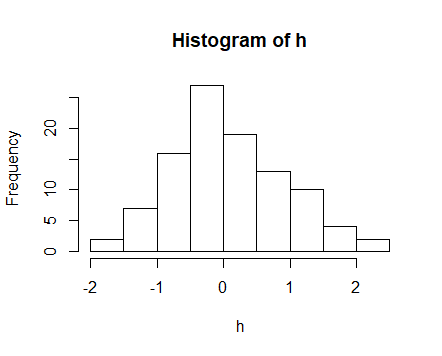


(3)

Similarly, generate another vector with 100 elements and plot its histogram. 

(4)

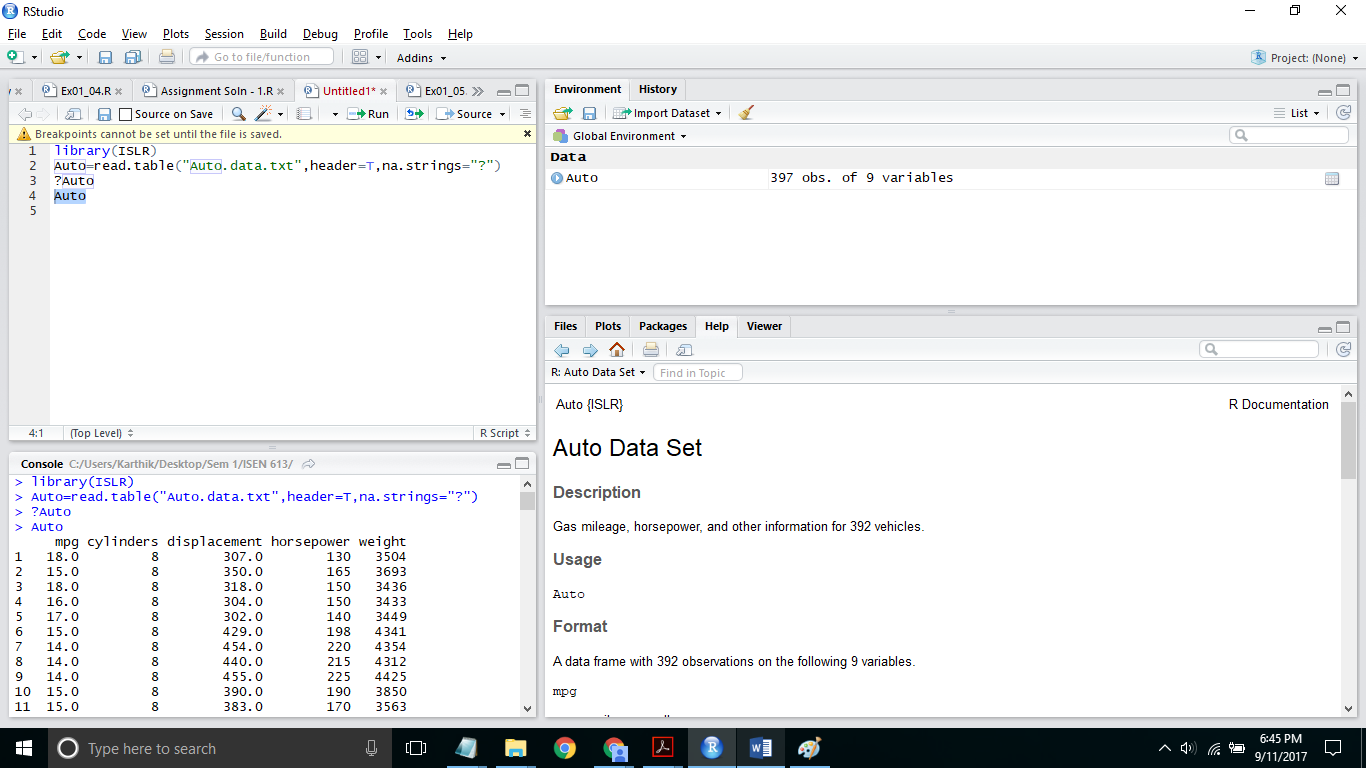
Provide screenshots of the R code used for the above questions as well as the plots in the report. Explain the plots in your own words.

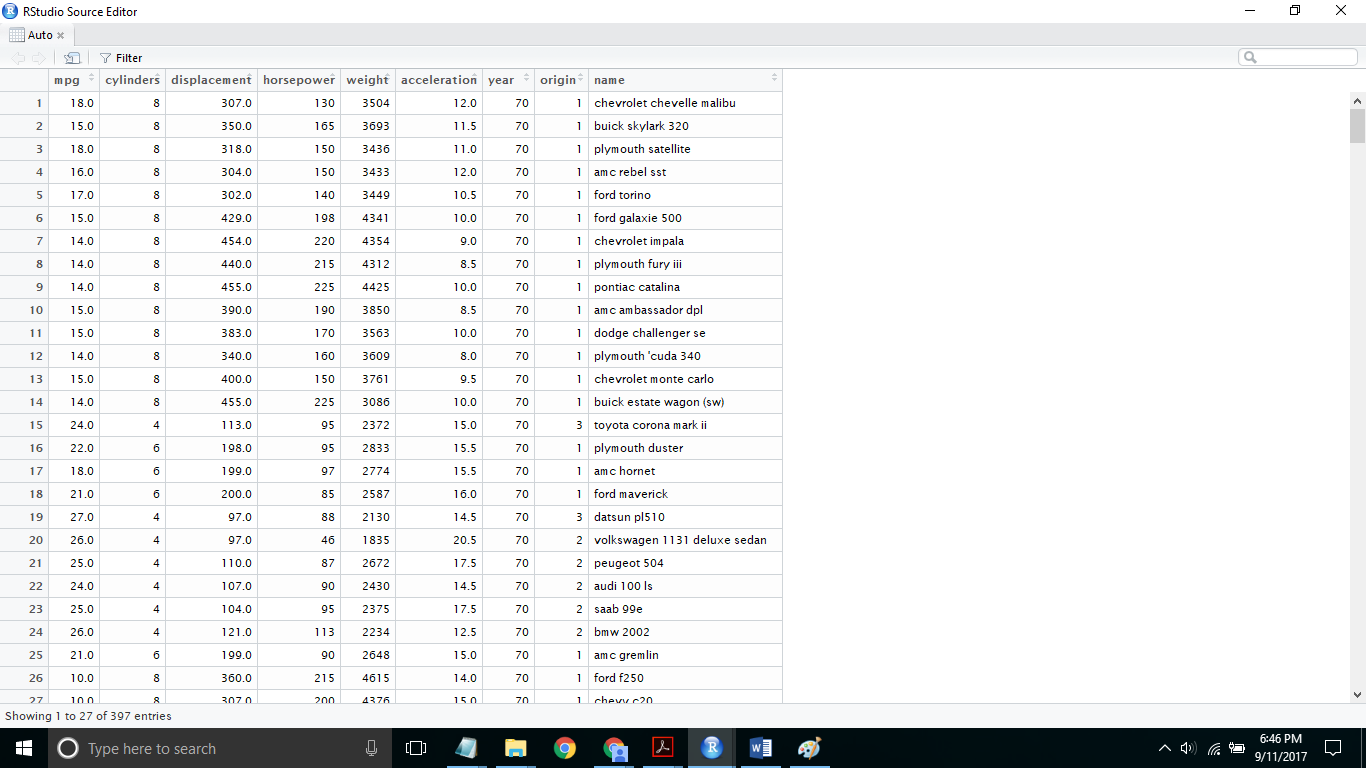


The frequency distribution of the random variables seems to follow a near Normal Distribution with mean at 0. Maximum number of observations or the frequency is around 0 as expected and reduces as one moves away from the mean which is 0 on either direction.

2)

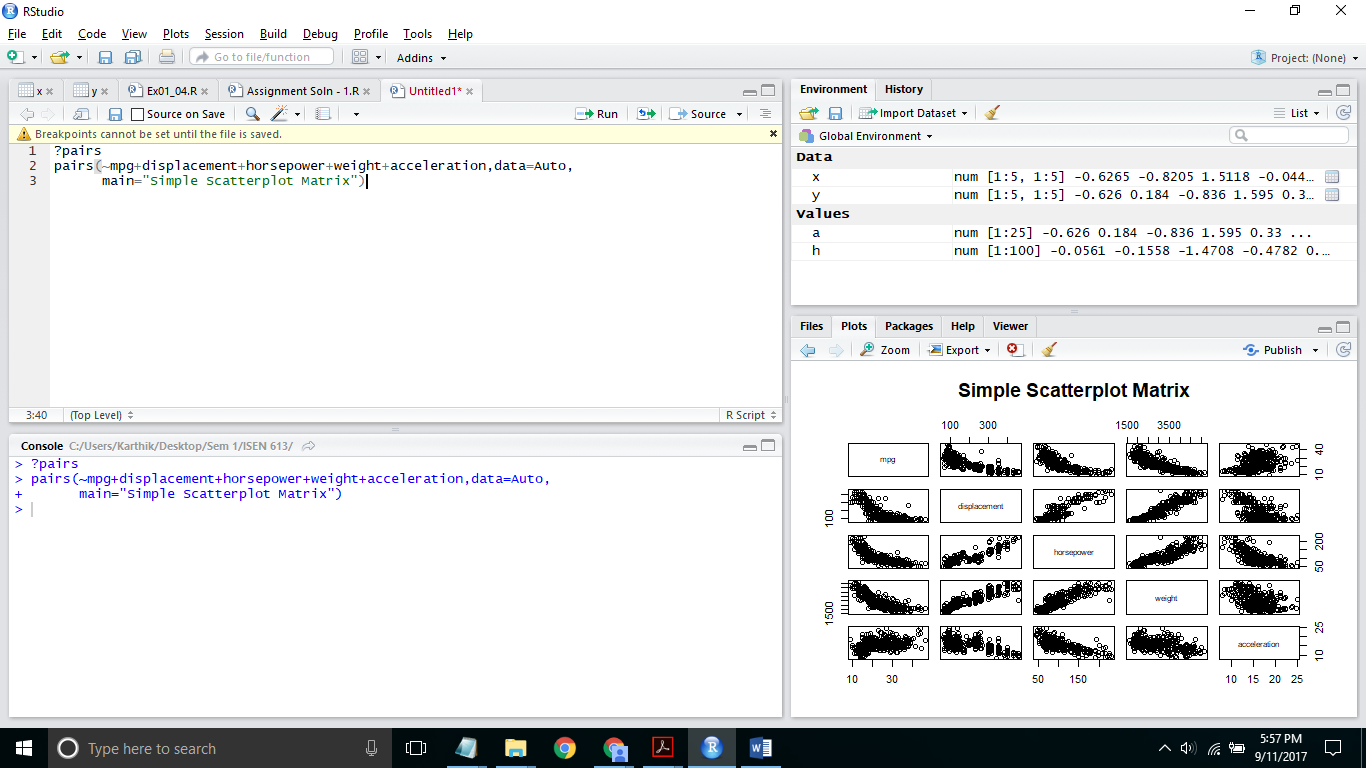
Upload the Auto data set, which is in the ISLR library. Understand information about this data set by either ways we introduced in class (like “?Auto” and names(Auto))

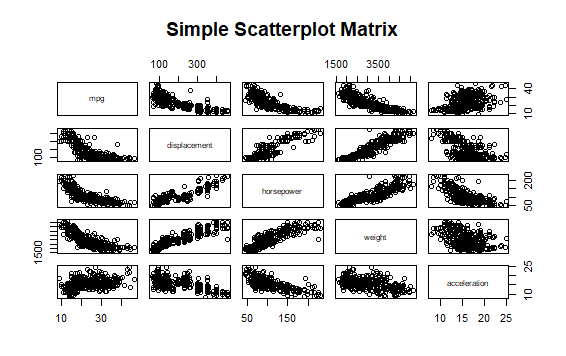




3)

Make a scatterplot between every pair of the following variables (try to plot all scatterplots in one figure; hint: use pairs() command): “mpg”, “displacement”, “horsepower”, “weight”, “acceleration”. By observing the plots, do you think the two variables in each scatterplot are ***correlated***? If so, how?





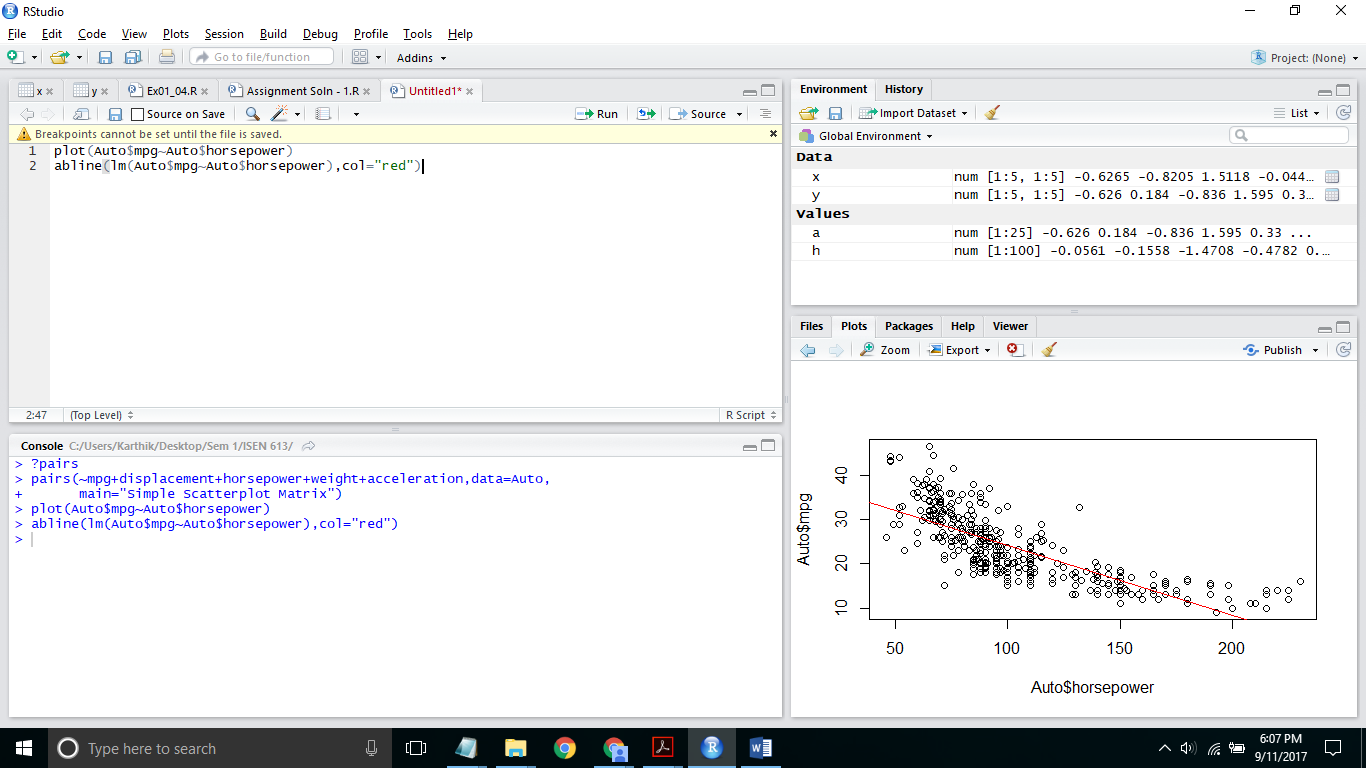
All the set of variables are approximately correlated to each other as follows either directly or inversely. Directly correlated because with the increase of one variable the other variable also increases.

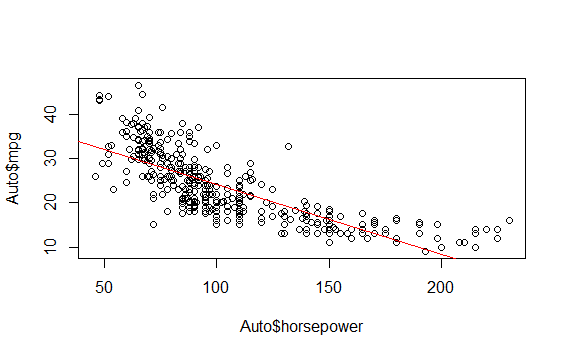
Inversely correlated because with the increase of one variable the other variable decreases.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | mpg | displacement | horsepower | weight | acceleration |
| Mpg | - | inversely | inversely | inversely | directly |
| Displacement | inversely | - | directly | directly | inversely |
| Horsepower | inversely | directly | - | directly | inversely |
| Weight | inversely | directly | directly | - | inversely |
| acceleration | directly | inversely | inversely | inversely | - |

4)

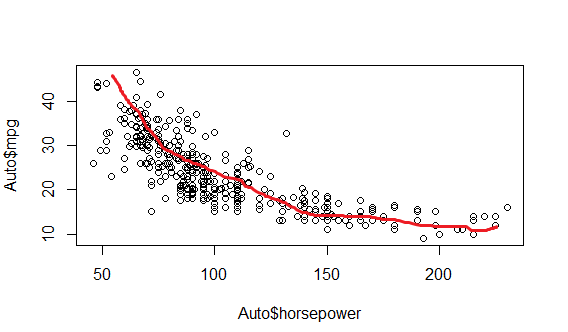
Draw a line on the scatterplot of mpg vs. horsepower to represent relationship between the two variables.





5)

Is there a better way to represent their relationship rather than the linear model you just drew? (No need to use mathematical formula. Just draw something on the figure)



It can be represented by a non-linear function to better represent the data with less variance and standard deviation.