#### STA 141A Homework 1 Haozhe Gu 999200555

**The codes and results derived by using these codes constitute my own work. I have consulted the following resources regarding this assignment: NONE**

1. For the first number of markov chain, I used 0.5 chance to produce 0 and 0.5 chance to produce 1. They, I just use for loop to iterate until the length meet the requirement. Input and return values are specified in the code.
2. I used a sapply loop to traverse through the whole sequence, recording position when see a run and end of run. I rbind the two arrays for runs of 0s and 1s, the first row is for 0s and the 2nd is for 1s. Array with fewer elements will be filled with -1 in order to have same dimension to be binded. Input and return values are specified in the code.
3. First part, I use sapply to find the starting points of all start sequence and end sequence. Then, by comparing their starting point, length, I can judge if a subsequence exists. Input and return values are specified in the code.
4. Enter Q4() to see the summary statistics.
5. I ) As the length of runs increase, the occurrence of both 0’s and 1’s run decrease as ecpected. Also, when the length of runs increase, after running the script for several times, it seems like there’s a constant ratio between occurrence of 0’s runs and 1’s runs.

ii) The counts of the occurrence of subsequences with starting motifs “0000” and end motif “1111” is around 40.

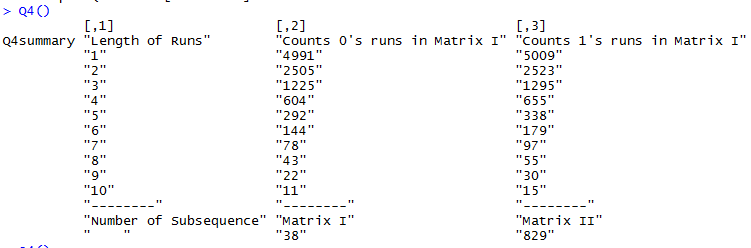
1. I) As the length of runs increase, the occurrence of both 0’s and 1’s run decrease as ecpected. Also, when the length of runs increase, after running the script for several times, it seems like there’s a constant ratio between occurrence of 0’s runs and 1’s runs.

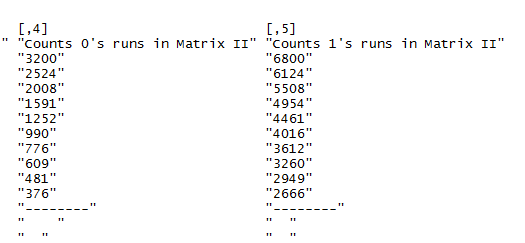
ii) The counts of the occurrence of subsequences with starting motifs “0000” and end motif “1111” is around 800.

1. From the Q4 Summary Statistics., Matrix A generally have less short length( runs than Matrix B, this is because 0.5 0.5 because it is easy to generate a different number that ruin the run.

The number of subsequence with specified starting and ending motifs for Matrix B are much more than that of Matrix A. It is reasonable because Matrix B have far more runs of “0000” and “1111” than Matrix A does. Both have approximately normal distributed location (too many data, so I didn’t put the figures on report)

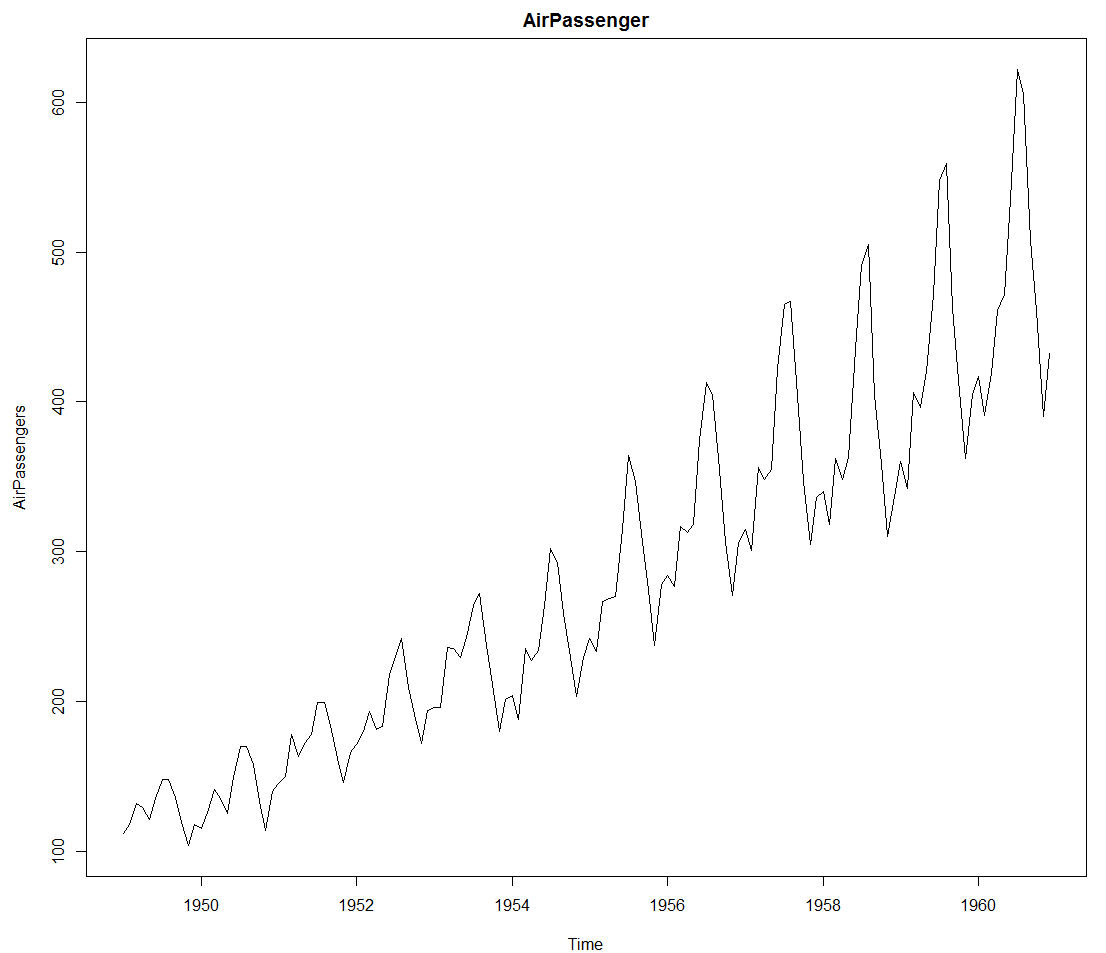
There seems to be a constant ratio between 0’s runs and 1’s run for each Matrix after testing for several times. It seems that the probability of generating 0 or 1 is regardless of the number before it when n is close to infinity.





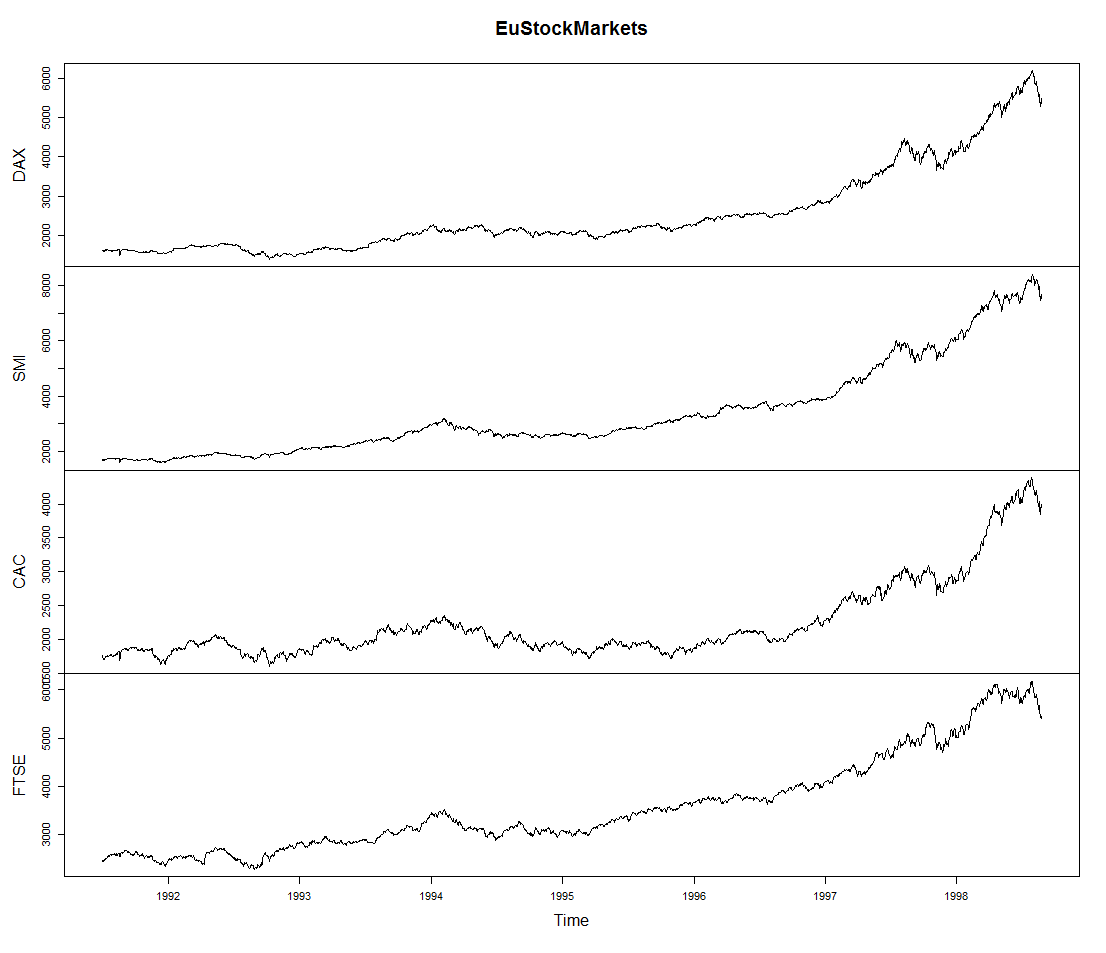
5, Graphs can be produced by Q5()

1. AirPassengers



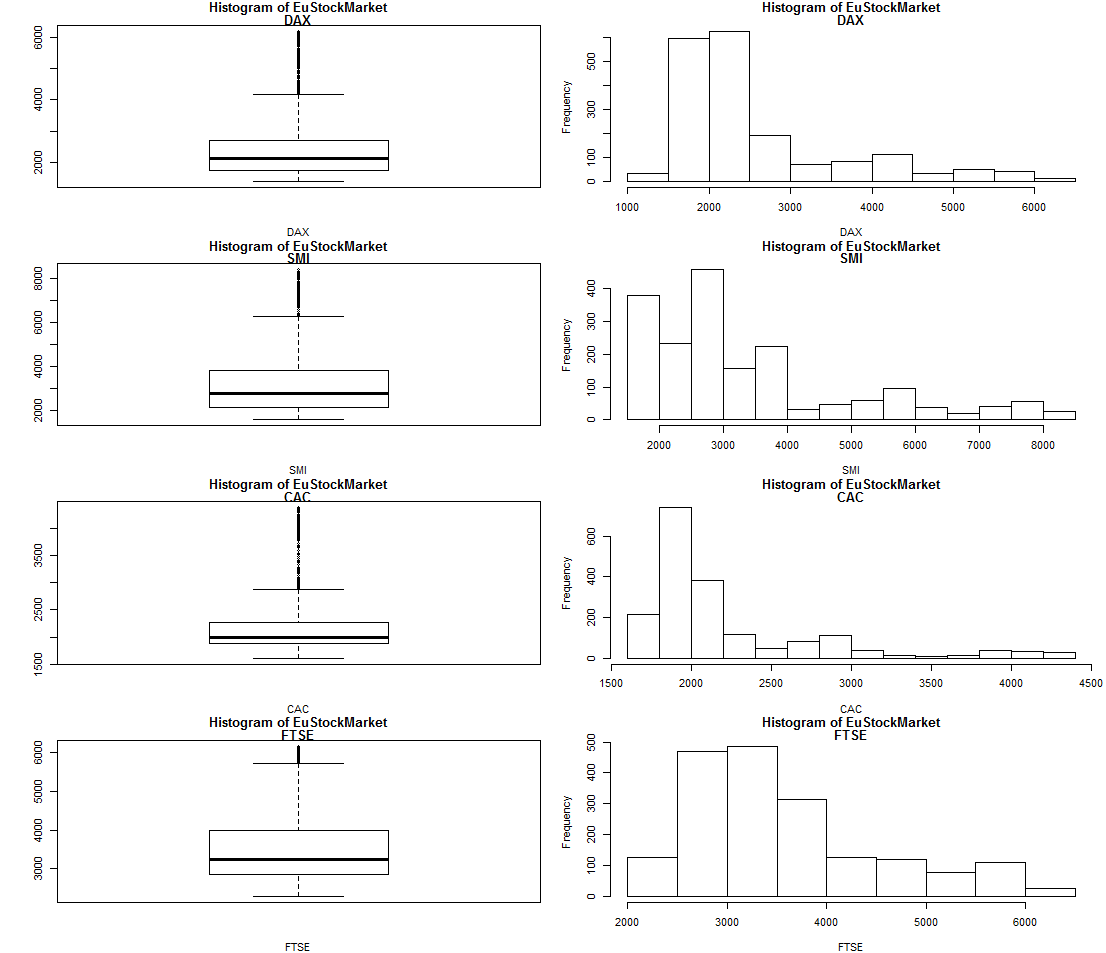
From the graph, there is clearly a time series in the dataset. The Seasonal component is quite obvious to have a frequency of 6 times per year. This is exactly saying that there is a two-monthly difference in air passengers number. From the graph, it seems like the overall trend of the number of air passengers are increasing as time advance. The number of air passengers reach its peak during July, August where people tends to have a travel plan. It decreases sharply after August where people are backed to their work.

1. EuStockMarkets

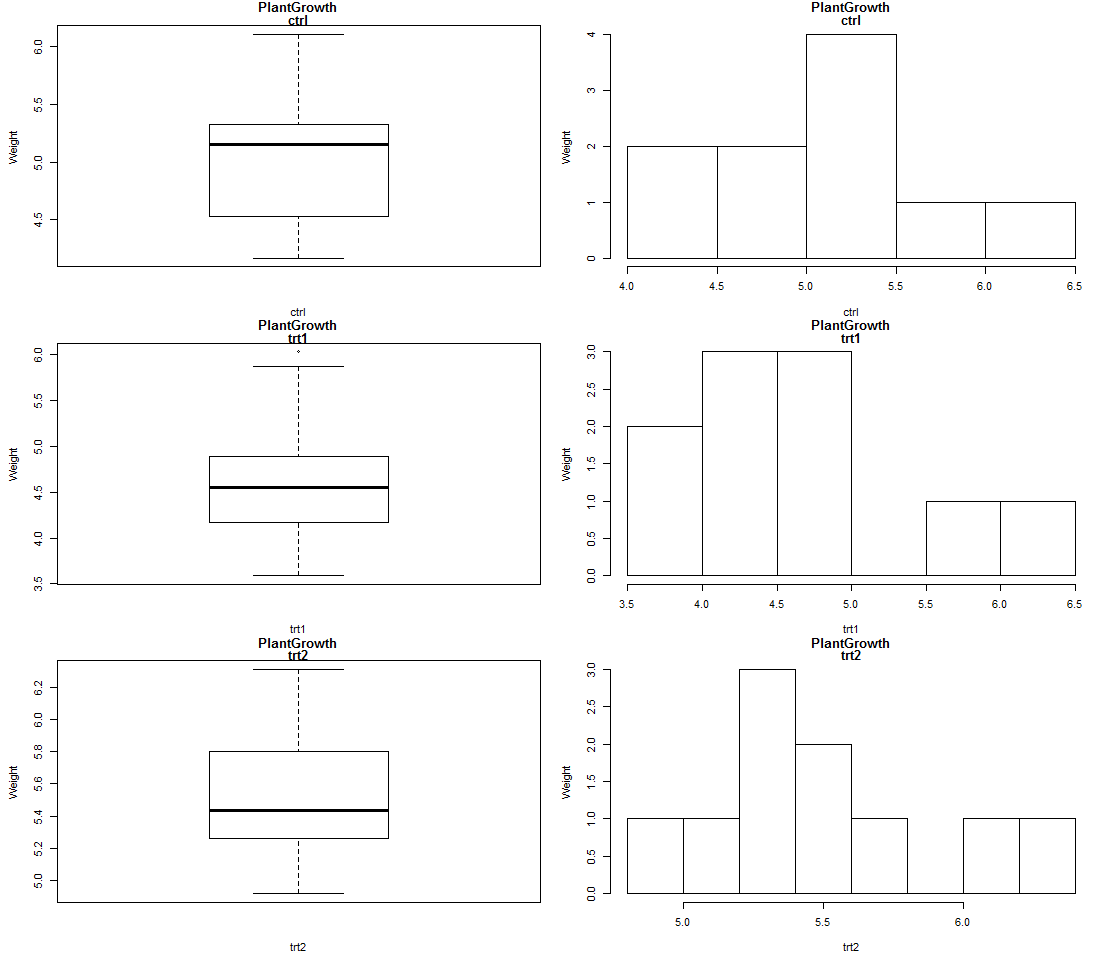
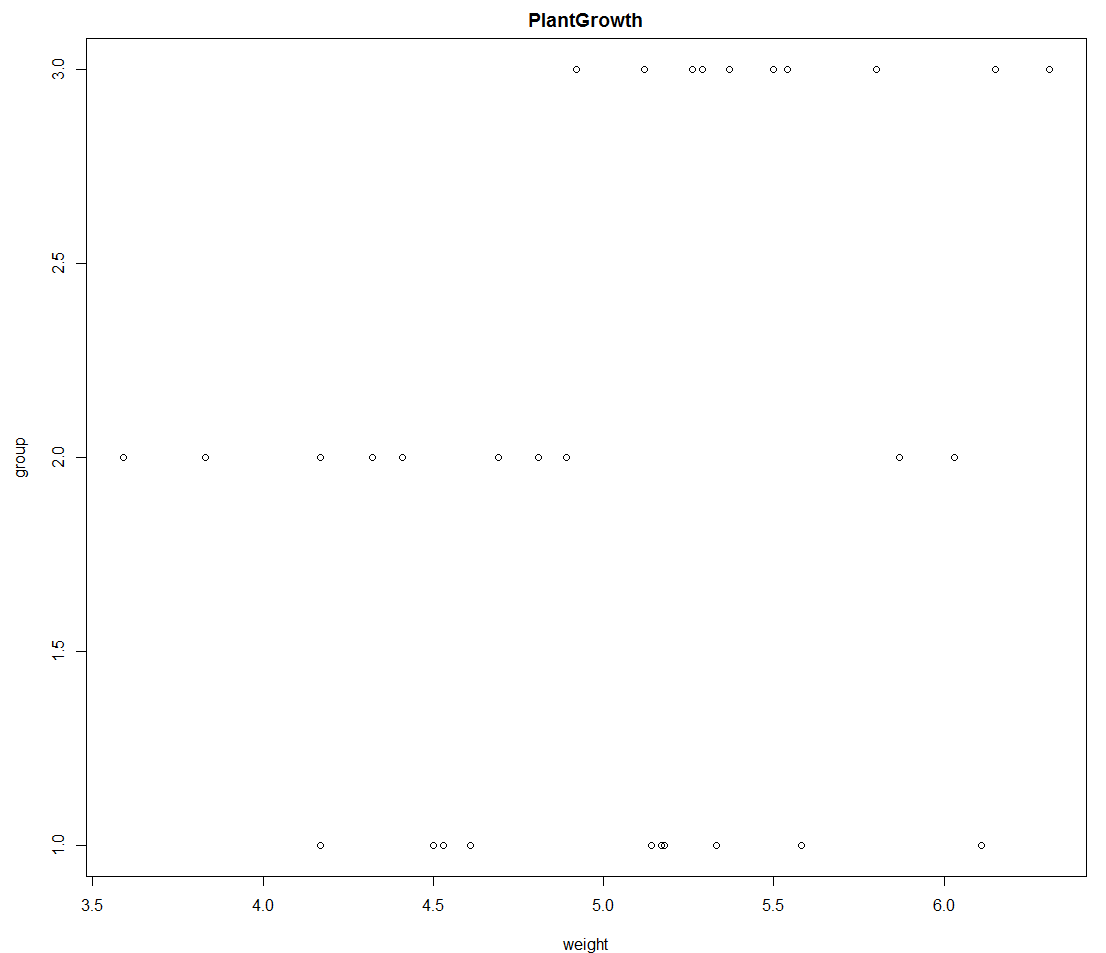


From the plot above, the general trend in the major European stock indices is increasing with time advance. All four plot seems to have similar shape where sudden increase and decline are happened at the same time. This might be that all four stock indices were influenced by the same economic growth boosts or depressions.

The plots below are the boxplots and histograms for single stock indices. From these plots, we know that FTSE are generally higher than all other indices which has a mean around 3000. All other indices have their mean around 2000. CAC has the lowest mean value among all four indices.

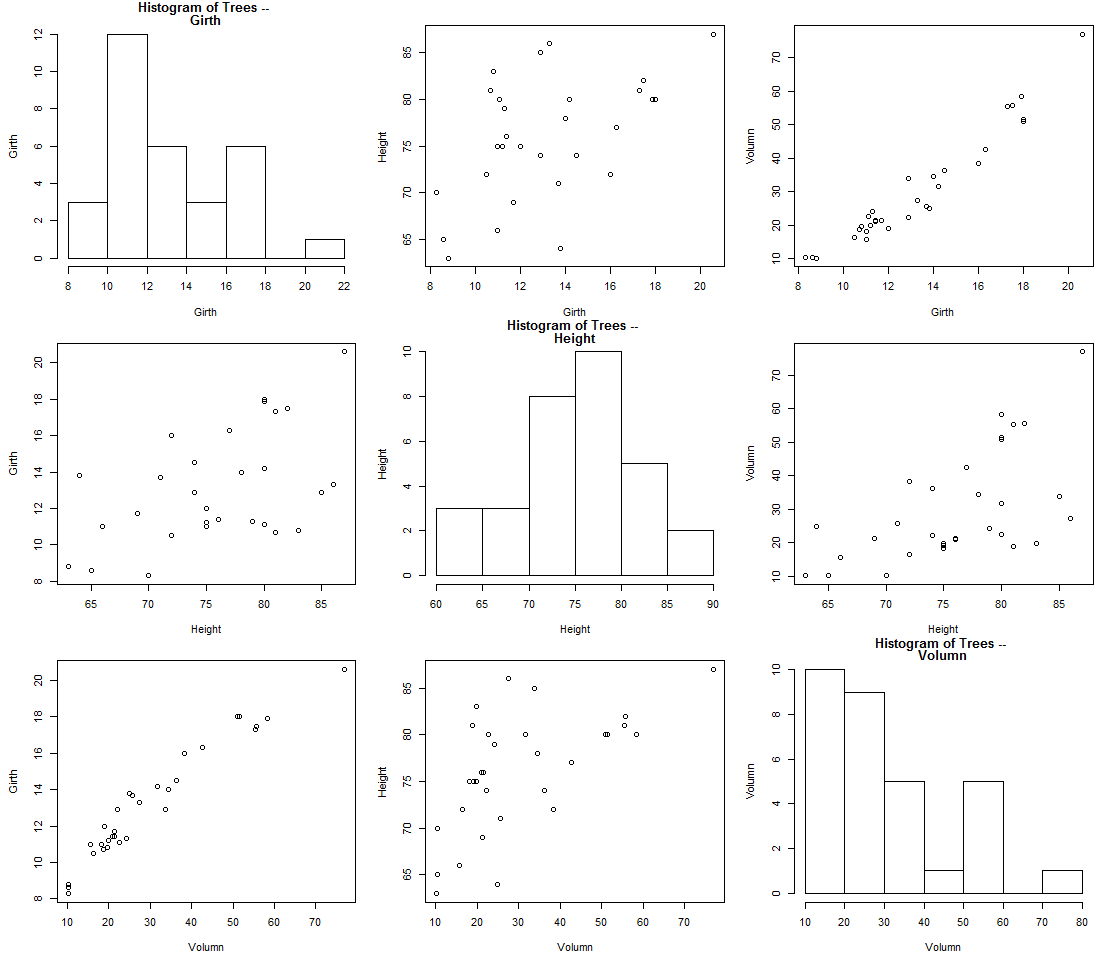


1. PlantGrowth



From the boxplot, it is clear that the group with treatment two have the highest mean weight at about 5.5 while the group treatment one have the lowest mean weight at about 4.5. Also, treatment two group have the highest weight at about 6.4.

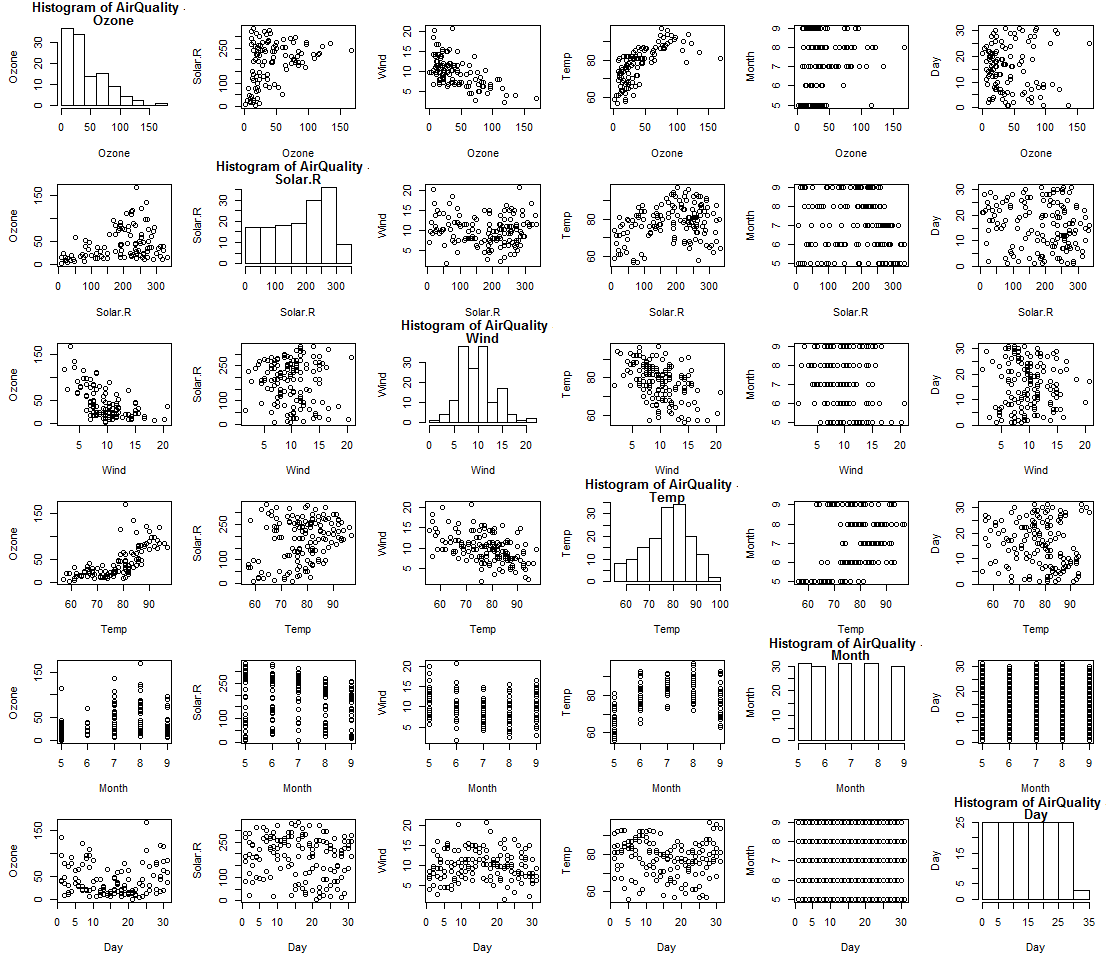
1. Trees



The graph above are 3\*3 pairwise comparisons for three variables. From the graph (1,3), we observe a clear linear trend in Girth of the tree and the Volume of the tree. This is straight forward; Girth is equal to 2\*PI\*Radius where Volume is also proportional to the Radius. Then, the graph of Height and volume no-doubly shows certain linear relation.

1. Airquality

The following graph shows inter-relationship between factors of Air quality. From graph (1,4), there might be a linear relationship between temperature and Ozone. Moreover, from graph (1,3), there might also be a negative linear relationship between wind and ozone. This make wind and temperature negatively related. There is also information like the mode of temperature throughout the year is about 80, August seems to be hottest month, etc.



### Appendix

