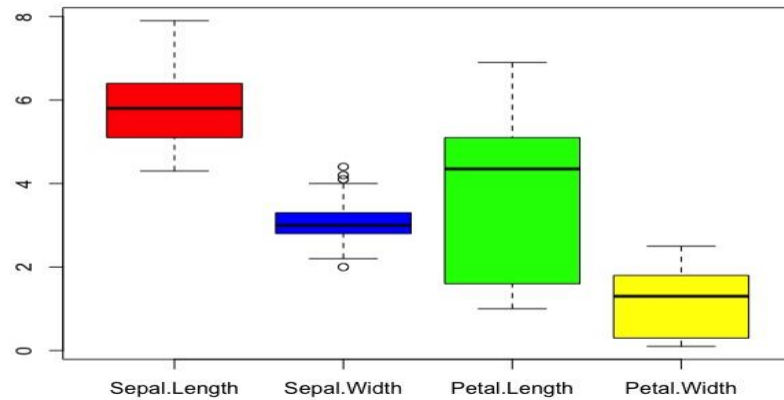


1. Compute standard deviation of each attribute. Compute boxplots for each attribute and interpret the results. Which attributes have the most variation. Are there any outliers in any of the attributes?

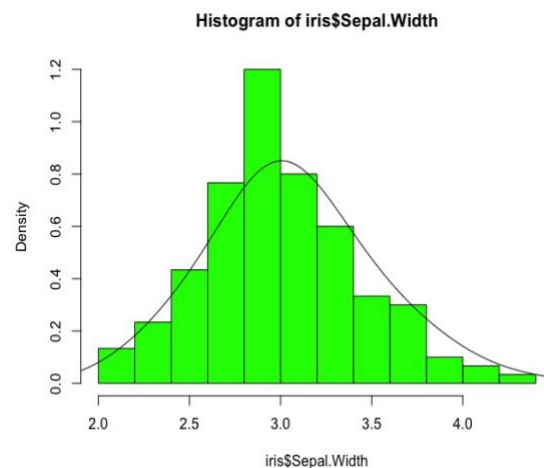
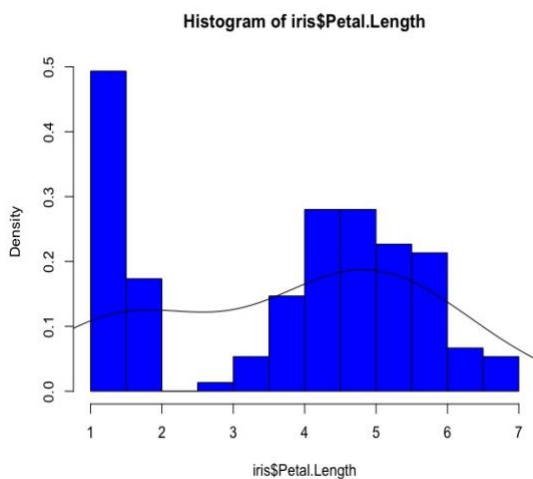
Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
0.8280661	0.4358663	1.765298	0.7622377

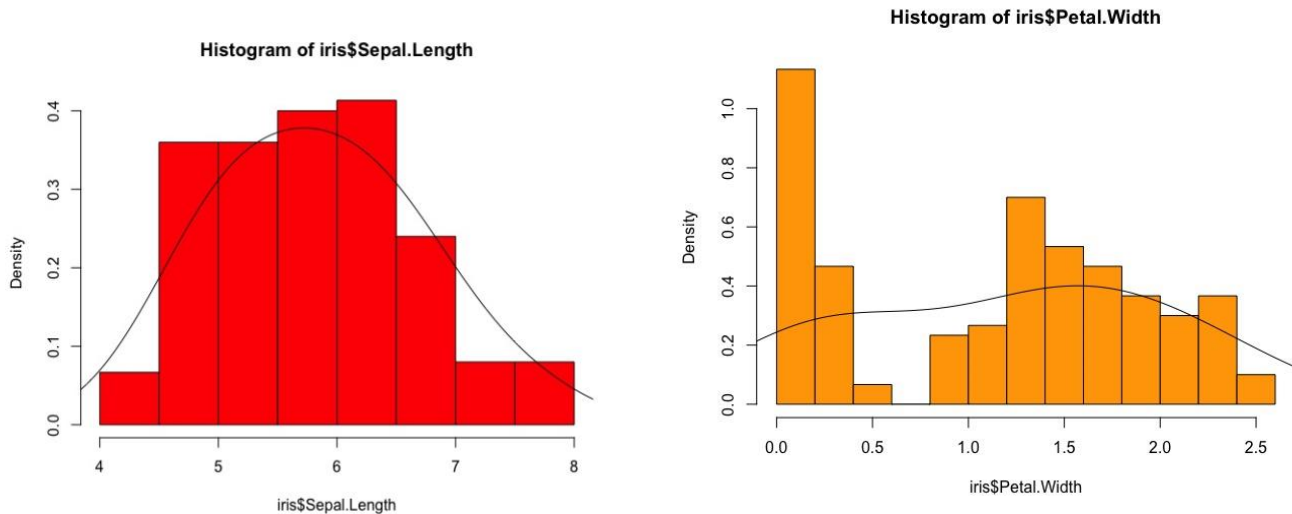


	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Min.	4.3	2	1	0.1
1st	5.1	2.8	1.6	0.3
Median	5.8	3	4.35	1.3
Mean	5.843	3.057	3.758	1.199
3rd	6.4	3.3	5.1	1.8
Max.	7.9	4.4	6.9	2.5

If we see the boxplot, the Petal.Length attribute is most variation. It is value mentioning that Sepal.Width contains three outliers, two on the high end of the values and a single outlier on the low end. This is because the majority of the data for Sepal.Width is concentrated around the center of the data.

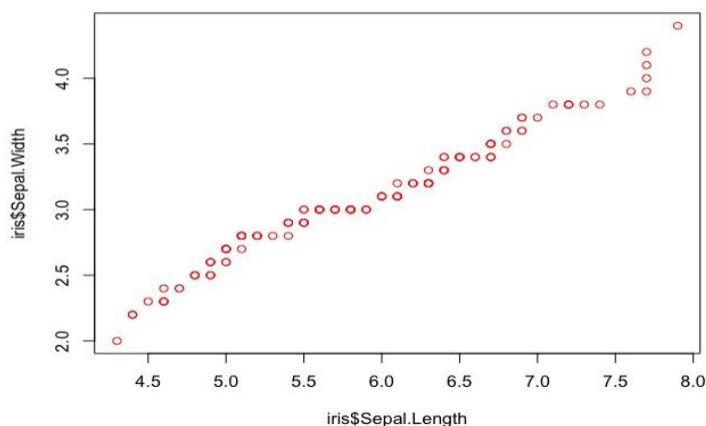
2. Compute histogram for each attribute. Based on histogram and boxplot of each attribute, which distributions look similar to a normal bell curve, which ones look different?





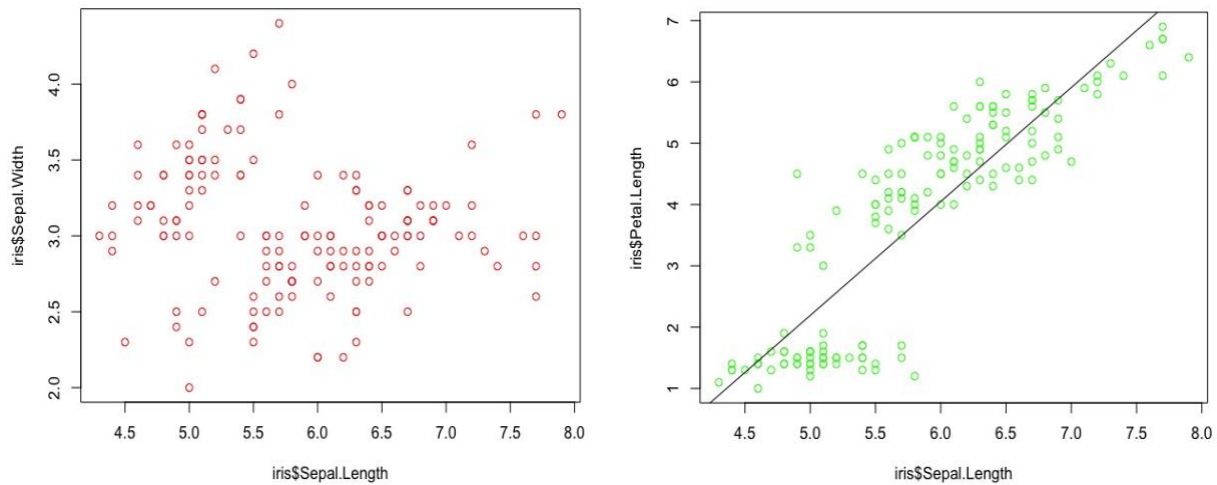
Histogram data given above, Sepal.Length and Sepal.Width appear to be distributed normally. Sepal.Width is clearly normal, while the Sepal.Length is less clear. From the boxplot, it can also say that the average value for Sepal.Length and Sepal.Width are located in the center of the box, which means that the data between Q1 and Q3 is equally distributed between median. Petal.Length and Petal.Width have a large number of data points of the histogram, indicating that there are less significant values in the data. The large volume of low-value data causes a low Q1 number, so the medians are far from Q2 and as shown in the boxplot for Q3.

3. Compute QQ plot between Sepal.Length and Sepal.Width and interpret the results with comparison to boxplots and histogram of these attributes.



In QQ plot, we can see the all data points forming in a line that's roughly straight, so the data distributions Sepal.Length and Sepal.Width are same. In a boxplot, these two attributes are the median value that located around the center of the box, which indicated Q1 and Q2. The histogram of these two attributes can be used to determine that the data is normally distributed. The QQ plot can only say that the distribution of these two attributes are the same, and they are usually distributed or not.

4. Compute scatterplots between Sepal.Length and Sepal.Width, and Sepal.Length and Petal.Length. Interpret the results. Is there any correlation? Is it positive or negative?



Sepal.Length and Sepal.Width does not seem to be associated significantly. Petal.Length and Sepal.Length does seem to be associated significantly, because the indication found by the line added to the graph. The two clustering of data points is visible in Sepal.Length and Petal.Length.

### References:

- <http://r4ds.had.co.nz>
- Scatter Plot. (n.d.). Retrieved January 31, 2018, from <http://www.r-tutor.com/elementary-statistics/quantitative-data/scatter-plot>
- Vidani, P. (2015, June 13). Data \* Science R. Retrieved January 31, 2018, from <http://things-about-r.tumblr.com/>
- Molecular Organisation and Assembly in Cells. (n.d.). Retrieved January 31, 2018, from [https://warwick.ac.uk/fac/sci/moac/people/students/peter\\_cock/r/iris\\_plots/](https://warwick.ac.uk/fac/sci/moac/people/students/peter_cock/r/iris_plots/)
- Ford, C. (2015, August 26). University of Virginia Library Sites. Retrieved January 31, 2018, from <http://data.library.virginia.edu/understanding-q-q-plots/>
- Standard Deviation. (n.d.). Retrieved January 31, 2018, from <http://www.r-tutor.com/elementary-statistics/numerical-measures/standard-deviation>
- Visual summaries. (n.d.). Retrieved January 31, 2018, from <http://statweb.stanford.edu/~jtaylo/courses/stats202/visualization.html>