COSC 4610/MSCS 5610 (Spring 2018)

Assignment 1

Due by Friday, Feb 2nd

**Important Submission Instructions**

* Upload your answers in DOC/PDF format to d2l.mu.edu
* No handwritten plots/answers please.
* Make sure to write your name on the document.
* You may use a programming language or a web tool to generate the plots. You must explain how you generated the plots. If you wrote code, include your source code with your submission.

**Questions**

In this assignment, you will analyze [iris](https://archive.ics.uci.edu/ml/datasets/iris) (click the link to open) dataset, which is uploaded in CSV format to D2L. When answering questions, you may ignore the Species column (the last column) in the dataset.

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

**Explain how you generated the results. If you wrote code, include your code below.**