# **Data Visualization [Data 502]**

### HW1

* **Part 1: Written Report**
* **Part 2: Software**
* **Data: Iris**

The purpose of this assignment is to provide you with some experience exploring and analyzing data **without** using an information visualization system. You will also gain some familiarity using various data softwares to import and manipulate data. This assignment **must be completed on your own**.

### Part 1: Written Report (one page max)

You will be using the Iris dataset. You should explore and analyze this data using Excel or simply by hand (drawing pictures is fine), but **do not use any visualization tools**; that means no charts or visual representation of any kind. Your goal here is to perform an exploratory analysis of the data set, better understand the data set and its characteristics, and develop insights about the Iris data.

Your report should consist of three small sections on a single page. First, list (bullet list of items) five "insights", chunks of knowledge, or deeper questions that you either encountered or gained while exploring the data. Even if you come up with more, only list your five best insights. An insight could be some understanding of the data and its characteristics that is not relatively obvious or intuitive. It is something that most people might not realize initially. Note that an insight or knowledge chunk simply may be a deeper question that arose in your mind while exploring the data. And your analysis may not have been sufficient to answer the question. Second, write one paragraph about the process you used to do the exploration and analysis. Did you load the data into Excel, work manually, or do both? What did you do in Excel? Did you draw pictures? Just tell me (briefly) what you did. Third, write one paragraph about challenges or problems that you encountered in doing the analysis this way. Did anything limit or frustrate you? If nothing did, perhaps there was something that was more difficult than you thought it should be. Nothing is perfect, so you should be able to list some potential issues here. So, to sum up, your assignment should have a bullet list of five items followed by two paragraphs.

Grading: I will evaluate the quality of the insights you listed. I am looking for things that I find interesting or perhaps unexpected. This is subjective. For the second and third sections, I will evaluate if you did what the assignment asked. Please proofread your submission before submitting it and make sure it is free of spelling and grammar issues.

### Part 2: Software

Setup RCloud and Tableau for Students. You will write a small code capable of importing the data and containing a small collection of functions that support basic queries. Your system must contain the following functions:

* read.csv() A function that loads the csv data from stored path
* write.csv()A function that writes an R dataset into a csv file into a path of your choice.
* str()A function that gives you the structure of the dataset.
* summary()A function that returns the basic statistics of the variables in a dataset
* describe()A function that returns the basic statistics of the variables in a dataset

You may require installing and loading additional libraries. (Hint\*)

List the structure of the dataset. Explain the difference between the summary and describe functions.

Split/Change the structure of the dataset in R to get descriptive statistics on each of the species types. (Hint\* might want to split or restructure the dataset)

Produce some basic boxplots in R for each of the species and the four measurements (sepal width, sepal length, petal width, petal length)

Import the dataset into Tableau and provide basic visuals within your capabilities.

Grading: Your assignment will be graded on whether it can perform the list of functions correctly. Please use good naming conventions and appropriate comments to help us read your code.

### Data Set

Iris data [available in base R]

The data set should be pretty self-explanatory. We also briefly reviewed it in lecture.