Introduction to R

Christopher Maronga

February 15, 2019

Matrices and Data frames

Matrices are used commonly as the mathematical domain of statistics and they are created using the function `matric()`

`Lets see how these works with examples in R.`, get more help *?matrix*

A **data frame** is the default way of storing data in R and it makes data analysis easier. Generally classifying it, a data frame can be said to be a list of qual-length vectors. This implies that it has rows and coloumns of equal length as well coloumn names.

A data frame is more general than a matrix in that different coloumns can contain different modes of data (character, numeric and so on).

A data frame equals to a **dataset** in other languages and its created using the function `data.frame()`.

`Lets see how these works with examples in R.`, get more help *?data.frame*

Reading data into R

Data can be read into R from different sources including from RDMS, websites via web-scrabing, from web-based databases via Application Programming Interface capabilities and many more..

In this section, we are going to focus on how we can read tabular data into R from text files, excell/csv formats and foreign software outputs such as `STATA` and `SPSS` data formats.

The concept of working directory and file path

1. The working directory is the **specifc folder** within your computer where your live R session is hosted/pointing to. This folders by default holds all your .RData files and .Rhistory files.
    - `getwd()` to get to know where your working directory is located, get help `?getwd`
    - `setwd()` to set a new location for your working directory, get hwlp `?setwd`
2. The **file path** is the specific folder address that contains your files of interest. A file path can be a folder within the working directory or outside the working directory

NOTE: The above concepts are very key when reading data into R, therefore a clear understanding of how they work will come in handy in importing data into R suing any method defined.

Read/Import data into R

- **General way** using the command **`read.table()`**.

You can read a variety of file extensions into R using the function `read.table()`. Some examples of applicable files to read include csv, tab delimited(.txt), space demilited and any flat file so long as it's defined interms of what character separates its coloumns.

NOTE: Knowing the extension of the file you are reading and how its coloumns are separated is essential in using `read.table()` command.

- **More specif way** using the command **`read.csv()`**.

-used to read *comma separated files ONLY* otherwise reffered to as `.csv`. More convinient for csv files in that it pre-populates arguments for you which you would have specified in `read.table()` such as `header=T` or even `sep=","`.

get help by `?read.table` or `read.csv`