R Programming - I

Data Types in R

- Character
- Numeric
- Integer
- Complex
- Logical (T/F)
- Date

Vector

- A Vector in R is a sequence of data elements of the same class
- Vector can be created by the function c()
- When you try to input different classes of data into a vector, R coerces the variables
- One class can also be explicitly coerced into another class

List

- A Vector in R is a special type of vector which can contain elements of different classes
- List can be created by the function list()

Matrices

- A matrix is a collection of data elements arranged in a two-dimensional rectangular layout
- The elements of the matrix should be of the same class
- Matrices are filled column wise by default

Factor

- Factors are used to represent categorical data
- A "factor" is a vector whose elements can take on one of a specific set of values
  - For example, "Sex" will usually take on only the values "M" or "F"
- Factors can be ordered or non-ordered
  - The above example Sex is an example of an un-ordered factor
  - When you consider designation, there is implicit ordering; “Engineer” is at a lower level than “Senior Engineer”

Missing Values

- Missing values can be found using the command is.na()
- To check whether an element is a number the command is.nan() is used
- Infinity denoted by Inf is considered as a number by R

Data Frame

- Data Frame is used to store data tables
- It is a list of vectors of equal length
- Can be created using data.frame() command
- Data Frame has special attributes
  - E.g. row.names

Attaching a data frame

- The attach() command helps us to access elements in a data frame with fewer keystrokes
- Using this command, a data frame is attached to R search path
- Once a data frame is attached, objects in the data frame can be accessed by simply giving their names (no need for name of data frame followed by $ sign)
- attach can lead to confusion and errors and therefore not recommended
- To remove a data frame from search path of R, use detach() command

Names

- R objects can have names
- Naming objects makes them more meaningful and helps in easy understanding of code
- Matrices can also have names and they are created using the function dimnames()

Reading & Writing data in R

- Tabular data can be read using the command read.table() or read.csv()
- For data in txt file, read.table() is used and for data in csv format, read.csv() is used
- The default separator for data in read.table() command is space and for read.csv() command is comma
- Each command has different arguments
- If arguments are not specified, R will take default arguments
- To write a data into a file, write.table() or write.csv() can be used

Subsetting

- The operator [ ] returns an object of the same class as the original
  - For e.g. when we subset a vector using [ ] we will get a vector
- The operator [[ ]] is used to extract elements of a data frame or matrix
- We can extract only a single element using [[ ]] and the returned object need not be a vector or a data frame
- To extract elements of a list or data frame by name, we use $