**R Programming** 

**Introduction:**

R is an open source tool, and an integrated suite of software facilities for data manipulation, calculation and graphical display. Among other things it has

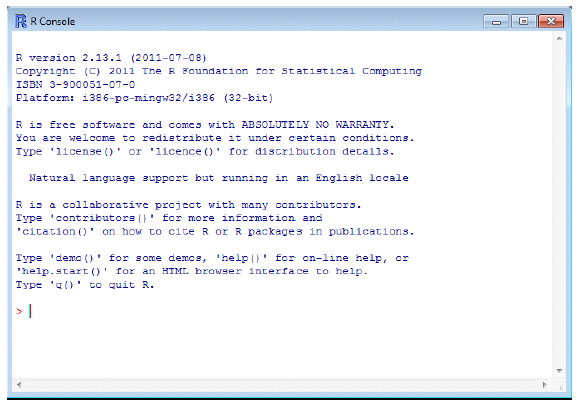
* An effective data handling and storage facility,
* A suite of operators for calculations on arrays, in particular matrices,
* A large, coherent, integrated collection of intermediate tools for data analysis,
* A well developed, simple and effective programming language (called ‘S’) which includes Conditionals, loops; user defined recursive functions and input and output facilities.

Technically R is an expression language with a very simple syntax. It is case sensitive. It provides an environment for statistical computing and graphics and also provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering ...) and graphical techniques, and is highly extensible. Community has developed friendly UI of R “R Studio”.

There are 6 data objects in R. They are vectors, lists, arrays, matrices, data frames and tables.

**Install R and R studio:**

1. Download the “R” software from any of the links at <http://cran.r-project.org/> and install. Choose the OS (Linux, Windows and Mac) that you are using on your system. Once you install the software, you see the work environment as shown below



1. Download r studio from <https://www.rstudio.com/products/rstudio/download/>

All the class room demonstrations are done using R -Studio. This version is more flexible and programmer friendly. Click on the below link to download. Once you install the software, you see the work environment as shown below



**R Packages:**

Packages are collections of R functions, data, and compiled code in a well-defined format. The directory where packages are stored is called the library. R comes with a standard set of packages. Others are available for download and installation. Once installed, they have to be loaded into the session to be used. Few famous packages in R are plyr, ddplyr, data.table, DataCombine, reshape2, rpart, C50, J48, RandomForest, ggplot2 etc..

**Additional Information:**

Generally there two types of programming languages

1. Domain specific language (DSL)
2. General language (GL)

R comes under DSL as it is designed for specific domain, statistics. R works on libraries/packages and it uses RAM for operations and calculations. Major disadvantage of R over python is it uses RAM which is very limited in our systems. We cannot process huge data using R. R is best suited for elementary research. With over 2 million users worldwide R is rapidly becoming the leading programming language in statistics and data science. Every year, the number of R users grows by 40% and an increasing number of organizations are using it in their day-to-day activities.

**Interview Questions:**

1. What are the different data objects in R?
2. What is R Base package?
3. Which data object in R is used to store and process categorical data?
4. What is the main difference between an Array and a matrix?
5. How do you install a package in R?

References:

* <http://www.r-bloggers.com/>
* <http://www.statmethods.net/>
* <http://rfunction.com/>