Statistical Software: R

By Rajni Gupta
Department of Mathematics
St. Stephen's Collge

What is R ?

R is a sophisticated computer language and environment for statistical computing and graphics that is more than just a program. It is a GNU project which is similar to the S-language and environment which was developed at Bell laboratories (formly AT&T, now Lucent Technologies) by John Chambers and colleagues and can be used for exploring and plotting data, as R is an open source(GPL=General public license) environment modeled after S and S-plus.

Why it is called R ?

1. The initial version of R was developed by ROSS IHAKA and ROBERT GENTLEMAN, from the statistical department of the university of AUCKLAND, New Zealand, in 1995. Now development of R is maintained by R-core development Team of which, as of August 2018, R is named partly after the first names of the first two R authors and partly as a play on the name of S.
2. S was created by John Chambers in 1976, while at Bell Labs

Open source software(OSS)

Open source software is a type of computer software in which source code is available for modification and is released under a license.

Statistical Features of R

1. R and its libraries implement a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and others.
2. R is easily extensible through functions and extensions, and the R community is noted for its active contributions in terms of packages.
3. Another strength of R is static graphics, which can produce publication-quality graphs, including mathematical symbols. Dynamic and interactive graphics are available through additional packages.

Programming Features

1. R is an interpreted language; users typically access it through a command-line interpreter.
2. Like other similar languages such as APL and MATLAB, R supports matrix arithmetic.
3. R's data structures include vectors, matrices, arrays, data frames (similar to tables in a relational database) and lists. Arrays are stored in column-major order.
4. R supports procedural programming with functions and, for some functions, object-oriented programming with generic functions.

Advantages of R as compared to other softwares

Benefits

1. R is freely available under GNU General Public License and pre-compiled binary versions are provided for various operating systems like Linux, Windows and Mac.
2. R has an excellent build in help system and Graphing capabilities.
3. R has a powerful easy to learn syntax with many built in statistical functions
4. R users may benefit from a large no. of programs written for S and most of them can be used directly with R.

5. The language is easy to extend with user written functions.

6. Students can easily migrate to the commercially supported programs (which requires purchasing a license) like SPSS, SAS, Minitab and S-Plus.

7. It is also possible to combine different statistical functions in a single program to perform more complex analyses.

Disadvantages

- There is no commercial support.
- The command language is a programming language, so students need to learn syntax issues etc.
- It has limited graphical interface(S-Plus has a good one). This means it can be harder to learn at the outset.

The R website

www.r-project.org is a good place to visit to obtain R-program

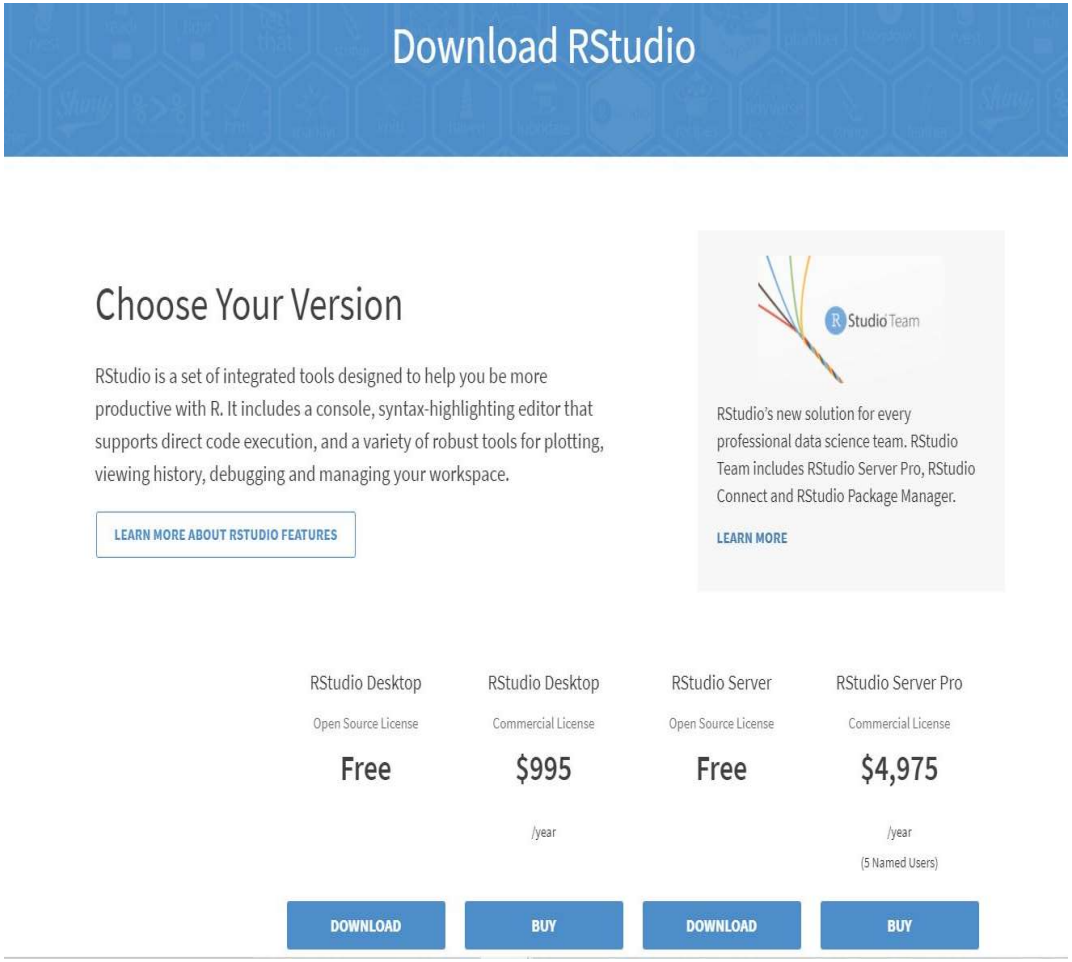
Downloading / Installing

The CRAN(comprehensive R archive network) is a network of websites that host the R program and that mirror the original R website.

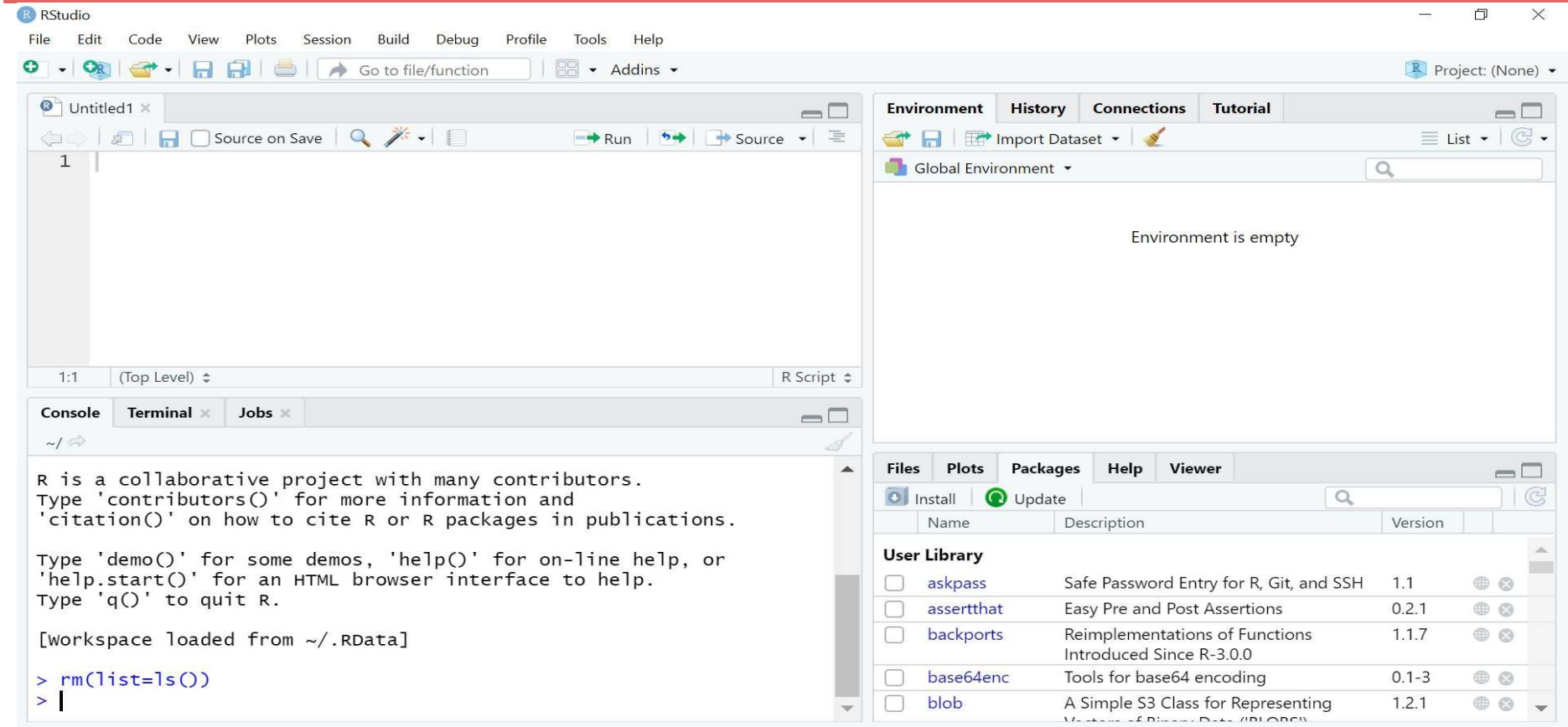
CRAN is the R website and holds downloads and documentation (e.g. Manuals, FAQs)

Downloading R Studio

- Go to www.rstudio.com and click on the “Download RStudio” button.
- Click on “Download RStudio Desktop”
- Click on the version recommended for your system.



Introduction to RStudio : After starting RStudio you will see a window divided into four windows.



Introduction to R studio

Editor Window: In the editor we simple write text files with the function calls we needed for the analysis. These scripts can be saved and return whenever needed. The commands in the editor are not executed. To send the currently selected line to the console and execute it, type CTRL+ENTER

The Console Pane: Here you can directly type commands after the > (prompt symbol) and execute them by hitting ENTER. This is actually R.

Notw: If you see + sign, R expect more output to finish and execute the command. Simply hit ESC to abort the command and return to >

The environment Pane: Here you can see which objects/value/data R know about. It's useful, because you can see the name and a summary of the objects. By clicking on them you can view the contents.

The Help/Package/Files/Plots Pane: Here are plots that you produce shown. Moreover, you can search within the help or browse through packages.

How R works!

If you just type function without parenthesis, R will display the content of the function.

R is an interpreted language, not a compiled one, meaning all the commands typed on keyboard are directly executed without any requirement to build a complete program like in most computer languages(C, Fortran, Pascal, etc.)

R's Syntax is very simple and intuitive. In order to be executed, a function always need to be written with (), even if there is nothing within them.

The End

Now refer Class notes!