

A Brief Introduction to R

The journey of statistical computing and graphics gets interesting when it comes to R but before we delve into the fundamental introduction to what R is, let the words of statistics, statistical computing, and graphics not deter you from starting. We are about to simplify them.

Let us start with data.

Data refers to simply raw information or rather unprocessed information

What are statistics?

Statistics highlight methodologies used in the collection, organization, presentation of raw information, and data, and enable its interpretation.

Scenario: Think of getting a whole sack of candy of different varieties meant to be organized and packed into their respective categories, where would you start? At this point your candy has been collected or brought to you, you now need to handle the organization in terms of grouping based on their attributes. Afterward, how would you present details of what you have, you do not need to understand what they are made of but just generalize what you have. Finally, how would you interpret what you have? How much candy do I have? How many varieties and how is best to divide or package them further?

Statistical computing now enables you to perform the above process on raw data; some processes may include linear, non-linear, and time-series analysis among others. Graphics are based on the visual interpretations of the data we seek to understand further.

What is R and how does it come in?

R is composed of a language and environment aimed to foster techniques targeting statistical computing and graphics in an intuitive, fun, and simple manner. R is a free software provided under the [Free Software Foundation GNU General public license](#).

R Environment

The term environment, with regards to programming, encompasses compiler and associated development tools with regards to a programming language. In this case, R Environment is taken as a fully planned and coherent system as opposed to a gradual incrementation of very specific and inflexible tools.

It is composed of a suite or myriad of software facilities that foster data manipulation, calculation and graphical display.

Its Composition:

- Effective Data handling and storage facilities
- A suite of operators for performing calculations on arrays in particular matrices
- A large coherent integrated collection of intermediate tools for data analysis
- Graphical facilities for data analysis and display either on screen or on hard copy
- Simple well-developed programming language (R) including loops, conditions, recursive functions(user-defined) facilitation of input/ output operations.

How does it work?

R is an interpreted language and not compiled, the user types in the commands or statements and later executed them with no need of building complete programs like in languages like C, Fortran, Pascal etc.

When R is running, variables, data, functions, and results are stored in active memory (no temporary files used) in the form of objects which have a name. All functions of R are done on objects.

Visual Representation of How R Works:

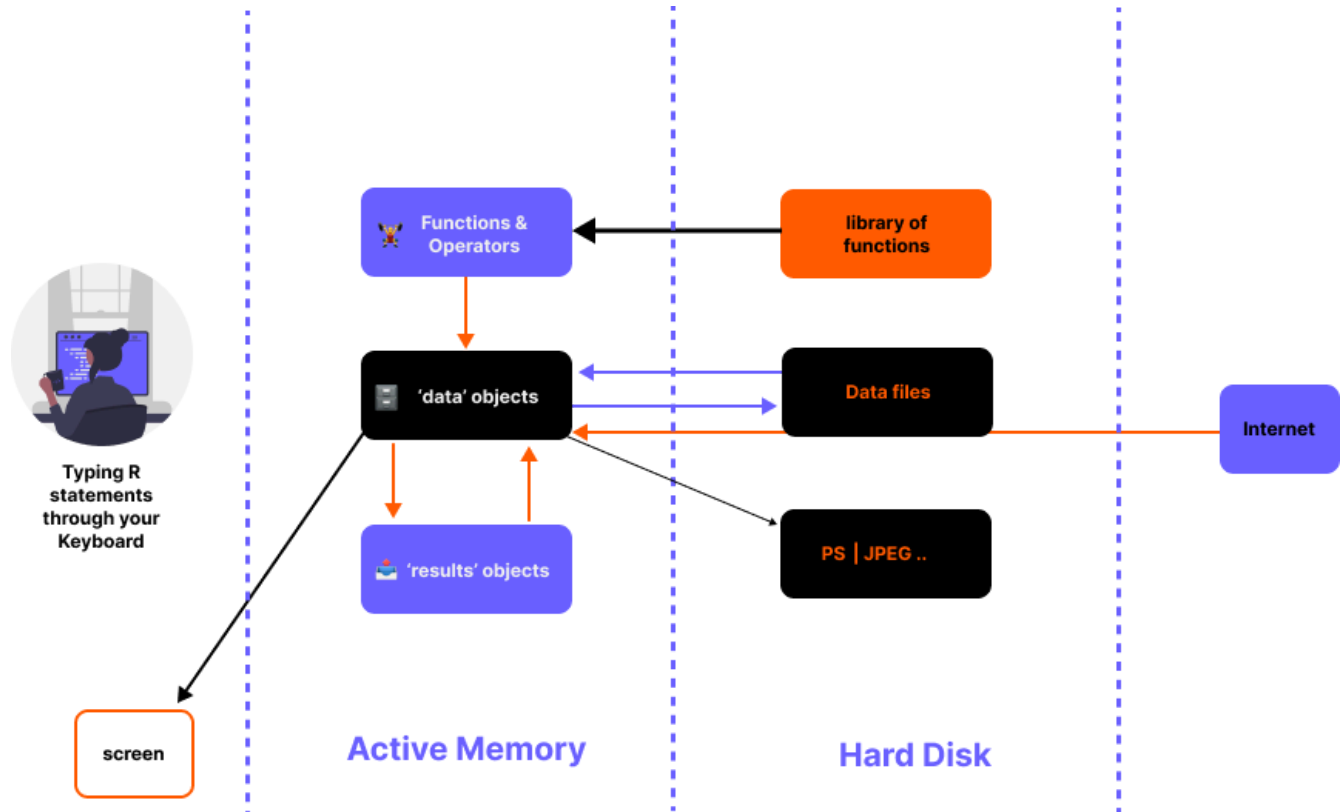


Figure 1: Schematic of How R Works

Advantages of R

- Simple and Intuitive
- Easy setup
- Not limited in terms of statistical tools, highly extensible hence one can build on top of the other
- It is Fun

References:

1. R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL:
<http://www.R-project.org>.