

Getting Started With R

Hamid Abdulsalam

About myself

- Data Scientist at Factual Analytics
- **Contact:** hamid@factual.ng

Why Study Data Science

- The new driving force behind industries is Data, hence every Industry needs data scientists to survive.
- Can be applied to solve problems in all sectors i.e health, Insurance, banking, maritime, energy, telecom e.t.c
- A very Lucrative Career

Why R ?

- Open source
- Over 10,000 packages and still counting
- Rich online Community i.e stackoverflow, stackexchange, rbloggers.
- Leading Language for data science and analytics
- Elegant plots

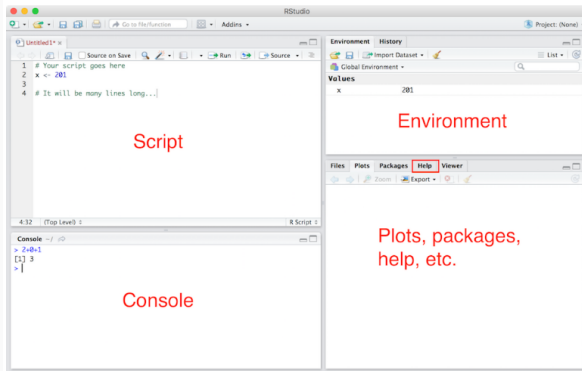
Installing R

To Install R and Rstudio, use the following urls.

- <https://cran.r-project.org/bin/windows/base/>
- <https://rstudio.com/products/rstudio/download/>

- R is an open-source programming language, this implies that it is free and continually improved upon by the R community. R allows you to read, analyze, and visualize datasets.
- With R Installed, you can run R programs via command line but running R programs via the command line can be tricky, Hence Rstudio is recommended.

- RStudio is an open-source integrated development environment (IDE) which allows users to write and execute code



- **Script Section:** A simple text editor for writing R- codes. To write a new R-script, select File > New File > R Script from the menu.
- **Console Section:** The Console allows you to execute one line of code at a time. It shows the printed results from the codes executed from the Script Section.
- **Environment Section:** The environment section is highly useful for debugging. It helps keep track of previous codes that have been executed.
- **File, Plots, packages and help Section:** The plots created from R-scripts are shown in this quadrant. You can also see the packages you have installed in your R or look up on files. If you need information about the usage of a function, you can access the official documentation of the function under the help section.

Variables

Variables are used to store information. Variables names must begin with a letter and they are case sensitive. It can contain any combination of letters, numbers, periods (.), or underscores (_).

- height and Height are different objects in R:

```
height <- 10
```

```
Height <- 50
```

```
height
```

```
## [1] 10
```

```
Height
```

```
## [1] 50
```

Objects in R

There are five basic classes of objects in R:

- character
- numeric (real numbers)
- integer
- complex
- logical (True/False)

The most basic object is a vector

- A vector can only contain objects of the same class e.g (Only Characters or only numbers). An exception to this is a list which can contain objects of different classes. Vectors are created using `c()` function in R

Numbers in R

- R treats numbers as numeric objects (i.e. double precision real numbers) e.g 2.00, 4.058,5.06 e.t.c
- If you explicitly want an integer, you need to specify the L suffix
- Ex: Entering 1 gives you a numeric object; entering 1L explicitly gives you an integer.
- For instance, a vector X which contains 2L,3L will be treated as Integer(Integers are whole number) while a vector which contains 2,3,4 will be treated as numeric

R objects can have attributes

- names, dimnames
- dimensions (e.g. matrices, arrays)
- class
- length

Attributes of an object can be accessed using the `attributes()` function.

Getting Help in R

- `help()` and `?`
- From the R “Console” you can use the `help()` function or `?`. For example, try the following two commands (which give the same result): `help(mean)` `?mean`
- `help.search()` and `??` The `help.search()` function scans the documentation for packages installed in your library.

For Instance, `help.search("^levneTest")` searches for help pages, vignettes, and code demos that have help “aliases,” “concepts,” or titles that begin (case-insensitively) with the characters “levneTest”