

Introduction to R

Bobbur Abhilash Chowdary

August 2018

Why R?

- It's open source!
- Need not run behind IT support to get the license renewed (SAS users?)
- It's one of the fastest growing languages along with Python for Machine Learning and AI
- Easy integration with GitHub and Latex
- Reproducible research(?)
- Imagine with one click your code analyzes your data and generates all the required tables & graphs, and finally generates a nice pdf/word document which you can send it to a journal. This can be done in R.

Installing R

- Download and install R from <https://www.r-project.org/>
- User interface of just R is very bad. So we need to install R Studio. There are alternative to R-Studio like Emacs etc. You can also try them.
- Download and install R-Studio from <https://www.rstudio.com/>. Free version is good enough.
- R-studio runs on top of R. R-studio relies on R to execute all the commands. R-studio cannot function without R.
- Don't change the order of installation. First R and then R-Studio. Otherwise sometime it can create problems.
- Try this in console of R-Studio to check if its working

```
1+2
```

```
## [1] 3
```

Optional - Git

Have you ever been frustated seeing files named Final_Version, New_Final_Version . . . ? Have you ever wondered what is the difference between Version_1 and Version_2?

Then what you need is a version control system. Git is a popular version control system. Luckily R-Studio has very good interface with Git and GitHub.

Once again Git is open source and free. One can download Git from <https://git-scm.com/>.

Git Clients : Git ~ RStudio : R. There several Git clients out there. You may use any of them (Sourcetree, GitUp, GitHub etc).

Rstudio does most of the basic stuff a Git clients does. Did I forget to say RStudio has good interface with Git?

Also create GitHub account so you can store your code online. For private repositories if you register with IIM Cal id its free.

Step 1 - Open Git Bash app and type the following commands

```
git config --global user.name 'abcd efgh'
git config --global user.email 'abcdexy@email.iimcal.ac.in'
git config --global --list
```

Step 2 - GO to GitHub.com and create a new repository. Choose default options. Open your repository to find readme.md file. Press “clone or download” button on top right corner and copy the url.

Step 3 - Open Rstudio >File >New Project >Version Control >Git . Paste the above url and chose the folder you want to install the project.

Step 4 - Open the readme.md file. Make some changes and save it. In your Git window in Rstudio (beside environment and history) select Readme.md and press commit. Type a commit message in new window and then press commit. Then press push to send the commit to GitHub.

Open GitHub.com to find the changes reflected in Readme.md file there.

RStudio Basics

‘Source’ pane shows the scripts. For a new script : RStudio >File >New File >R Script

Script has .R extension. It is the place where you will write all your code. Select the code and press ‘Run’ option on top right corner to execute it. Shortcut: CTRL + ENTER

The ‘Console’ pane is where you will find the result of executing the code you have written in script.

‘Environment’ pane is where you will find the datasets you have loaded into the memory.

‘Files’ pane contains the list of files in the current working directory

Intro to R

- Clear Console

```
cat("\014")
```

- Display current working directory and set current working directory

```
getwd()
setwd("G:/My Drive/Thesis/BGs")
```

- Installing Packages and Loading Packages. Need to install a library only once per computer but have to load it everytime you restart.

```
install.packages("data.table")
library(data.table)
```

- head and tail functions provide the first and last ‘n’ lines of a given dataset. Here we use a popular and default dataset already stored in R - mtcars

```
var1 <- head(mtcars,10)
var2 <- tail(mtcars,10)
```

- Removing one variable or all variables from ‘Environment’

```
rm(var1)
rm(list = ls())
```

- Reading files

```
data <- read.csv(file = "Input/data.csv", header = TRUE) data <- read.table("Input/data.txt", header = TRUE, sep = "|", quote="")
```

Slide with R Output

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

Slide with Plot

```
plot(pressure)
```