Statistics Bootcamp using R

DAY 1 INTRODUCTION TO STATISTICS IN BUSINESS

1.2 INTRODUCTION TO R

GU Zhan (Sam)
Institute of Systems Science
National University of Singapore

issgz@nus.edu.sg

Agenda

Day 1: Introduction to Statistics in Business

- Basic Vocabulary of Statistics & Data Types
- **Introduction to R**
- Data Collection & Summarization

Learning objectives

- Understand what R can do
- Understand R studio interface
- Run simple R commands
- Understand the common mistakes

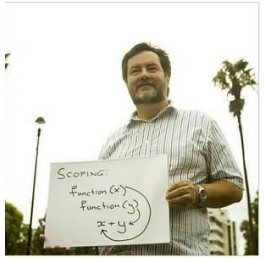
What is R?

project considered in 1992

initial version released in 1995

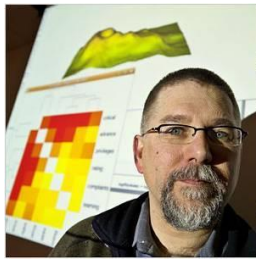
stable beta version in 2000

named after first names of first 2 authors



was created at Auckland New Zealand

was created by Ross Ihaka & Robert Gentleman



Source: https://data-flair.training/blogs/r-interview-questions-and-answers/

Why should we use R?

A Open source programming language

B Serves as a glue language

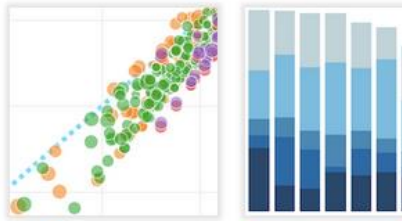
C R groups are noted for its energetic contributions

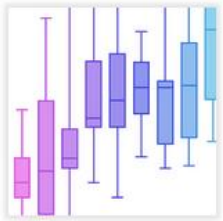
D R has gotten faster over time

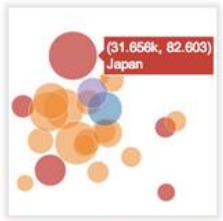
Source: https://data-flair.training/blogs/r-interview-questions-and-answers/

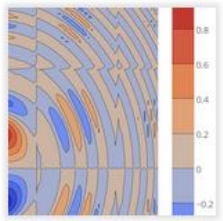
Other equivalent software : **Commercial:** SAS, SPSS, JMP etc.
Open Source: Python etc.

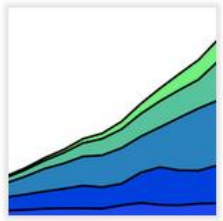
R has great graphic capabilities

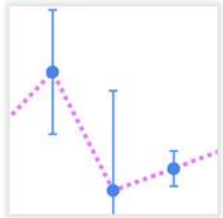


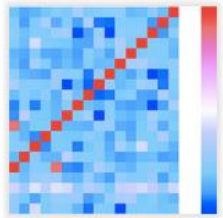


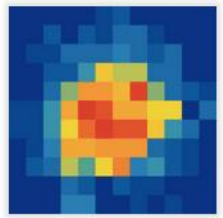


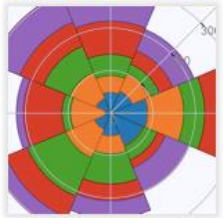


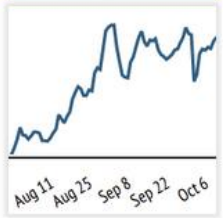


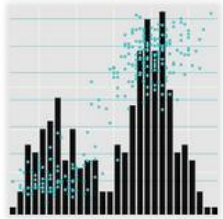


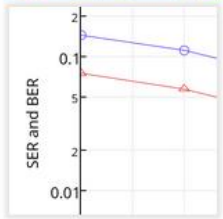


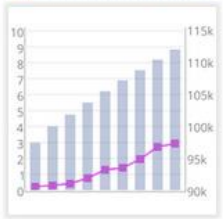


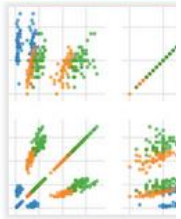


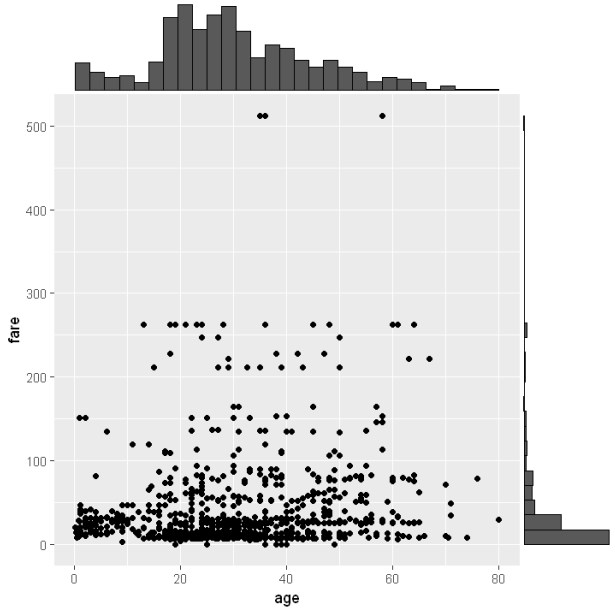


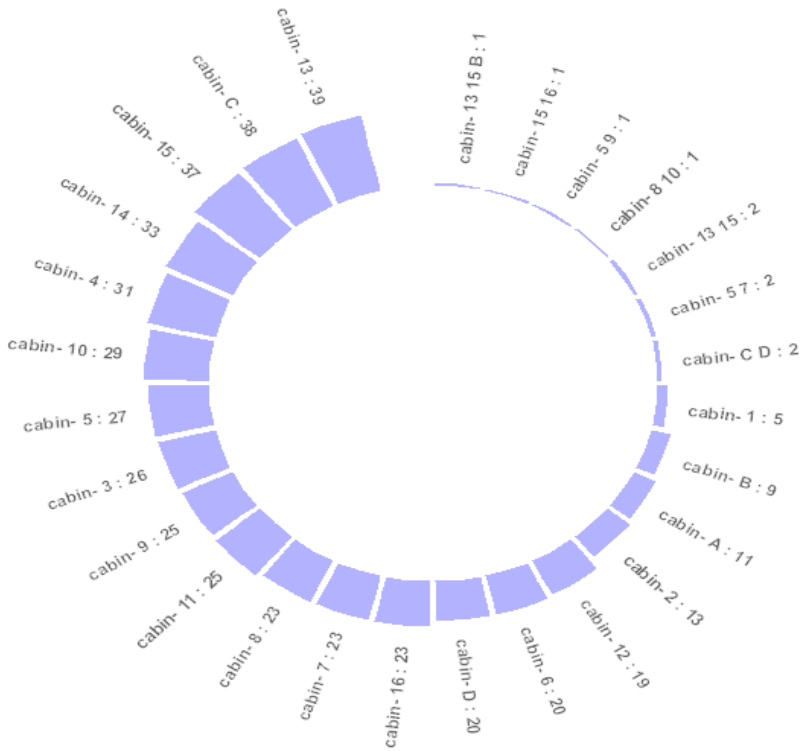












How applicable is R?

A R is used in wide range of industry such as finance, e-commerce, semiconductors, manufacturing and many more

B Considered as an alternate execution of science

C R is one of the most prevalent language though Python is also gaining popularity

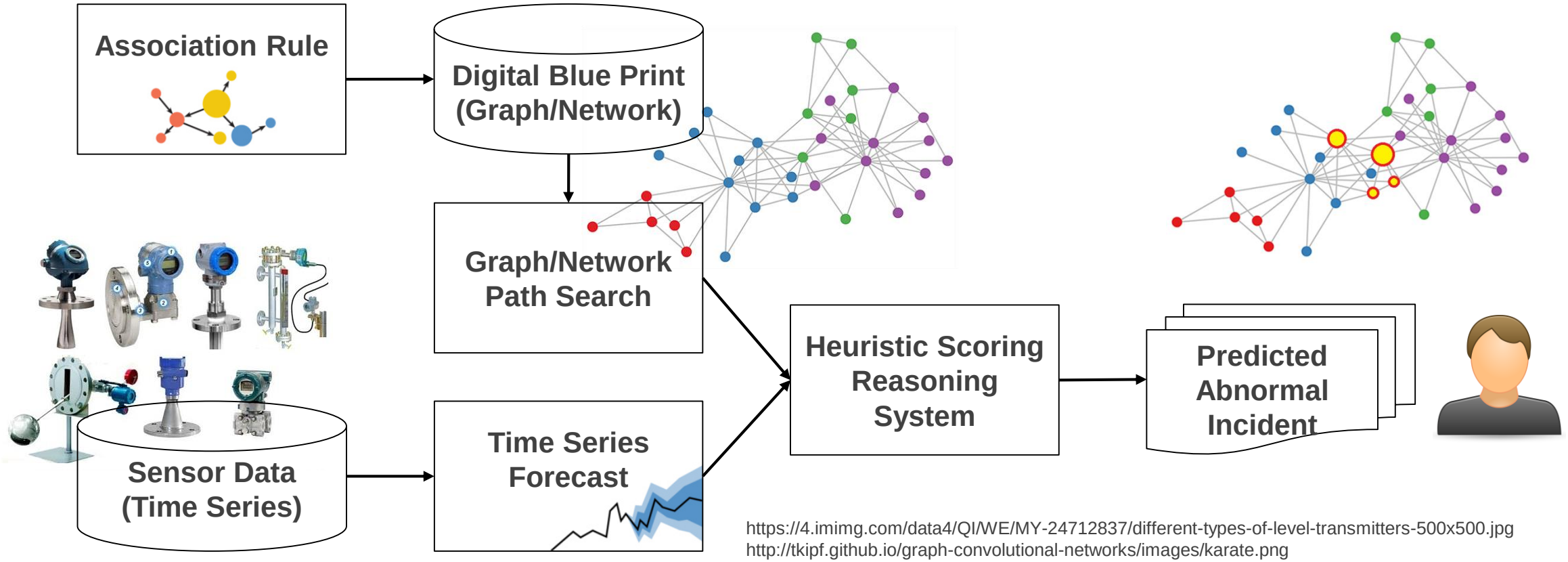
D R helps in data importing and cleaning

E R is one of the best open source tools for data science today

Source: https://data-flair.training/blogs/r-interview-questions-and-answers/

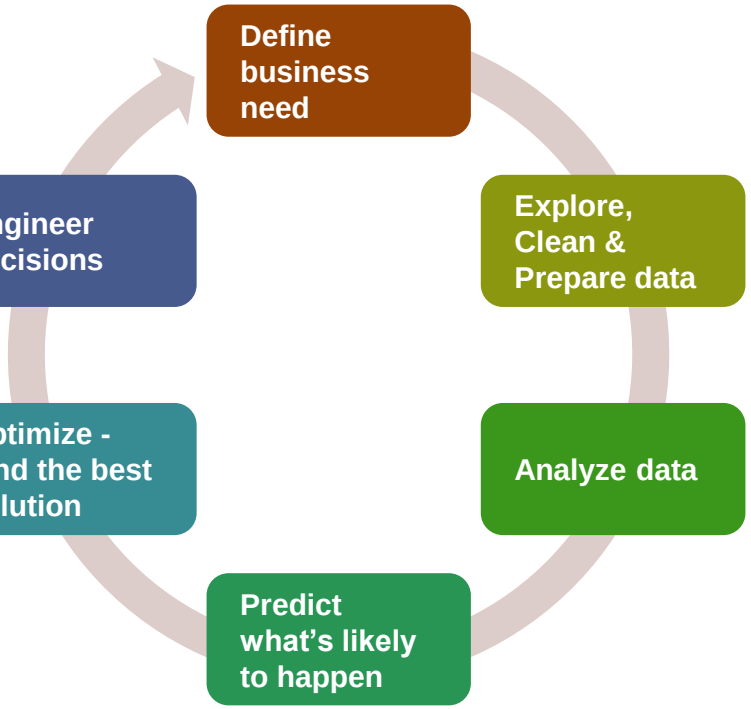
How applicable is R?

Use case: Plant Abnormal Situation Prediction (Predictive Maintenance)
Sub-systems: Time Series Forecast, Association Rule, Graph Search, Heuristic Score



https://4.imimg.com/data4/QI/WE/MY-24712837/different-types-of-level-transmitters-500x500.jpg
http://tkipf.github.io/graph-convolutional-networks/images/karate.png
https://www.microsoft.com

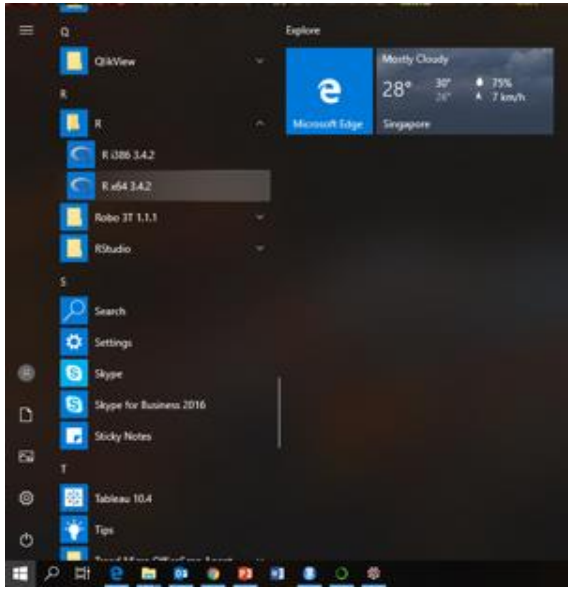
R is a vital part of the analytics process life-cycle



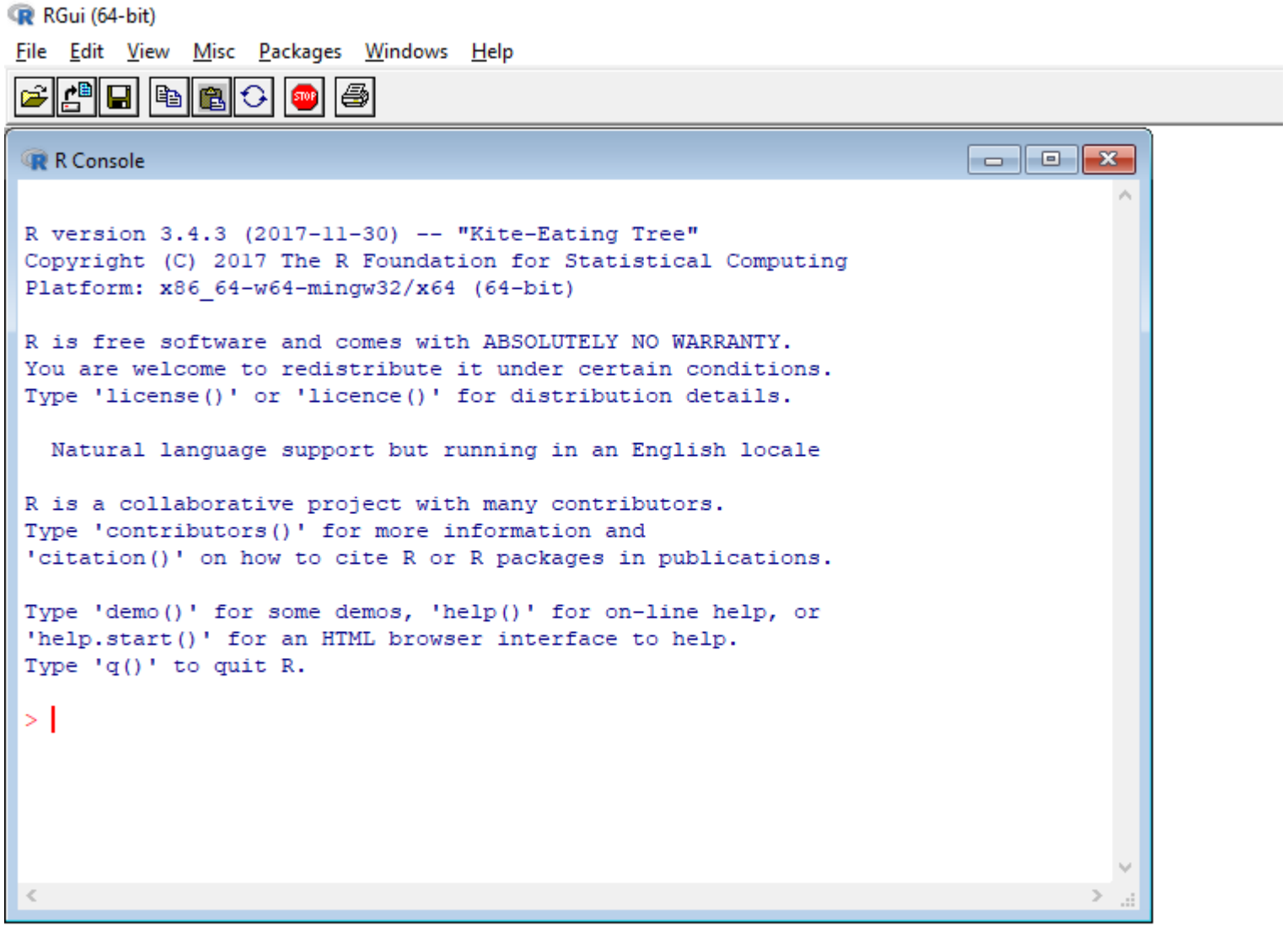
- Enable decision engineering
- Data ingestion
- Data cleaning, feature creation
- Optimization of models
- Visual and non-visual analysis of data
- Statistical comparison and analysis between groups (test and validate)
- Build model and make prediction

Obtaining R

- Freely available from the Comprehensive R Archive Network (CRAN)
- Downloadable at http://CRAN.R-project.org/mirrors.html
- Runs on Windows, Mac and Linux
- To access the installed R, for windows user, go to *Start Menu*, select 'All apps', go to folder 'R', click 'R x64 3.x.x'

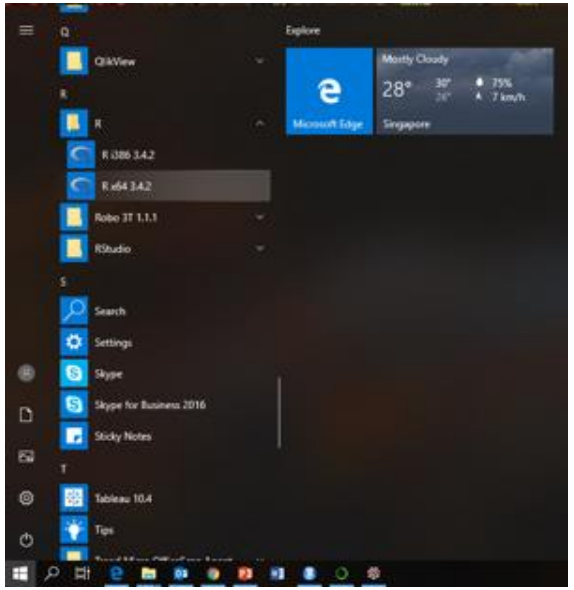


** It is possible to run R on USB stick, when you don't have administrative right to the computer.*

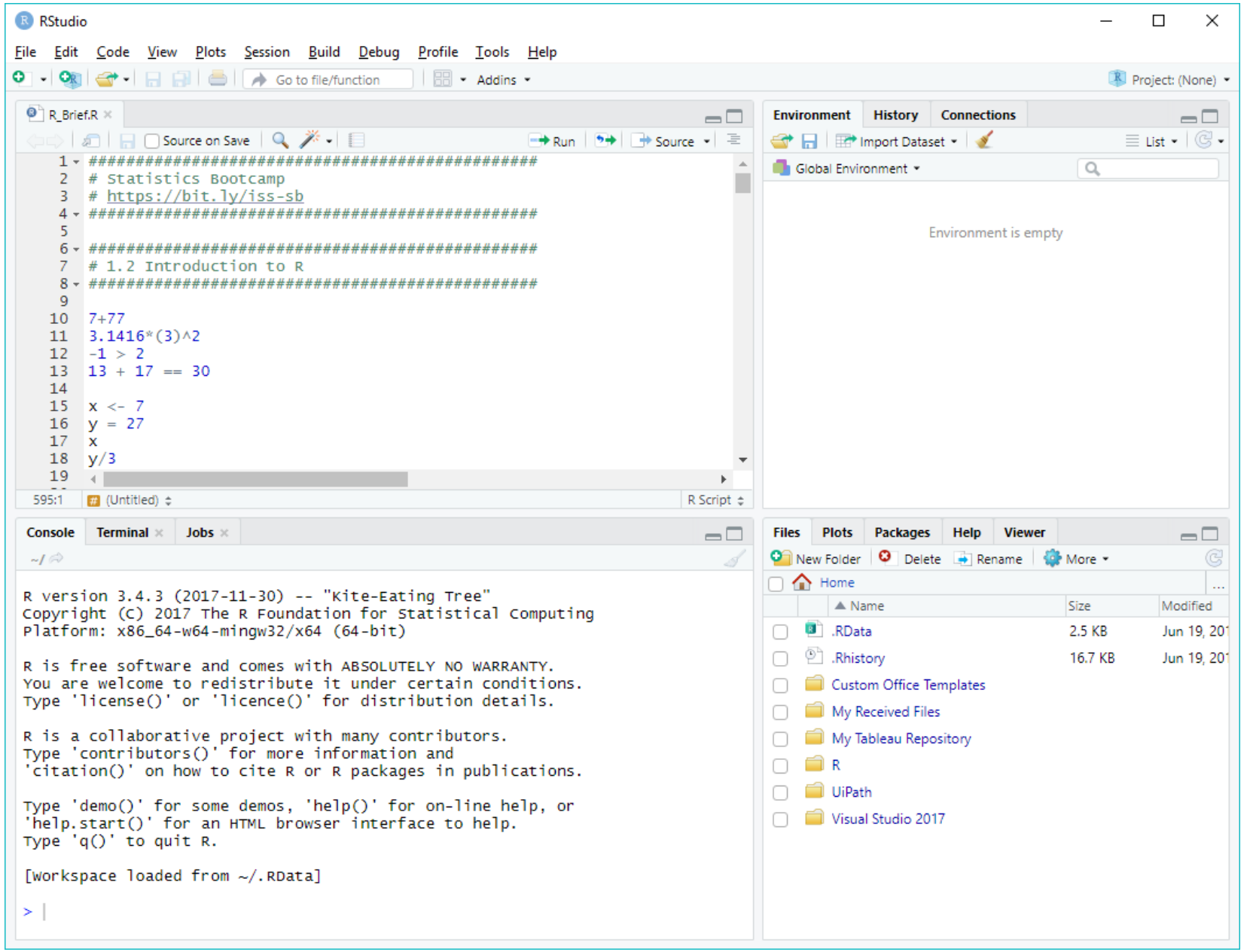


Installing R studio

- An integrated development environment (IDE) built for R
- Downloadable at http://www.rstudio.com
- Runs on Windows, Mac and Linux, or over the web using RStudio Server
- To access the installed R, for windows user, go to *Start Menu*, select 'All apps', go to folder 'RStudio', click 'RStudio'



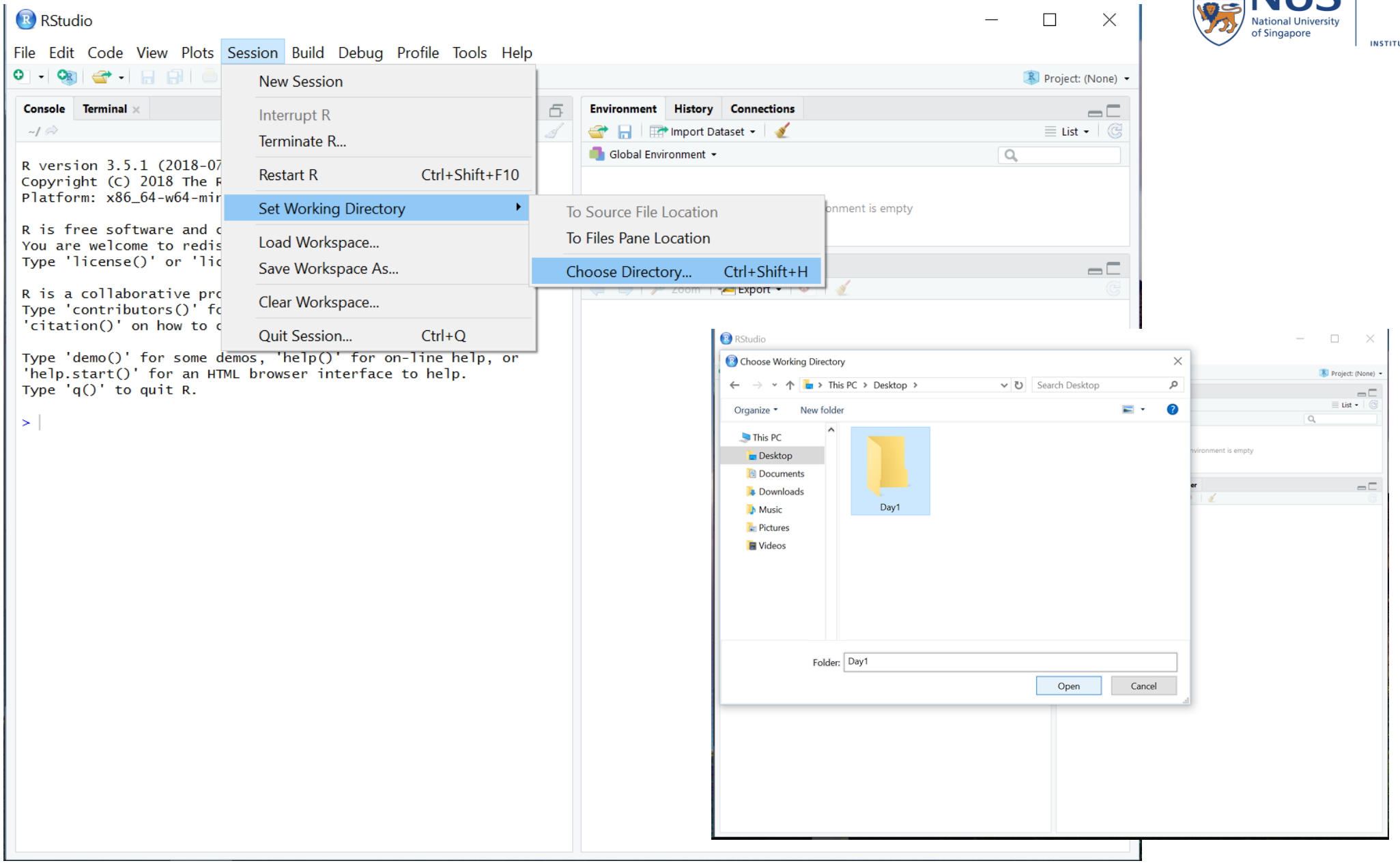
** It is possible to run R and R studio on USB stick. You need to install R before installing R studio.*



Set working directory

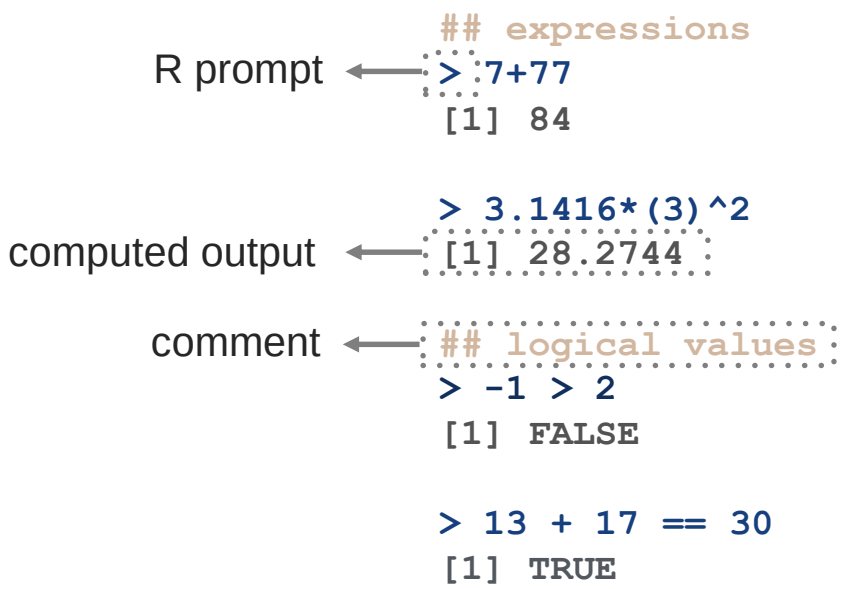
- This step is often needed when you start RStudio

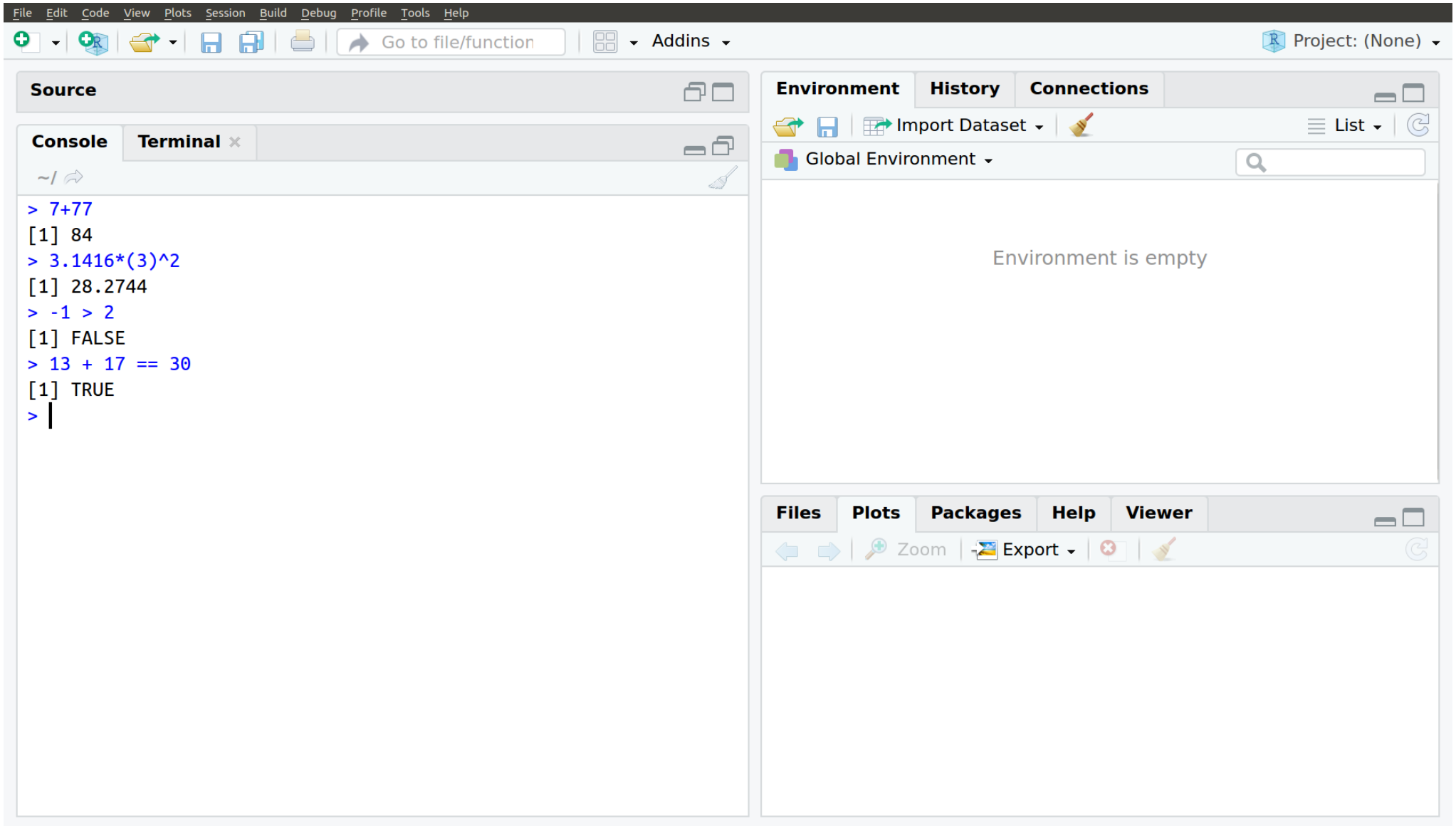
- Not setting the correct working directory is a common mistake among new learners



How R works

1. Type expressions at the R prompt
2. Expressions evaluated by the R interpreter
3. Then computed values are printed below the prompt
4. Some examples:





Basic R

- Variable: store value (number, string, etc) for access later
- Assignment operator: `<-` or `=`

```
> x <- 7
> y = 27
> x
[1] 7

> y/3
[1] 9

> a = 'Learning statistics'
> B <- 'Learning R'
```

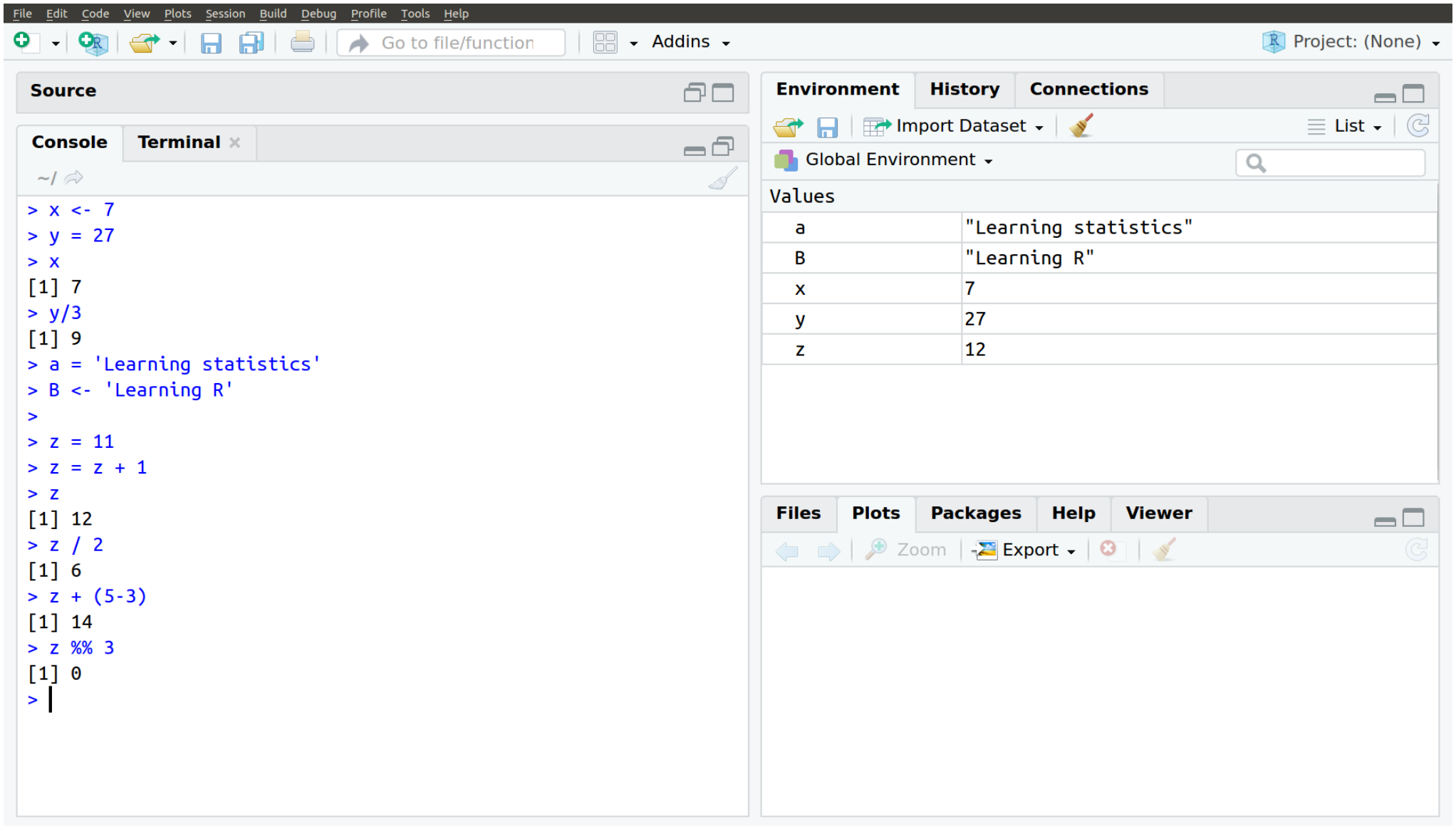
- `+`, `-`, `/` , `<-`, `=`, `%%`, `==`, these are called operators:

```
> z = 11
> z = z + 1
> z
[1] 12

> z / 2
[1] 6

> z + (5-3)
[1] 14

> z %% 3
[1] 0
```



Function

- A piece of code written to carry out a specified task
- It can or can not accept arguments
- It can or can not return one or more values

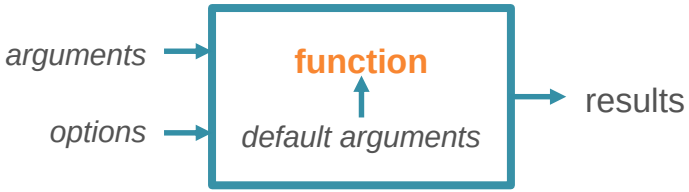
```
> sum(2,4,5)

[1] 11
> sqrt(121)

[1] 11
```

- Some arguments have names. For example to display the help page for ‘stats’ package, do:

```
> help(package="stats")
```



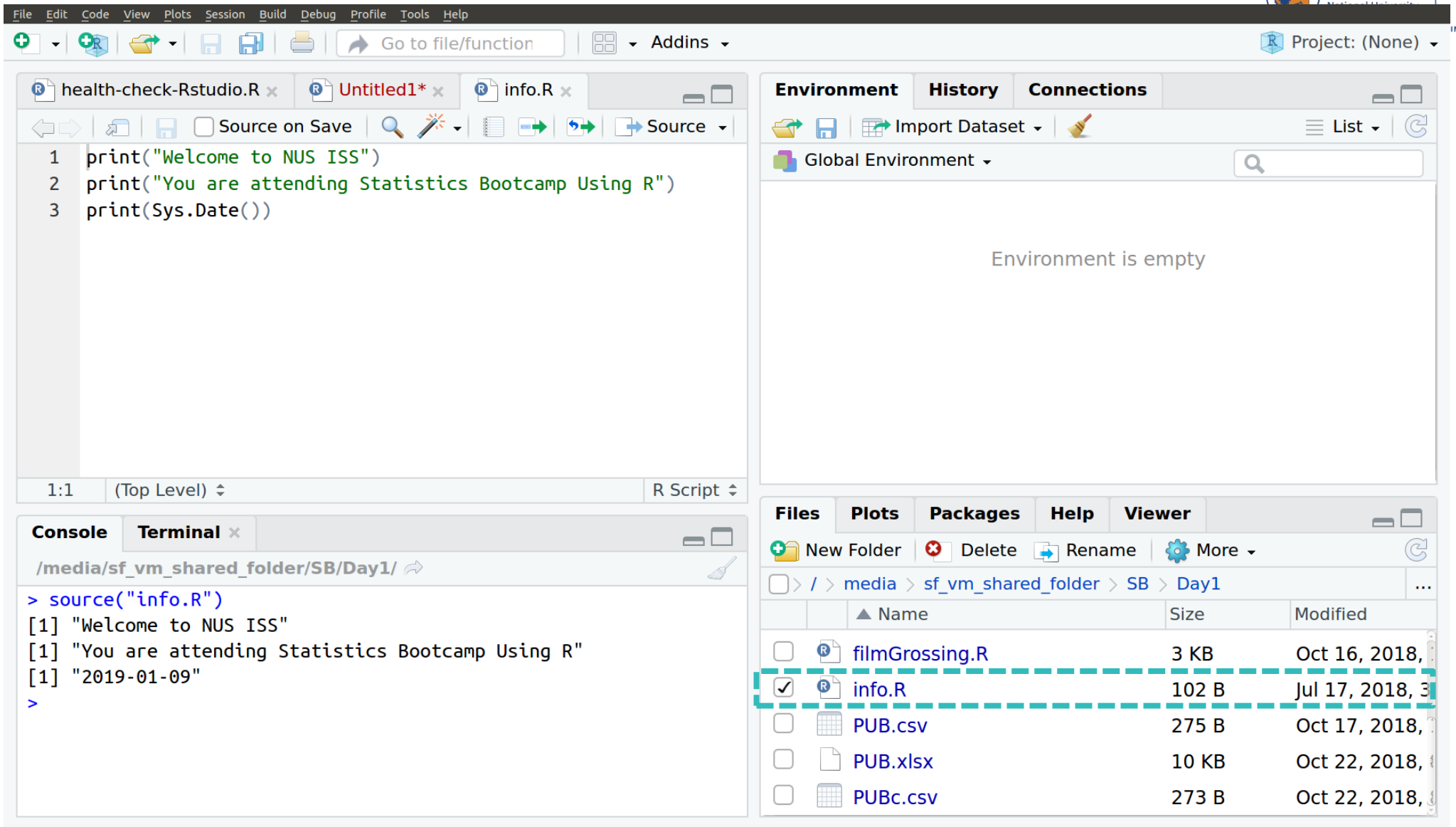
R script

- For most of the time, we run R interactively. Enter commands at the prompt and observe the result.

- At times, we may want to run a groups of R commands at one go. In this case, we write R commands in a plain text file (with ".R" extension) and run the script.

- To run a script file named 'info.R', type:

```
> source("info.R")
```



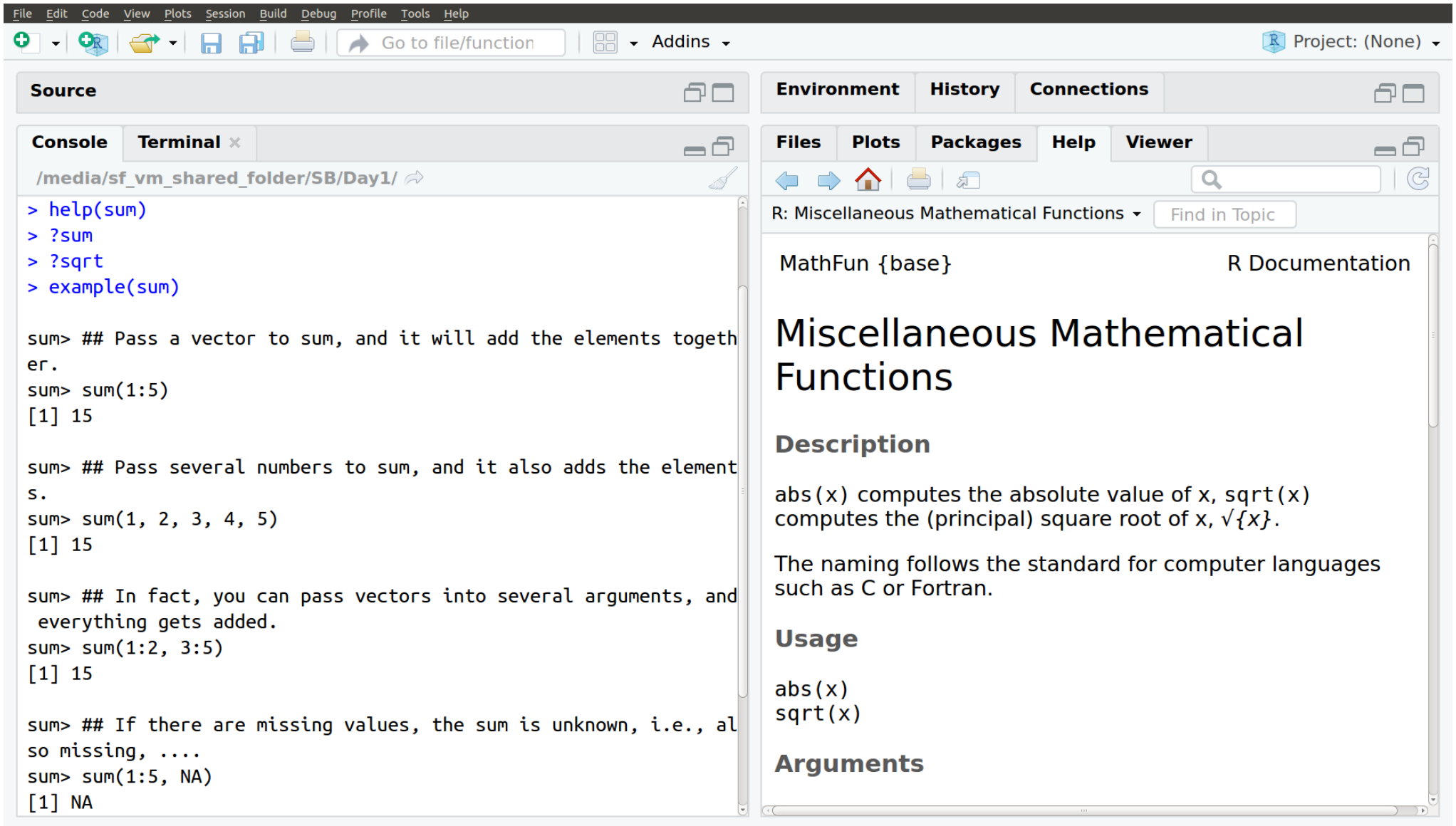
Getting help

- General help (Menu > Help > R Help)
- To get a help on a particular function, type

```
> help(sum)
> ?sum
> ?sqrt
```

- Run the examples of an R function if provided

```
> example(sum)
```



Common mistakes in R programming

	Example	
	√	X
Using the wrong case	`help()`	`Help()`
Forgetting to use quote marks	`install.packages("lattice")`	`install.packages(lattice)`
Forgetting to include the parentheses in a function call	`help()`	`help`
Using the \ in a pathname on Windows	`c:/mydata` `c:\\mydata`	`c:\mydata`
Using a function from a package that's not loaded	`install.packages("lattice")` `contourplot()`	`contourplot()`

Useful R functions

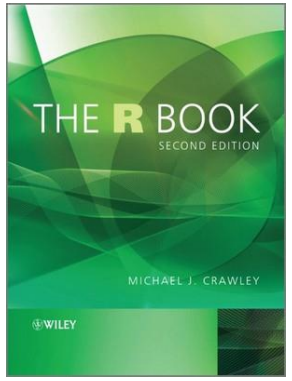
Function	Purpose
`library()`	Load add-on packages.
`install.packages()`	Download and install packages from CRAN-like repositories or from local files.
`library(help="package_name")`	Show documentation for a package.
`sum()`	Compute the sum of values.
`sqrt()`	Compute the square root.
`rep()`	Replicate a value.
`help()`	Access the primary interface to the R help systems.
`??function`	To find out which package carries this function.
`example()`	Run the examples of an R function
`source()`	Submit a script to the current R session

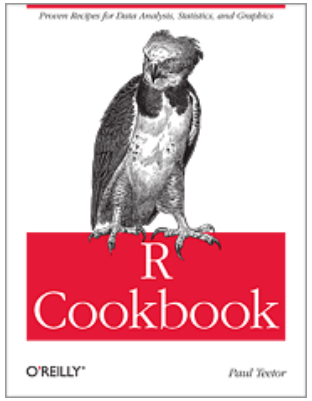
Useful tips

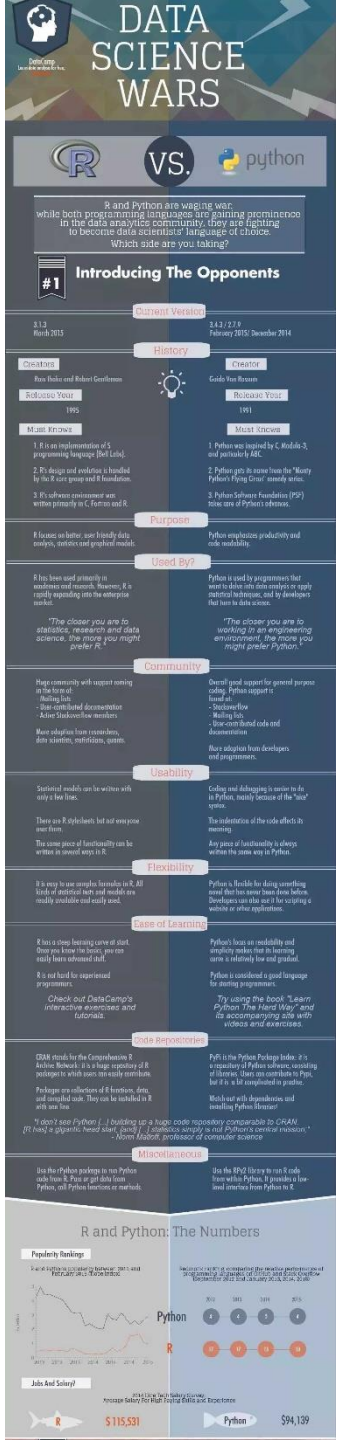
Tip	Purpose
Use arrow keys	To navigate between previously typed in commands
Control L	Clears the console
library(help="package_name")	Show documentation for a package.
Typing inside quotes	Both quotes are automatically added
Setting working directory	See screenshot on page 23
Keyboard shortcuts	Alt + Shift + K
Open a Terminal Window	Shift + Alt + r
If you see a “+” instead of a “>”	Your command is incomplete. Check your syntax.

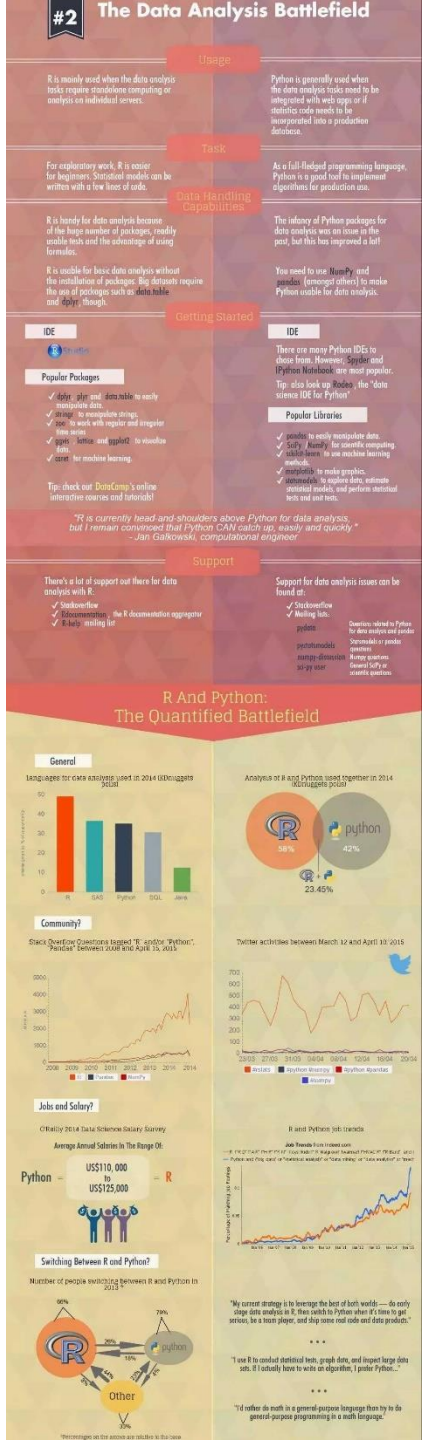
Other resources

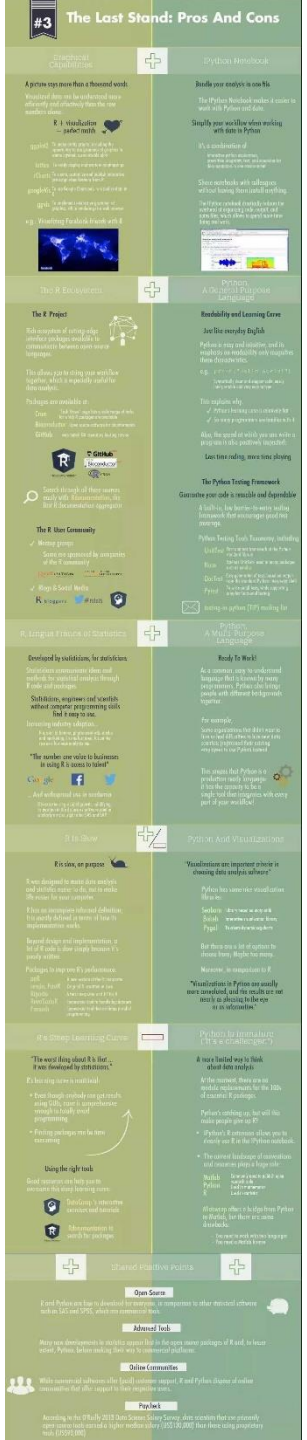
http://www.r-project.org
http://www.r-bloggers.com
http://www.rstudio.com
http://stackoverflow.com
http://stats.stackexchange.com
http://quora.com
http://www.google.com
https://github.com/telescopeuser/S-SB-Workshop











< Data Science Wars: R vs Python >

https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis

End of Lecture Notes