

# Big Data: Data Analysis Boot Camp

## What is R?

Chuck Cartledge, PhD

19 January 2018

# Table of contents (1 of 1)

- 1 Intro.
- 2 What is R?
  - The language
- 3 RStudio
  - Basic how-tos (left side)
  - Basic how-tos (right side)
- 4 R Basics
  - Variables
  - Functions
- 5 Hands-on
- 6 Q & A
- 7 Conclusion
- 8 References
- 9 Files

# What are we going to cover?

We're going to talk about:

- What is the language R?
- What GUI do I use to write and execute R programs?
- What are some basic variable types in R?





## The official definition.

*“R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.*

*R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, . . . ) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.”*

*CRAN Staff [2]*



# R is available for almost all major operating systems.

- Linux (and its variants)
- (Mac) OS X
- Windows



Get the R environment and a command line interface.

Download from: <https://cloud.r-project.org/>

Source code is available for custom OSs.

<https://github.com/wch/r-source>

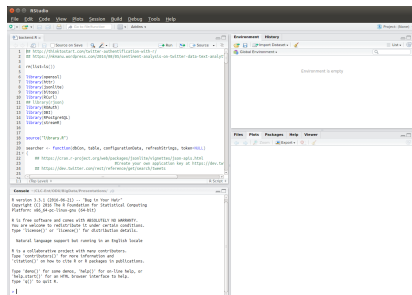
Basic how-tos (left side)

# A complete IDE

A complete, integrated R development environment.

- 1 Text editor
- 2 R console
- 3 Variable list and contents
- 4 Tabbed display for different uses

See software overview and design document for version and download information.





Basic how-tos (left side)

# Same image.

The screenshot displays the RStudio environment. The main editor window contains the following R code:

```

1 # https://thekostart.com/twitter-authentication-with-r/
2 # https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=la())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(bttops)
10 library(Rcurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(OBS)
14 library(RPostgreSQL)
15 library(stream)
16
17
18 source("library.R")
19
20 searcher <- function(@Con, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-opts.html
23   ## Create your own application key at https://dev.tw
24   ## https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }
  
```

The Environment pane on the right shows "Environment is empty". The Console pane at the bottom displays the R startup message:

```

Console ~CLC-Eno/ODU/BigData/Presentations/ ./
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
  
```

# Editor

- “Smart” editor
- CTRL + O to open a file
- CTRL + S to save a file
- CTRL + A to highlight contents
- CTRL + Enter to transfer contents to Console
- Multiple files can be opened at once

The screenshot displays the RStudio interface. The main window is a script editor containing R code for loading libraries and defining a function. The console window at the bottom shows the R version and copyright information. The package manager window on the right is empty.

```
## ---- Basics on Base ----
1 # RStudio is based on the R Foundation's source code, licensed under the
2 # GPL. (https://www.r-project.org/licenses/2015-clr)
3 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
4 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
5 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
6 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
7 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
8 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
9 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
10 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
11 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
12 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
13 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
14 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
15 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
16 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
17 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
18 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
19 # RStudio is licensed under the RStudio License. (https://www.rstudio.com/licenses/rstudio-license/)
20 source("library.R")
21
22 searcher = function(token, table, configurationData, refreshToken, tokenURL) {
23   ## https://cran.r-project.org/web/packages/searcher/index.html
24   ## https://www.rstudio.com/products/rstudio/docs/1/working-with-data-science/
25 }
26
```

Console: RStudio - RStudio (https://www.rstudio.com/)

```
R version 3.5.1 (2018-06-21) -- "Bug to Your Heart"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```





Basic how-tos (left side)

# Same image.

The screenshot shows the RStudio environment with the following components:

- Source Editor (Left):** Contains R code for loading libraries and defining a function. The code is:
 

```

1 # http://thekostart.com/twitter-authentication-with-r/
2 # https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=la())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(bitops)
10 library(Rcurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(OAuth)
14 library(RPostgreSQL)
15 library(stream)
16
17
18 source("library.R")
19
20 searcher <- function(@Conn, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-opts.html
23   ## Create your own application key at https://dev.tw
24   ## https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }
27
28 # Top Level: 3
      
```
- Environment (Right):** Shows "Global Environment" and "Environment is empty".
- Console (Bottom):** Shows the R startup message:
 

```

Console ~CLC-Env/ODU/BigData/Presentations /C/
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or # packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
      
```

# Console

- Interprets R commands
- Commands from editor, other panels, or manually entered
- Execution errors appear here
- Contents of print function appear here

```

R version 3.5.1 (2018-06-22) -- "Tap to your step!"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
For more details see the 'COPYING' file in the source distribution.

In natural language: concept not being taught in an English lecture.

R is a collaborative product with many contributors.
Please 'contribute!' for more information and
instructions on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
  
```



Basic how-tos (left side)

# Same image.

The screenshot shows the RStudio interface. The script editor on the left contains the following R code:

```

1 ## https://thekostart.com/twitter-authentication-with-r/
2 ## https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=la())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(bitops)
10 library(RCurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(OBS)
14 library(RPostgreSQL)
15 library(stream)
16
17
18 source("library.R")
19
20 searcher <- function(@Conn, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-opts.html
23   ## Create your own application key at https://dev.tw
24   ## https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }
27
  
```

The console window at the bottom, highlighted with a red border, displays the R startup message:

```

<C.LC-End/ODU/BigData/Presentations />
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

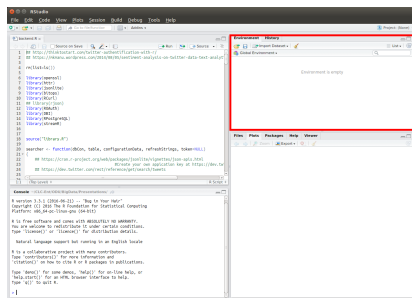
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
  
```

# Variables

- Displays contents of selected environment (including variables)
- Display history of console commands
- Can save and load data from data files





Basic how-tos (right side)

# Same image.

The screenshot displays the RStudio environment. The script editor on the left contains the following R code:

```

1 ## http://dlaktostart.com/twitter-authentication-with-r/
2 ## https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=la())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(bttops)
10 library(RCurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(OBS)
14 library(RPostgreSQL)
15 library(stream)
16
17
18 source("library.R")
19
20 searcher <- function(@Conn, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-opts.html
23   ## Create your own application key at https://dev.tw
24   ## https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }
  
```

The console at the bottom shows the R startup message:

```

Console ~CLC-Eno/ODU/BigData/Presentations/ ./
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

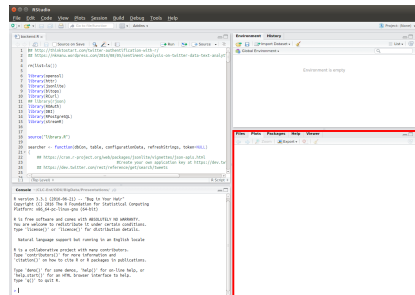
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
  
```

The Environment pane on the right, highlighted with a red box, shows "Environment is empty".

# Tabbed display

- Displays files in the current directory
- Displays plots from the console
- Allows packages to be added, or removed from the console
- Provides help/man pages for R functions and packages





Basic how-tos (right side)

# Same image.

The screenshot displays the RStudio environment with the following components:

- Script Editor:** Contains R code for loading libraries and defining a function. The code includes:
 

```
1 # http://dlaktostart.com/twitter-authentication-with-r/
2 # https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=la())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(dtpops)
10 library(RCurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(OAuth)
14 library(RPostgreSQL)
15 library(stream)
16
17
18 source("library.R")
19
20 searcher <- function(@Con, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-opts.html
23   #Create your own application key at https://dev.tw
24   ## https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }
```
- Environment Pane:** Shows "Global Environment" and "Environment is empty".
- Console:** Displays the R startup message:
 

```
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

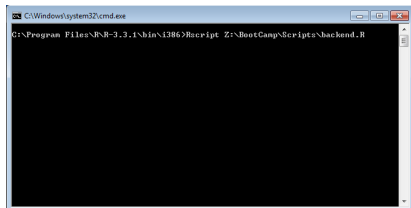
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```
- Files Pane:** A red box highlights the "Files" pane, which is currently empty.

# Starting an R script in the background

The image shows a Windows environment.

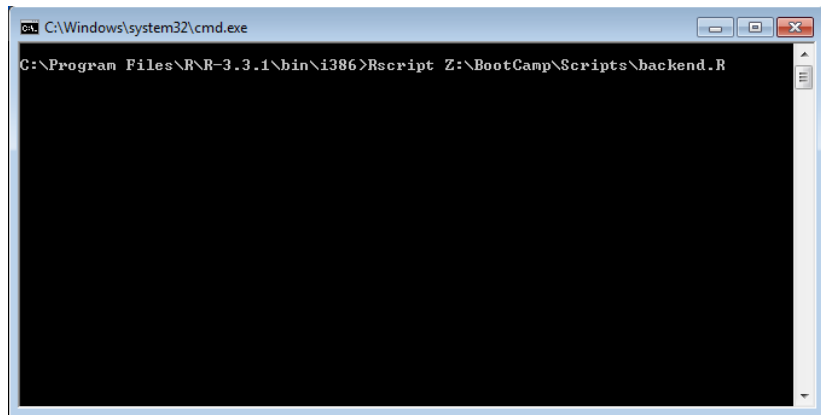
A \*nix environment command is:  
`Rscript backend.R &`



```
C:\Windows\system32\cmd.exe
G:\Program Files\R\R-3.3.1\bin\x86>Rscript Z:\BootCamp\Scripts\backend.R
```



# Same image.



```
C:\Windows\system32\cmd.exe  
C:\Program Files\R\R-3.3.1\bin\i386>Rscript Z:\BootCamp\Scripts\backend.R
```

# Basic help with functions[1]

## 1 Based on subject

```
1 help.search("data input")
```

## 2 Based on pattern matching

```
1 apropos("lm")
```

## 3 Looking for a specific item

```
1 find("lm")
```

## 4 About a specific item

```
1 ?lm  
2 ??lm
```

## 5 Example of a function

```
1 example(lm)
```

## 6 Source code for a function

```
1 lm
```

## 7 Demonstration of a function

```
1 demo(persp)
```

## 8 Demonstration of a function

```
1 vignette("moveline", package="  
  grid")
```

## 9 Contents of a library

```
1 library(help=spatial)
```

## 10 Install a new library

```
1 install.packages("Kfn")
```

# Variable types (part 1 of 2)[3]

## 1 Variable names:

- Names are case sensitive
- Names cannot begin with numbers or special symbols
- Names cannot have internal spaces

## 2 Scalars (simple values)

```
1 variable <- 3
```

## 3 Vector of values

```
1 variable <- c(3, 5, 7, 10)
```

## 4 Matrix of values

```
1 variable <- matrix(1:20, nrow=5,  
                    ncol=4)
```

## 5 Access matrix values

```
1 b <- variable[3,4]  
2 b <- variable[,4]  
3 b <- variable[3, ]
```

## 6 Array of values

```
1 variable <- array(1:20, dim=c  
                  (2,2,5))
```

## Variable types (part 2 of 2)[3]

- 1 Data frames (each column must have the same number of values)

```
1 L3 <- LETTERS[1:3]
2 fac <- sample(L3, 10, replace = TRUE)
3 d <- data.frame(x = 1, y = 1:10, fac = fac)
```

- 2 Lists (each list entry can have a different number of values)

```
1 pts <- list(x = cars[,1], y = cars[1:23,2])
```

- 3 Saving variables

```
1 save(x, y, file = "xy.RData")
```

- 4 Loading variables

```
1 load(file = "xy.RData")
```

# Functions are supported

- 1 Have the same naming conventions as variables
- 2 Have three parts:
  - 1 Optional pass parameters (named, evaluated, unnamed)
  - 2 Text of the function
  - 3 The environment where and while the function executes

```
1 functionName <- function(p1, p2=1234, ...)  
2 {  
3   p3 <- p1 + p2  
4   return(p3)  
5 }  
6  
7 functionName(3)  
8 1237  
9  
10 functionName(3, 6677)  
11 6680
```

- 3 The last value evaluated is returned.
- 4 Statements grouped by "curly braces" or semicolons.

# Some simple exercises to get familiar with R and RStudio

- 1 Create a variable and assign it the value 3
- 2 Print your variable
- 3 Create a function that takes one parameter and returns the square of that value
- 4 Use your function to compute the square of 45
- 5 Print the value of the passed parameter inside the function
- 6 Open the file `library.R` and explain what the function `dumpObject` does

## Q & A time.

Q: Do you know what the death rate around here is?

A: One per person.



## What have we covered?

- Covered a little bit of R's background
- Looked at RStudio, a cross platform GUI for working with R
- Looked at some R basics (variable types and functions)




Next: Look at some of R's built-in datasets and specifically; the iris dataset



## References (1 of 1)

- [1] Michael J. Crawley, [The R Book](#), John Wiley & Sons, 2012.
- [2] CRAN Staff, [What is R?](#),  
<https://www.r-project.org/about.html>, 2017.
- [3] Simon Walkowiak, [Big Data Analytics with R](#), Packt Publishing Ltd., 2016.

# Files of interest

- 
- 1 Software installation