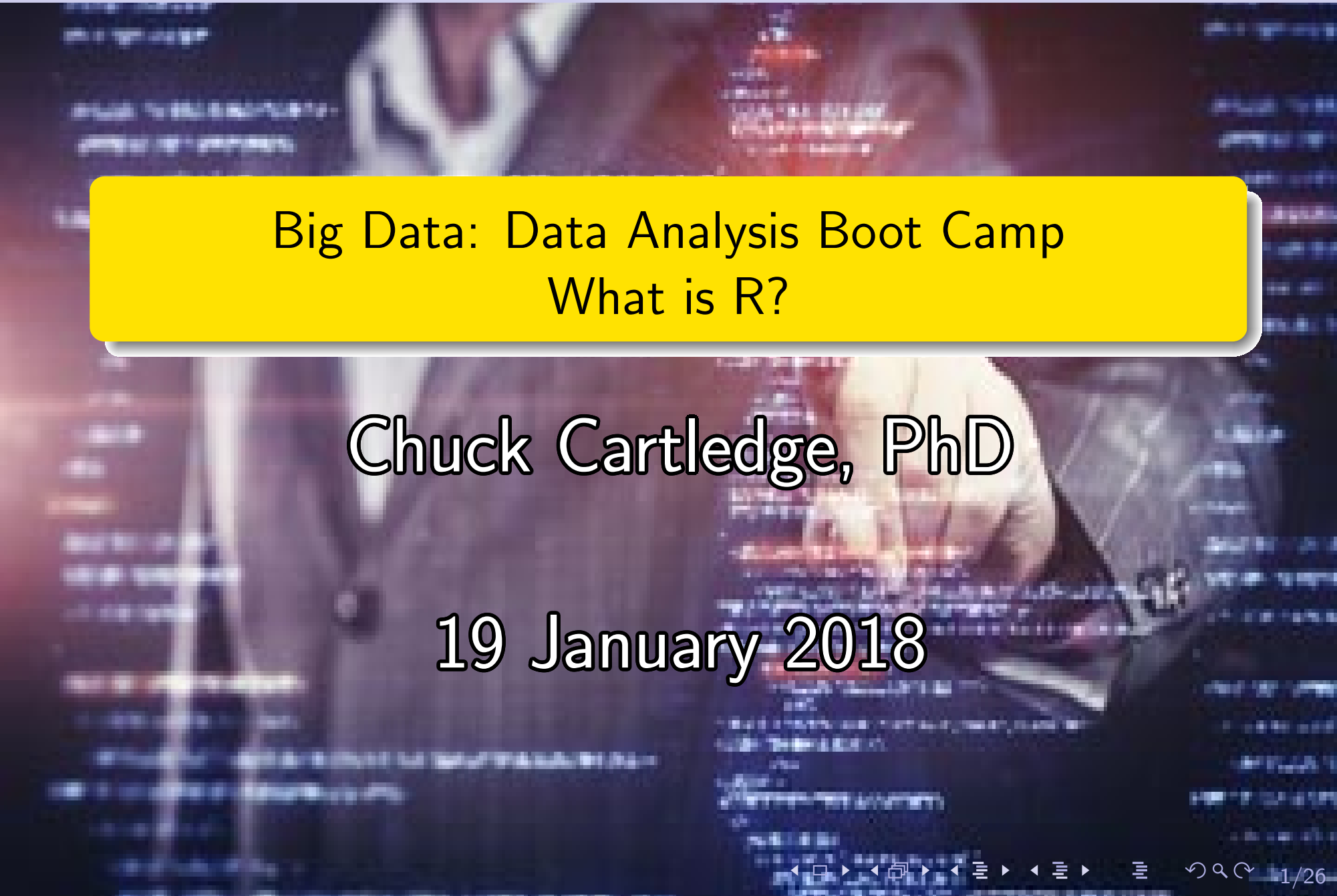


Big Data: Data Analysis Boot Camp
What is R?

Chuck Cartledge, PhD

19 January 2018

Table of contents (1 of 1)

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

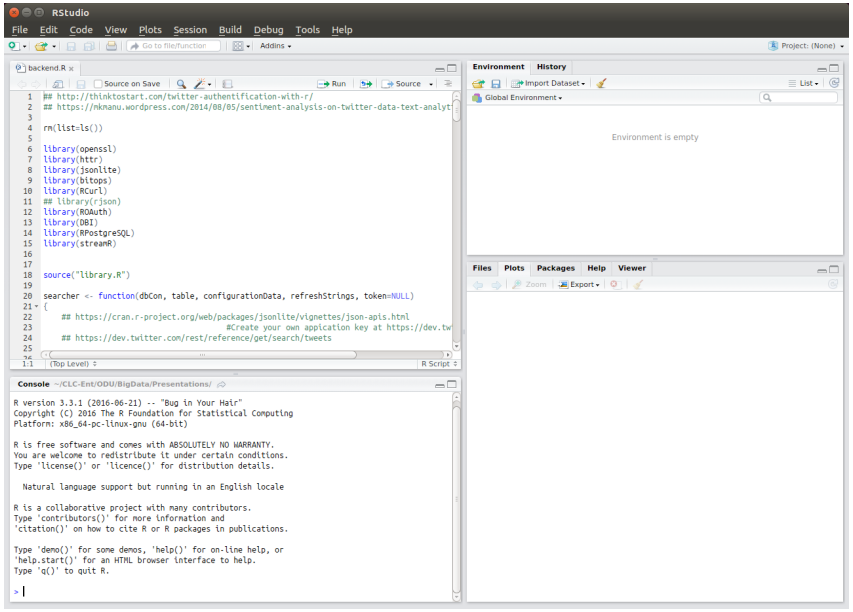
Get the R environment and a command line interface.
Download from: `https://cloud.r-project.org/`
Source code is available for custom OSs.
`https://github.com/wch/r-source`

A complete IDE

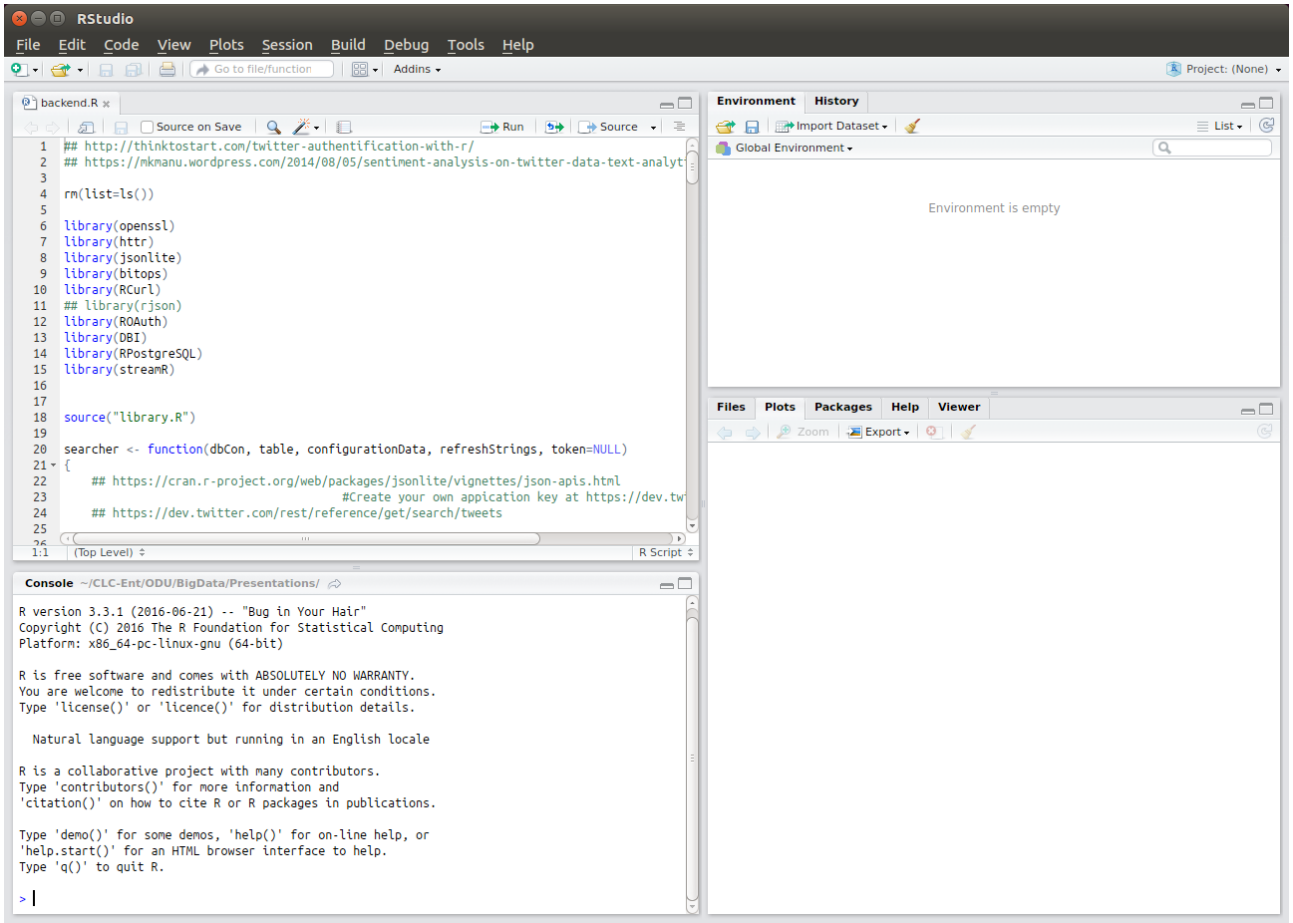
A complete, integrated R development environment.

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.

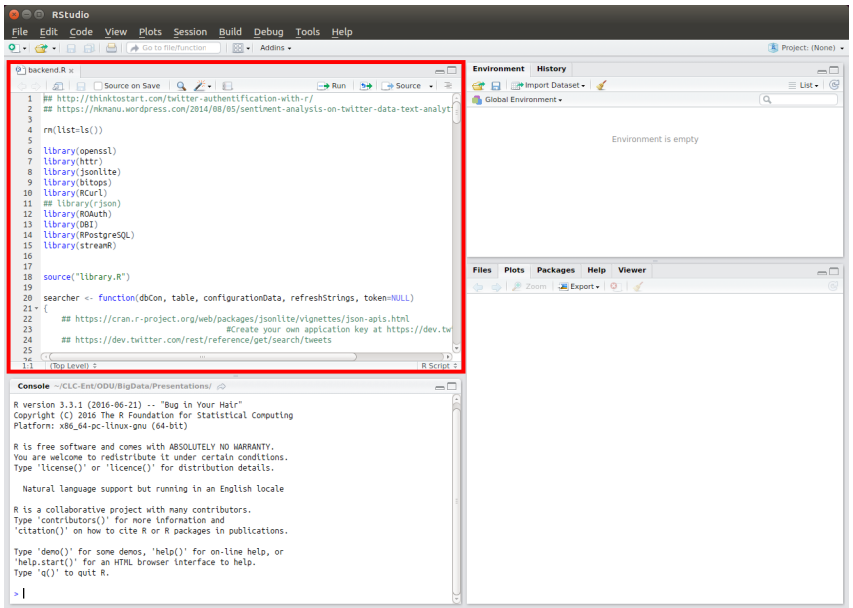


Same image.

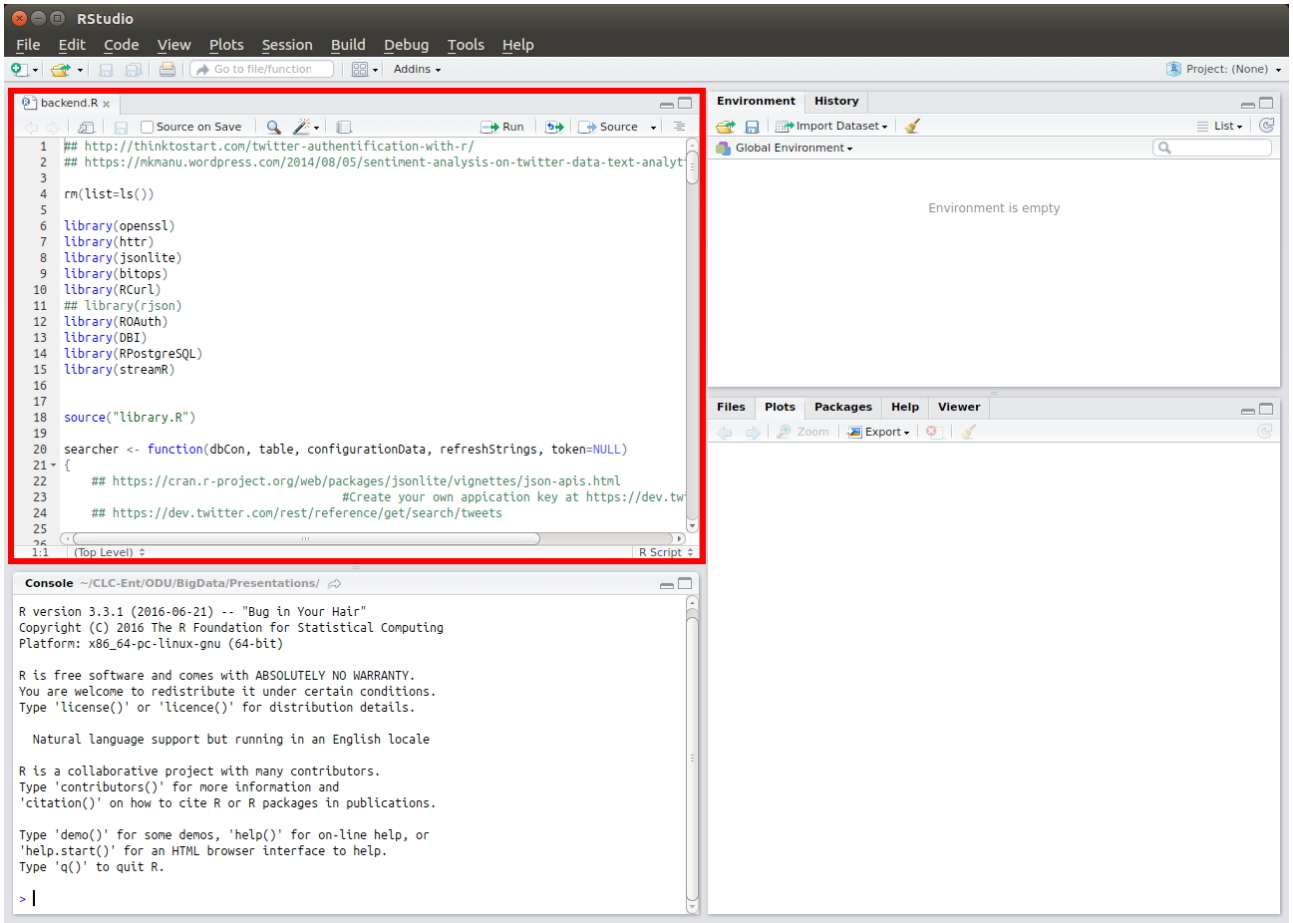

RStudio
File Edit Code View Plots Session Build Debug Tools Help
backend.R
Source on Save
Run
Source
http://thinktostart.com/twitter-authentification-with-r/
https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
rm(list=ls())
library(openssl)
library(httr)
library(jsonlite)
library(bitops)
library(RCurl)
library(rjson)
library(ROAuth)
library(DBI)
library(RPostgreSQL)
library(streamR)
source("library.R")
searcher <- function(dbCon, table, configurationData, refreshStrings, token=NULL)
{
https://cran.r-project.org/web/packages/jsonlite/vignettes/json-apis.html
#Create your own application key at https://dev.tw
https://dev.twitter.com/rest/reference/get/search/tweets
(Top Level)
R Script
Environment History
Import Dataset
List
Global Environment
Environment is empty
Files Plots Packages Help Viewer
Zoom
Export
Console ~/CLC-Ent/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

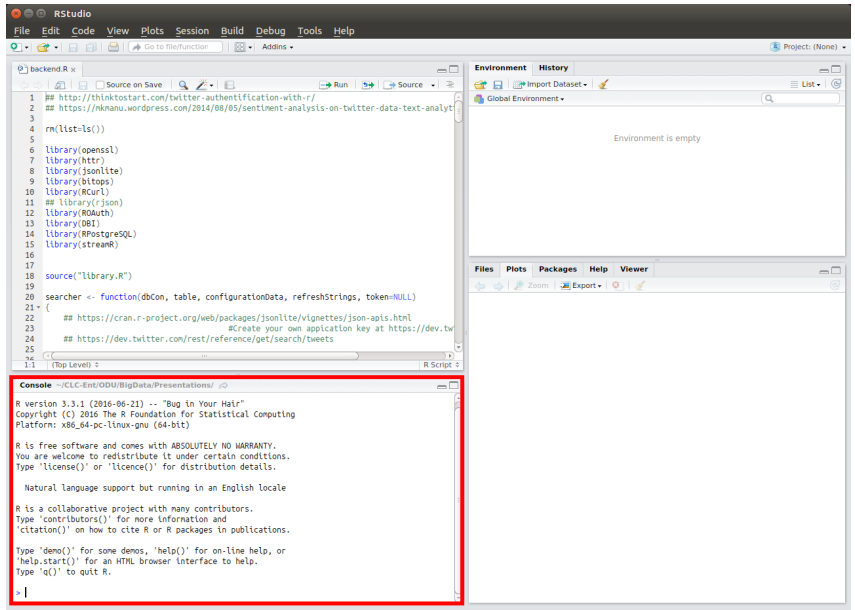


Same image.

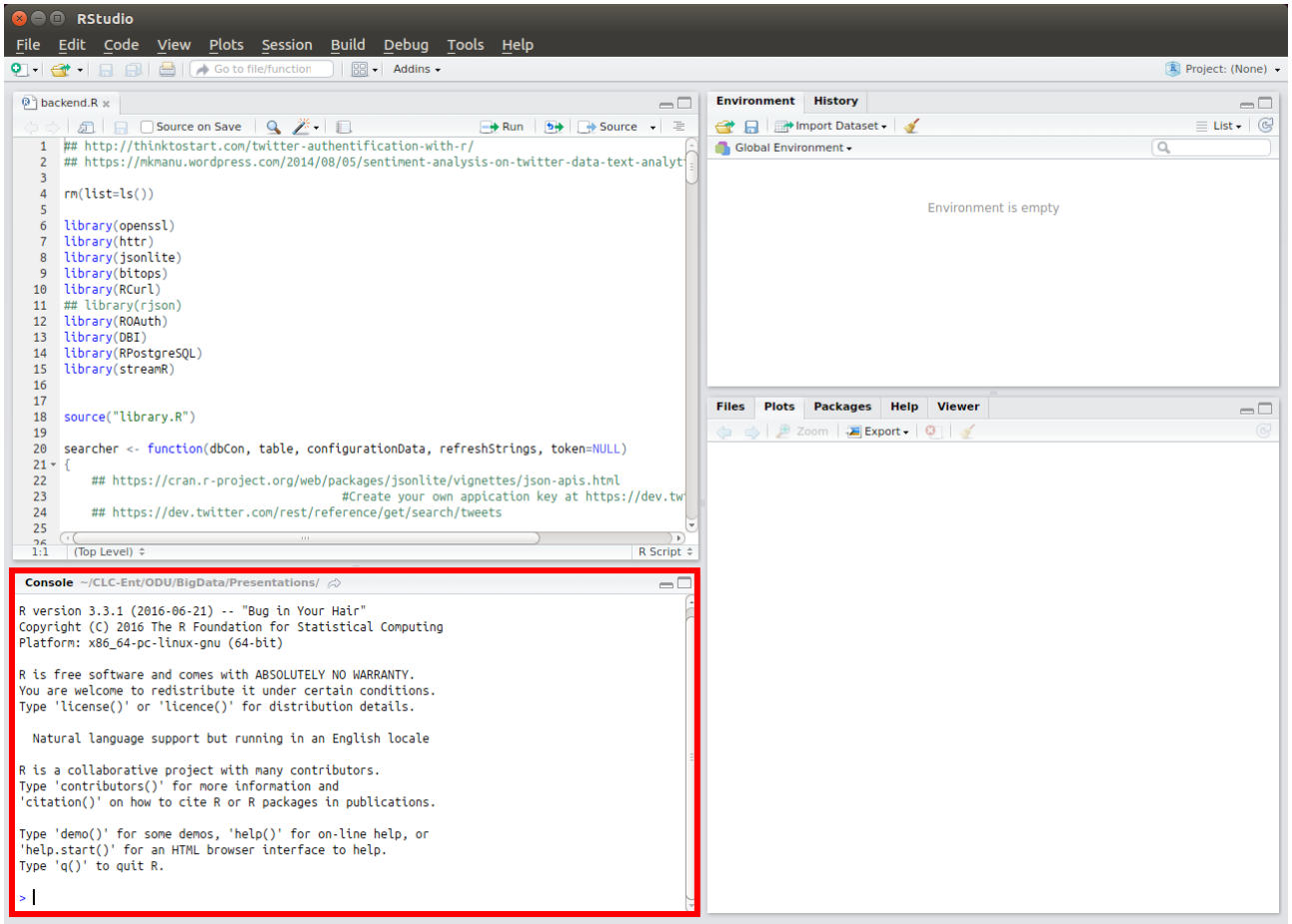

RStudio
File Edit Code View Plots Session Build Debug Tools Help
backend.R
Source on Save
Run
Source
http://thinktostart.com/twitter-authentification-with-r/
https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
rm(list=ls())
library(openssl)
library(httr)
library(jsonlite)
library(bitops)
library(RCurl)
library(rjson)
library(ROAuth)
library(DBI)
library(RPostgreSQL)
library(streamR)
source("library.R")
searcher <- function(dbCon, table, configurationData, refreshStrings, token=NULL)
{
https://cran.r-project.org/web/packages/jsonlite/vignettes/json-apis.html
#Create your own application key at https://dev.tw
https://dev.twitter.com/rest/reference/get/search/tweets
(Top Level)
R Script
Environment History
Import Dataset
List
Global Environment
Environment is empty
Files Plots Packages Help Viewer
Zoom
Export
Console ~/CLC-Ent/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.
>

Console

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

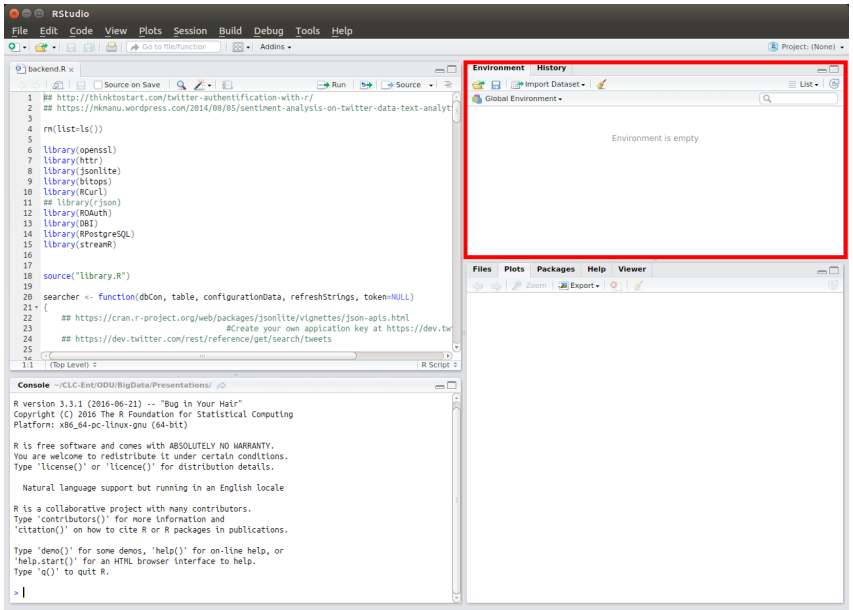


Same image.

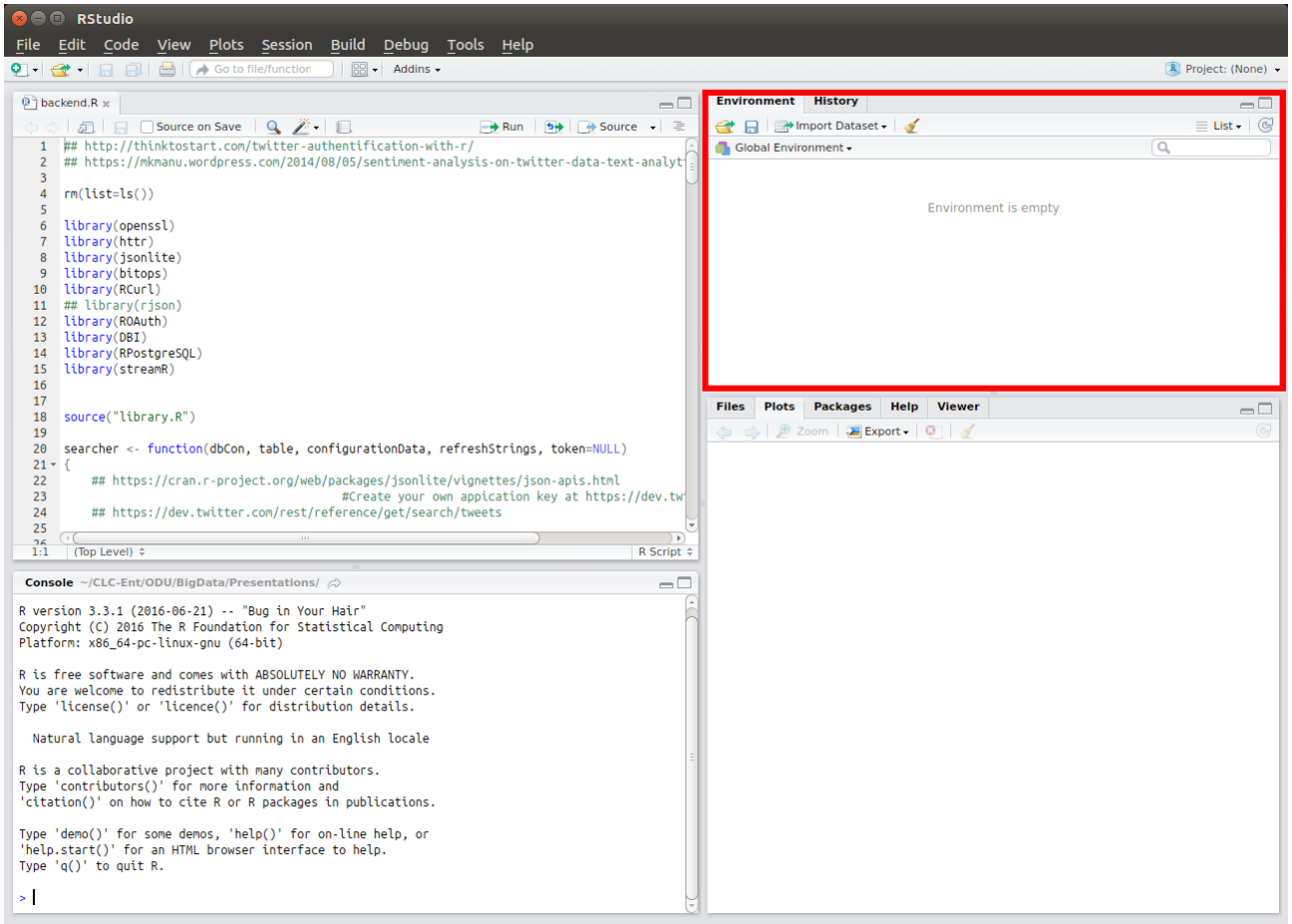

RStudio
File Edit Code View Plots Session Build Debug Tools Help
backend.R
Source on Save
Run
Source
http://thinktostart.com/twitter-authentification-with-r/
https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
rm(list=ls())
library(openssl)
library(httr)
library(jsonlite)
library(bitops)
library(RCurl)
library(rjson)
library(ROAuth)
library(DBI)
library(RPostgreSQL)
library(streamR)
source("library.R")
searcher <- function(dbCon, table, configurationData, refreshStrings, token=NULL)
{
https://cran.r-project.org/web/packages/jsonlite/vignettes/json-apis.html
#Create your own application key at https://dev.tw
https://dev.twitter.com/rest/reference/get/search/tweets
(Top Level)
R Script
Console ~/CLC-Ent/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.
>
Environment History
Import Dataset
List
Global Environment
Environment is empty
Files Plots Packages Help Viewer
Zoom
Export

Variables

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

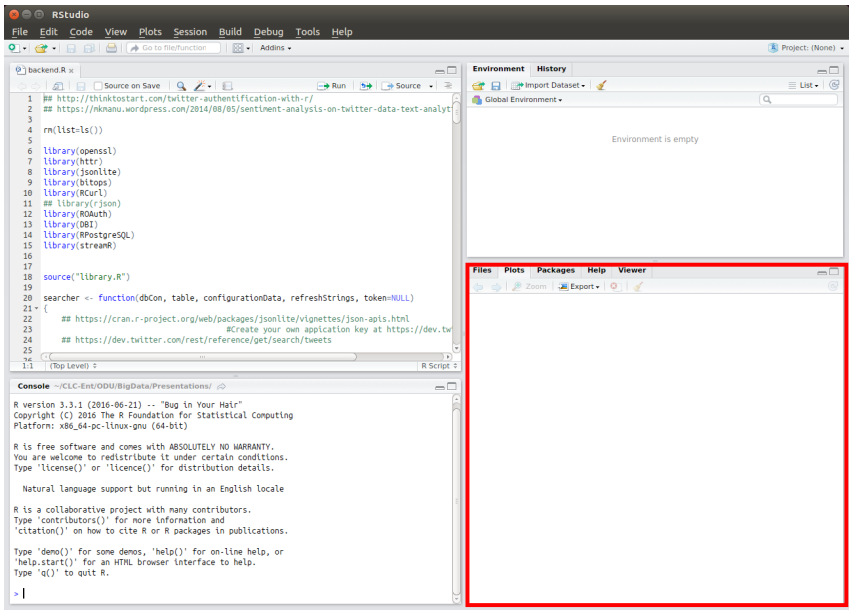


Same image.

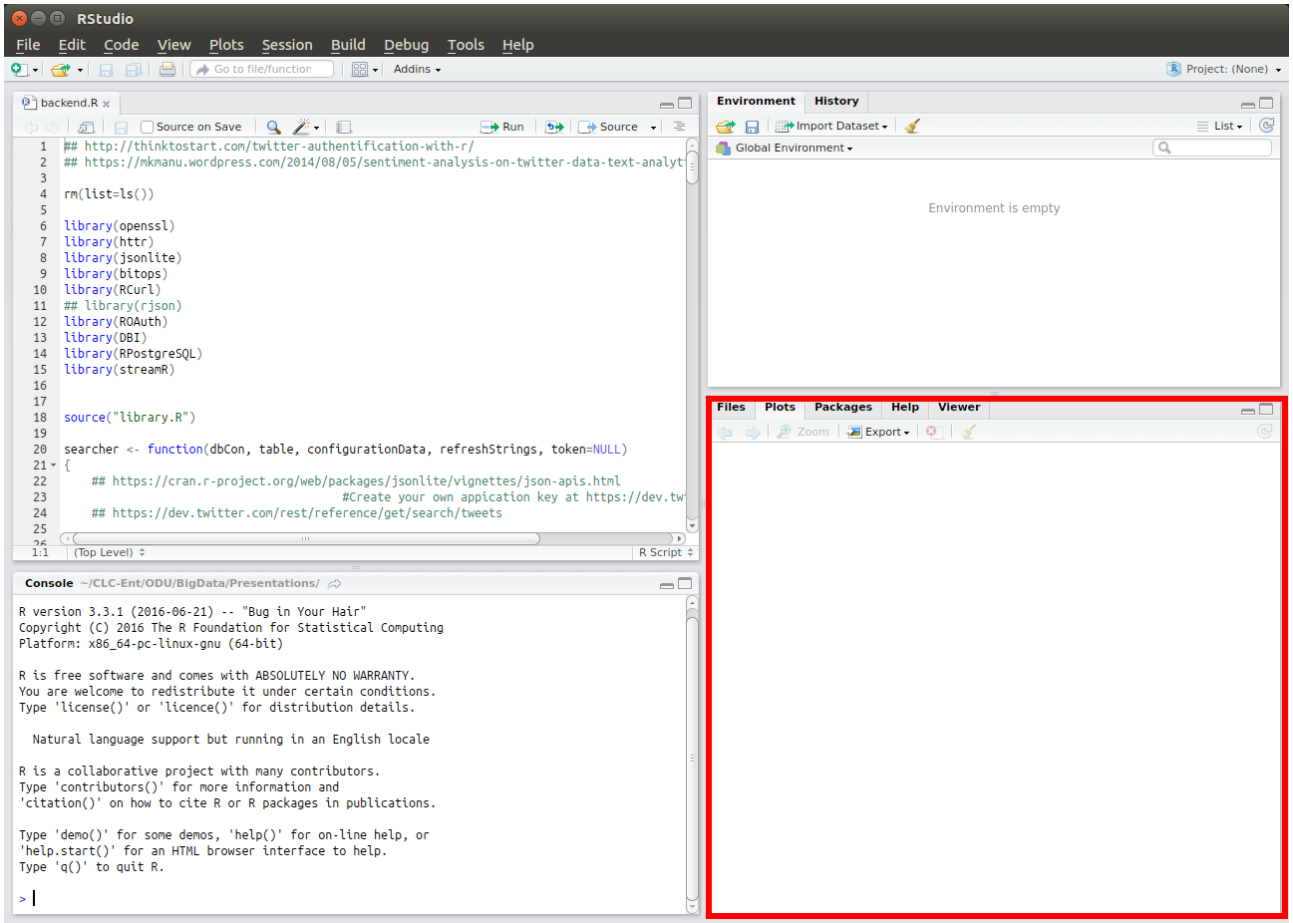

RStudio
File Edit Code View Plots Session Build Debug Tools Help
backend.R
Environment History
Import Dataset
Global Environment
Environment is empty
Files Plots Packages Help Viewer
Export
Console ~/CLC-Ent/ODU/BigData/Presentations/
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 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.

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

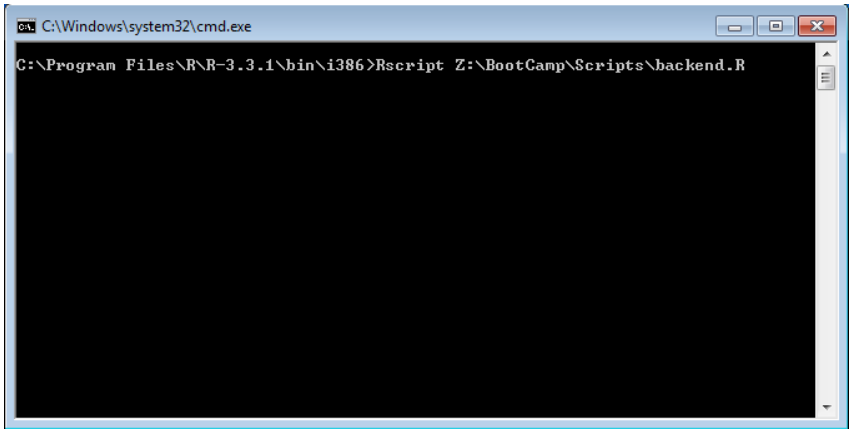


Same image.

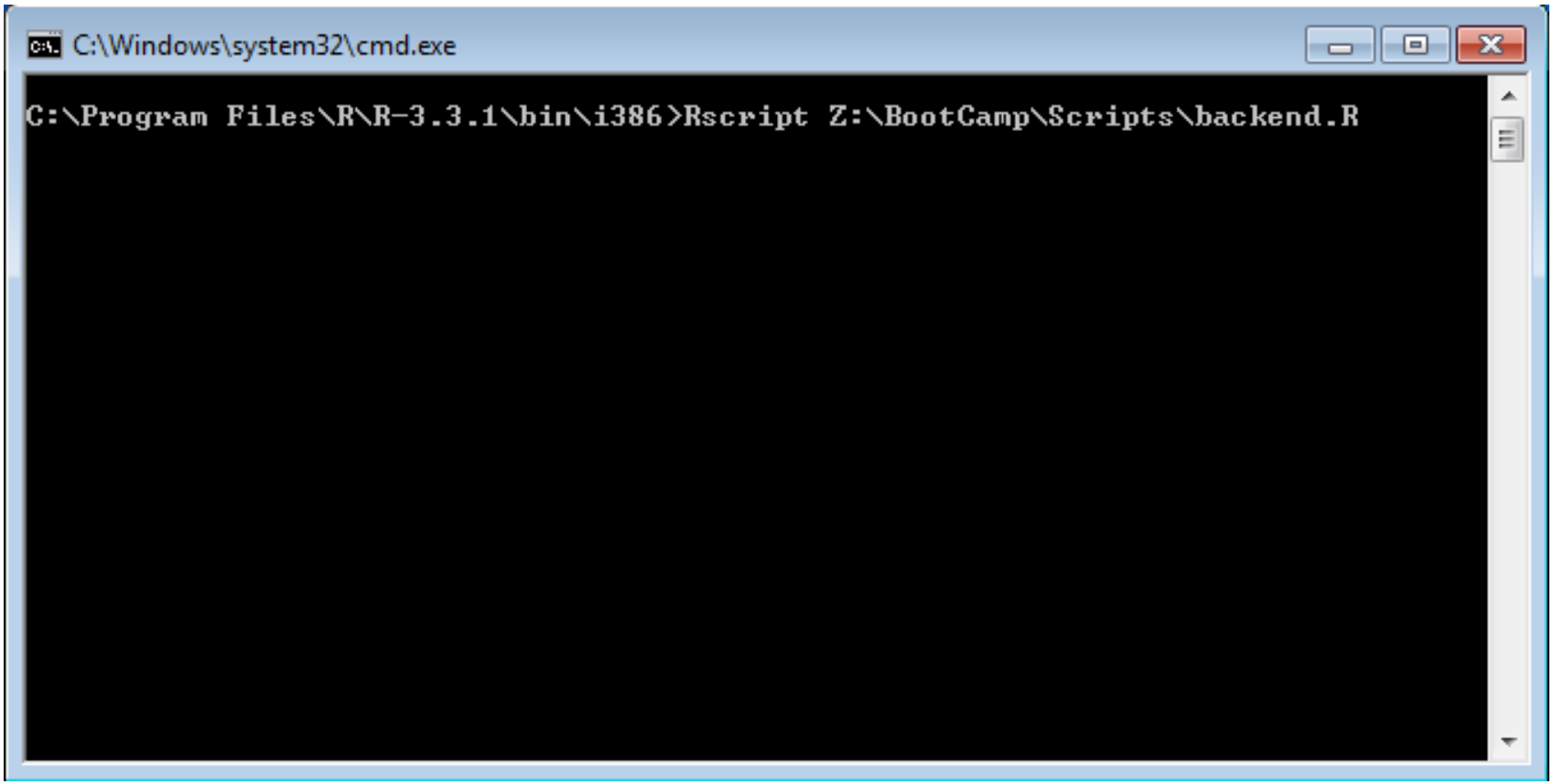

RStudio
File Edit Code View Plots Session Build Debug Tools Help
backend.R
Source on Save
Run
Source
http://thinktostart.com/twitter-authentification-with-r/
https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
rm(list=ls())
library(openssl)
library(httr)
library(jsonlite)
library(bitops)
library(RCurl)
library(rjson)
library(ROAuth)
library(DBI)
library(RPostgreSQL)
library(streamR)
source("library.R")
searcher <- function(dbCon, table, configurationData, refreshStrings, token=NULL)
{
https://cran.r-project.org/web/packages/jsonlite/vignettes/json-apis.html
#Create your own application key at https://dev.tw
https://dev.twitter.com/rest/reference/get/search/tweets
(Top Level)
R Script
Environment History
Import Dataset
List
Global Environment
Environment is empty
Files Plots Packages Help Viewer
Zoom
Export
Console ~/CLC-Ent/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.
>

Starting an R script in the background

The image shows a Windows environment.
A *nix environment command is:
`Rscript 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

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

2. Based on pattern matching

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

3. Looking for a specific item

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

4. About a specific item

```
?lm
??lm
```

5. Example of a function

```
example(lm)
```

6. Source code for a function

```
lm
```

7. Demonstration of a function

```
demo(persp)
```

8. Demonstration of a function

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

9. Contents of a library

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

10. Install a new library

```
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)

```
variable <- 3
```

3. Vector of values

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

4. Matrix of values

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

5. Access matrix values

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

6. Array of values

```
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)

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

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

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

3. Saving variables

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

4. Loading variables

```
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

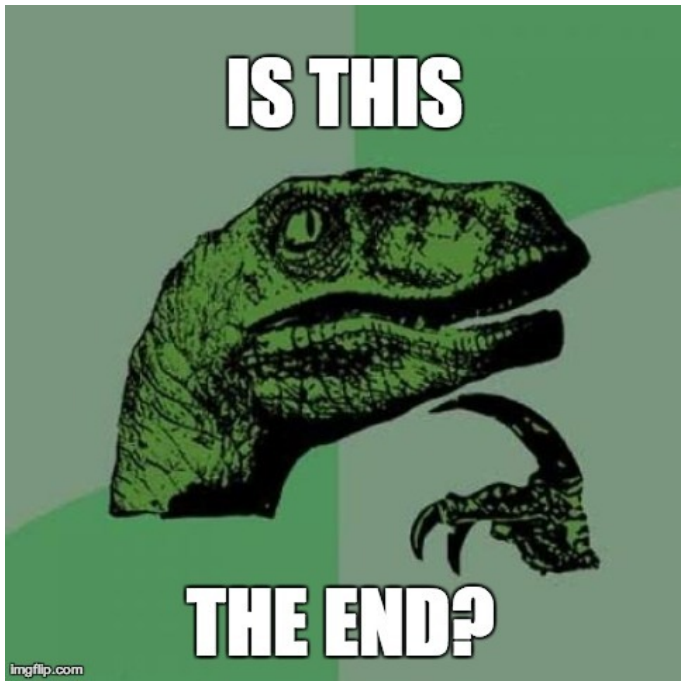
1. Create a variable and assign it the value 3
2. Print your variable
3. Create a function that takes one parameter and returns the square of that value
4. Use your function to compute the square of 45
5. Print the value of the passed parameter inside the function
6. Open the file `library.R` and explain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