

Big Data: Data Wrangling Boot Camp

R Sentiment Analysis

Chuck Cartledge, PhD

18 September 2016

Table of contents I

1 Introduction

2 Preview

3 RStudio

4 Q & A

5 Conclusion

What are we going to cover?

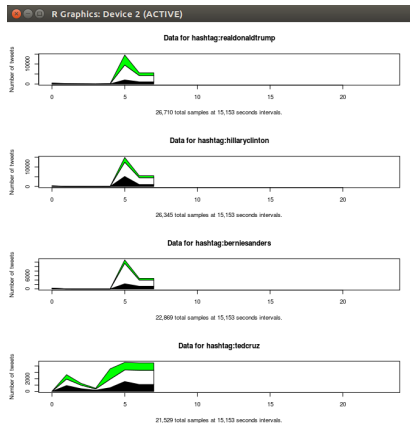
- Look to the future
- Data wrangle tweets using R
- Wrap-up the boot camp



Things that will be happening today

Things that we will be doing.

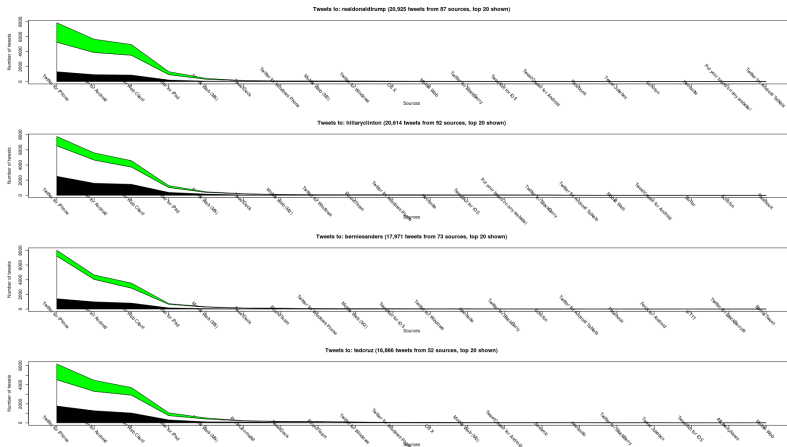
- 1 Data wrangle tweets using R
- 2 Conduct sentiment analysis on tweets
- 3 Look at the sentiments in different ways



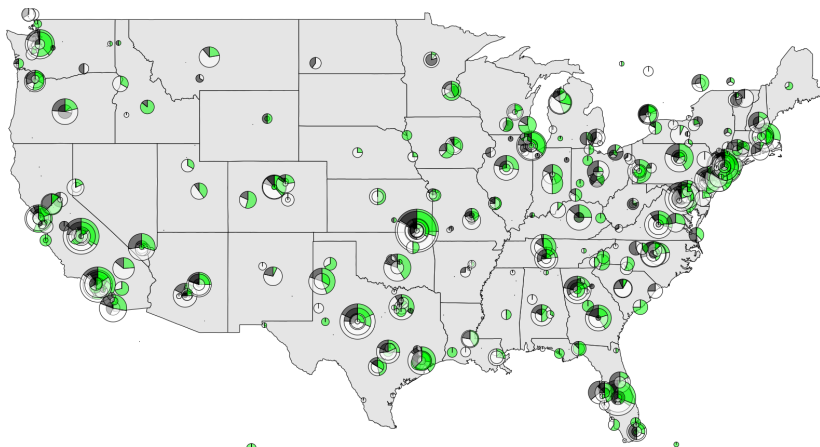
Sentiments over time.

Things that will be happening today

Sentiment by sending device



Sentiment by geographic location

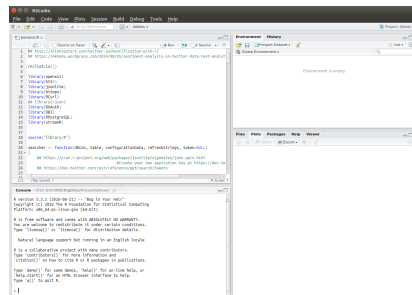


A complete IDE

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.

The screenshot shows the RStudio interface. The main editor window contains the following R code:

```

1 ## http://thinkstart.com/twitter-authentication-with-r/
2 ## https://nmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=ls())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(biteps)
10 library(RCurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(DBI)
14 library(RPostgreSQL)
15 library(streamR)
16
17
18 source("library.R")
19
20 searcher <- function(dbConn, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-apis.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 window at the bottom displays the R startup message:

```

R version 3.3.1 (2016-06-21) -- "Bug In Your Hat"
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.

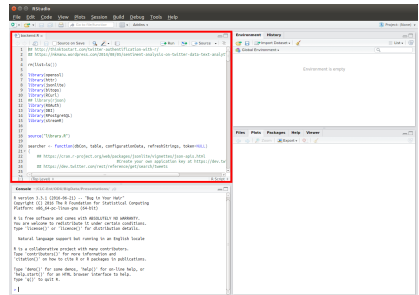
> |

```

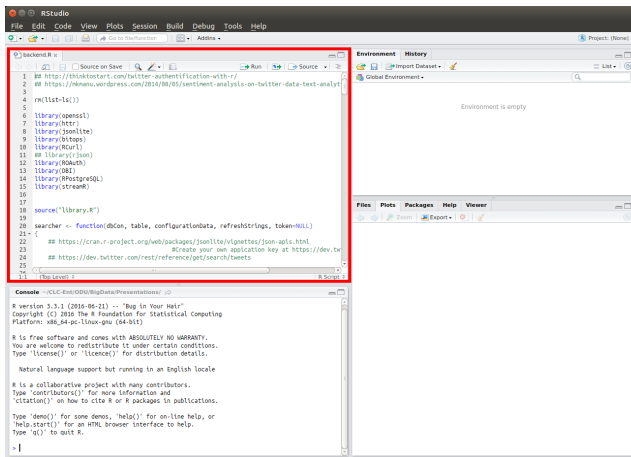

Basic how-tos

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

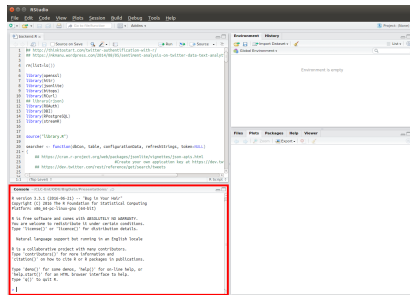


Same image.

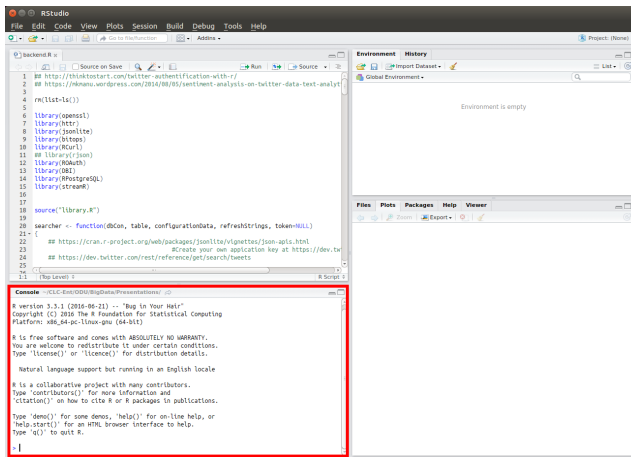


Console

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



Same image.



The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. The left sidebar contains the Environment and History panels, which are currently empty. The right sidebar contains the Files, Plots, Packages, Help, and Viewer panels. The main window is divided into two panes. The top pane shows the source editor with the following R code:

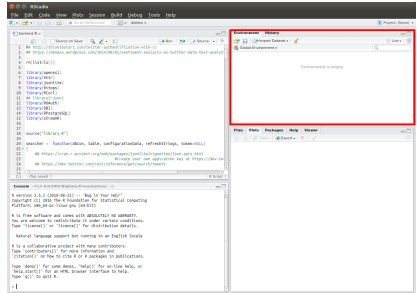
```
1 ## http://thinkstart.com/twitter-authentication-with-r/  
2 ## https://nkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt  
3  
4 rm(list=ls())  
5  
6 library(openssl)  
7 library(httr)  
8 library(jsonlite)  
9 library(biteps)  
10 library(RCurl)  
11 ## library(rjson)  
12 library(ROAuth)  
13 library(DBI)  
14 library(RPostgreSQL)  
15 library(streamR)  
16  
17  
18 source("library.R")  
19  
20 searcher <- function(dbConn, table, configurationData, refreshStrings, token=NULL)  
21 {  
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-apis.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  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100
```

The bottom pane shows the R console with the following output:

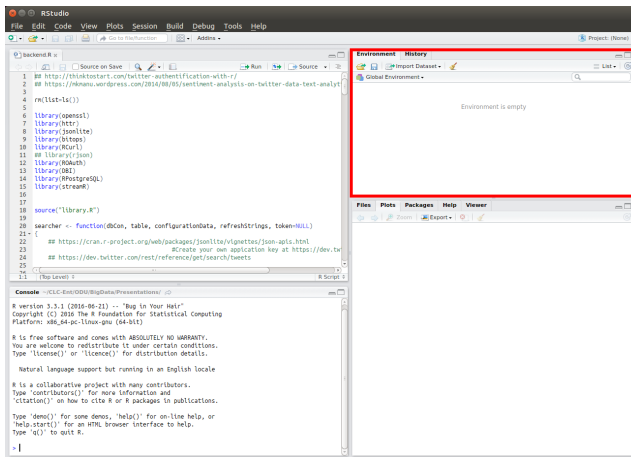
```
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"  
Copyright (c) 2016 The R Foundation for Statistical Computing  
Platform: x86_64-pc-linux-gnu (64-bit)  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' 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

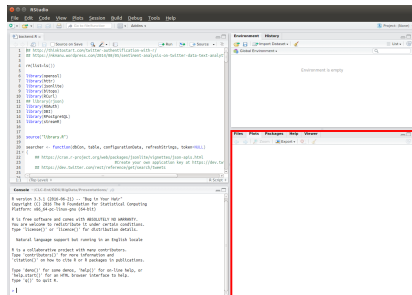


Same image.

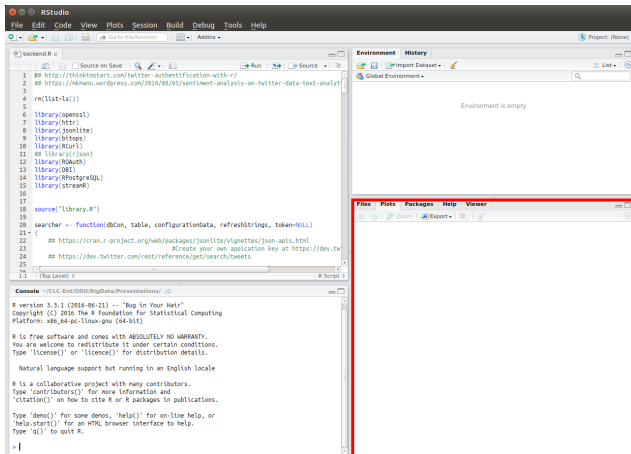


Tabbed display

- 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.

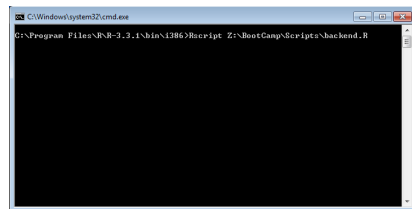


Starting an R script in the background

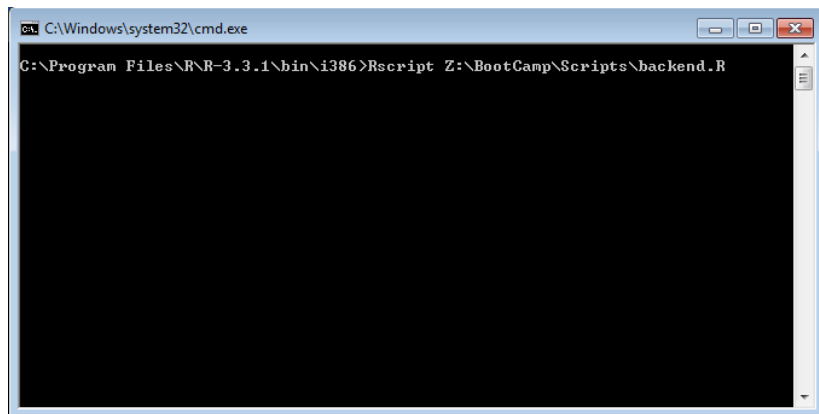
The image shows a Windows environment.

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
```

Q & A time.

“‘The Answer to the Great Question . . . Of Life, the Universe and Everything . . . is . . . forty-two,’ said Deep Thought, with infinite majesty and calm.”

Douglas Adams, The Hitchhiker's Guide to the Galaxy



What have we covered?

- A preview of today's activities
- An overview of RStudio



Next: Hands on analyzing tweets with R.