

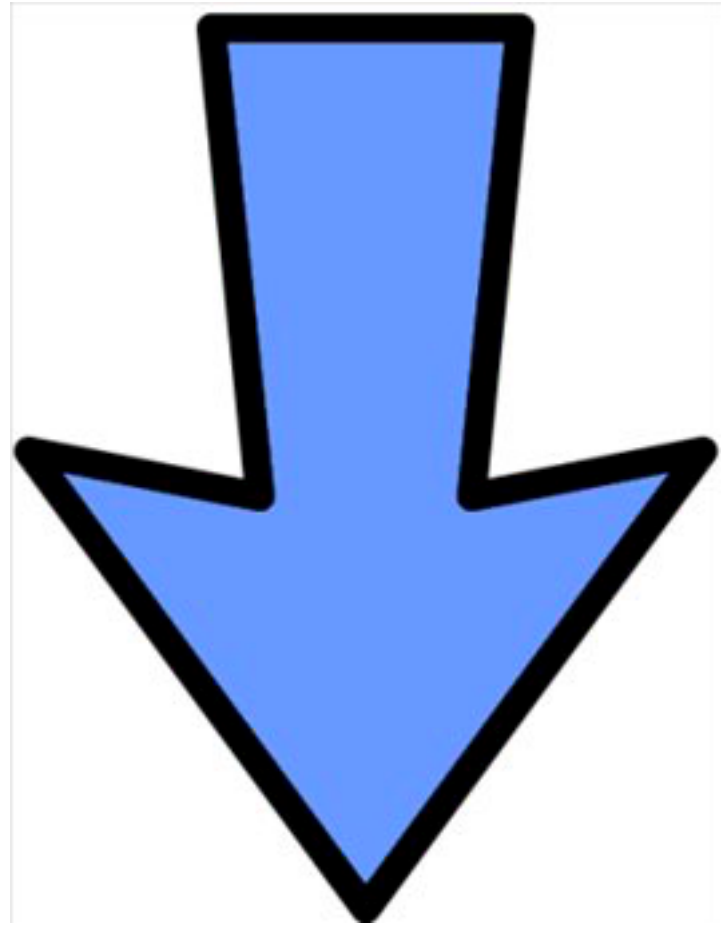
Twitter Web Scraping Tutorial

Getting your data from Twitter to R

By (Thomas) William Bell

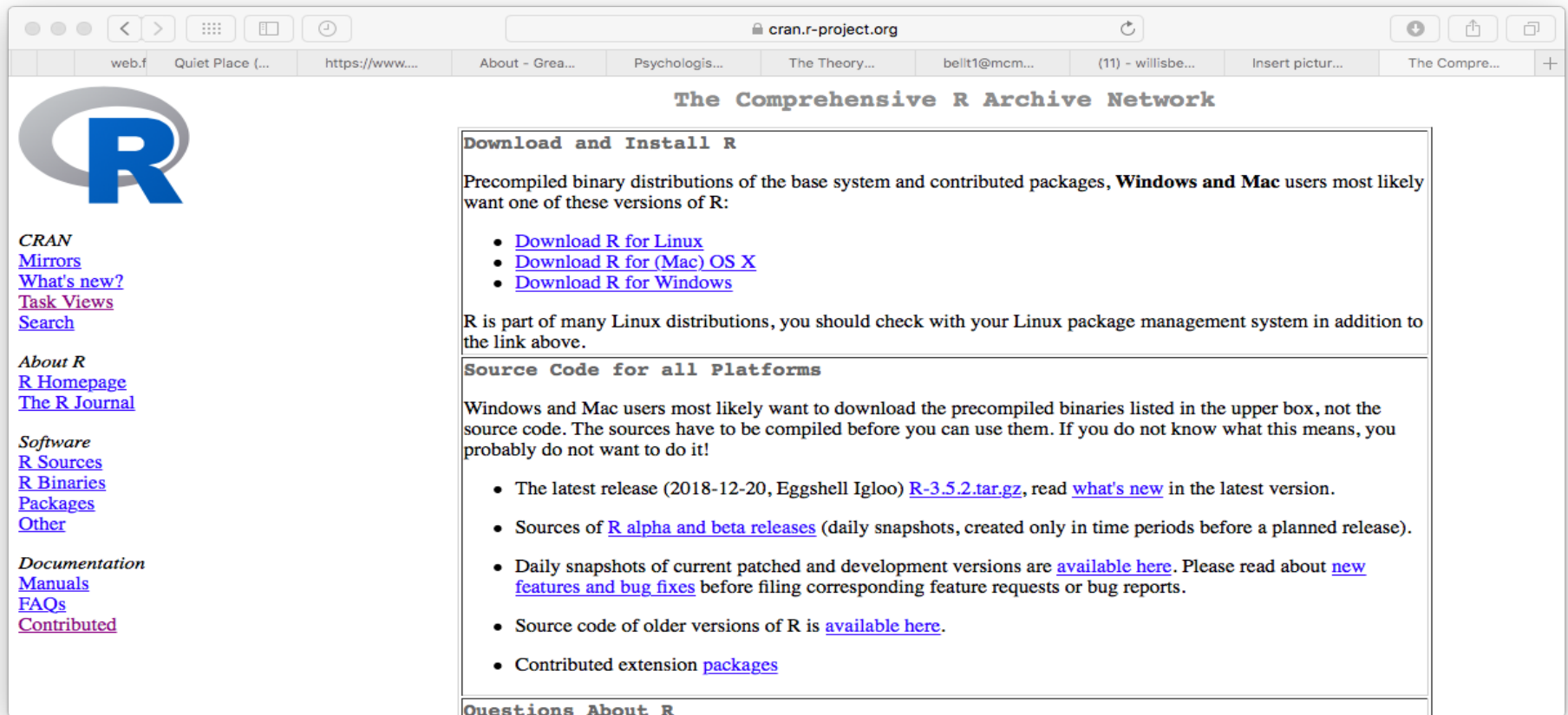
The Plan

1. Get R and R Studio
2. Set up a twitter account
3. Get developer permissions
4. Say you're building an 'app'
5. Install the TwitterR package
6. Scrape data



Getting R

1. Go to <https://cran.r-project.org>



The screenshot shows a web browser window with the address bar displaying cran.r-project.org. The browser's tab bar shows several open tabs, including "web.f", "Quiet Place (...)", "https://www....", "About - Grea...", "Psychologis...", "The Theory...", "bellt1@mcm...", "(11) - willisbe...", "Insert pictur...", and "The Compre...". The main content area of the browser displays the CRAN website. At the top, the CRAN logo is on the left, and the title "The Comprehensive R Archive Network" is centered. Below the logo, there are links for "CRAN", "Mirrors", "What's new?", "Task Views", and "Search". To the right of the logo, there is a section titled "Download and Install R" which contains a paragraph about precompiled binary distributions for Windows and Mac users, followed by a bulleted list of links: "Download R for Linux", "Download R for (Mac) OS X", and "Download R for Windows". Below this, a paragraph states that R is part of many Linux distributions and suggests checking with a Linux package management system. Further down, there is a section titled "Source Code for all Platforms" which contains a paragraph about downloading precompiled binaries versus source code, followed by a bulleted list of links: "The latest release (2018-12-20, Eggshell Igloo) R-3.5.2.tar.gz", "Sources of R alpha and beta releases", "Daily snapshots of current patched and development versions are available here", "Source code of older versions of R is available here", and "Contributed extension packages". At the bottom of the page, there is a section titled "Questions About R".

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

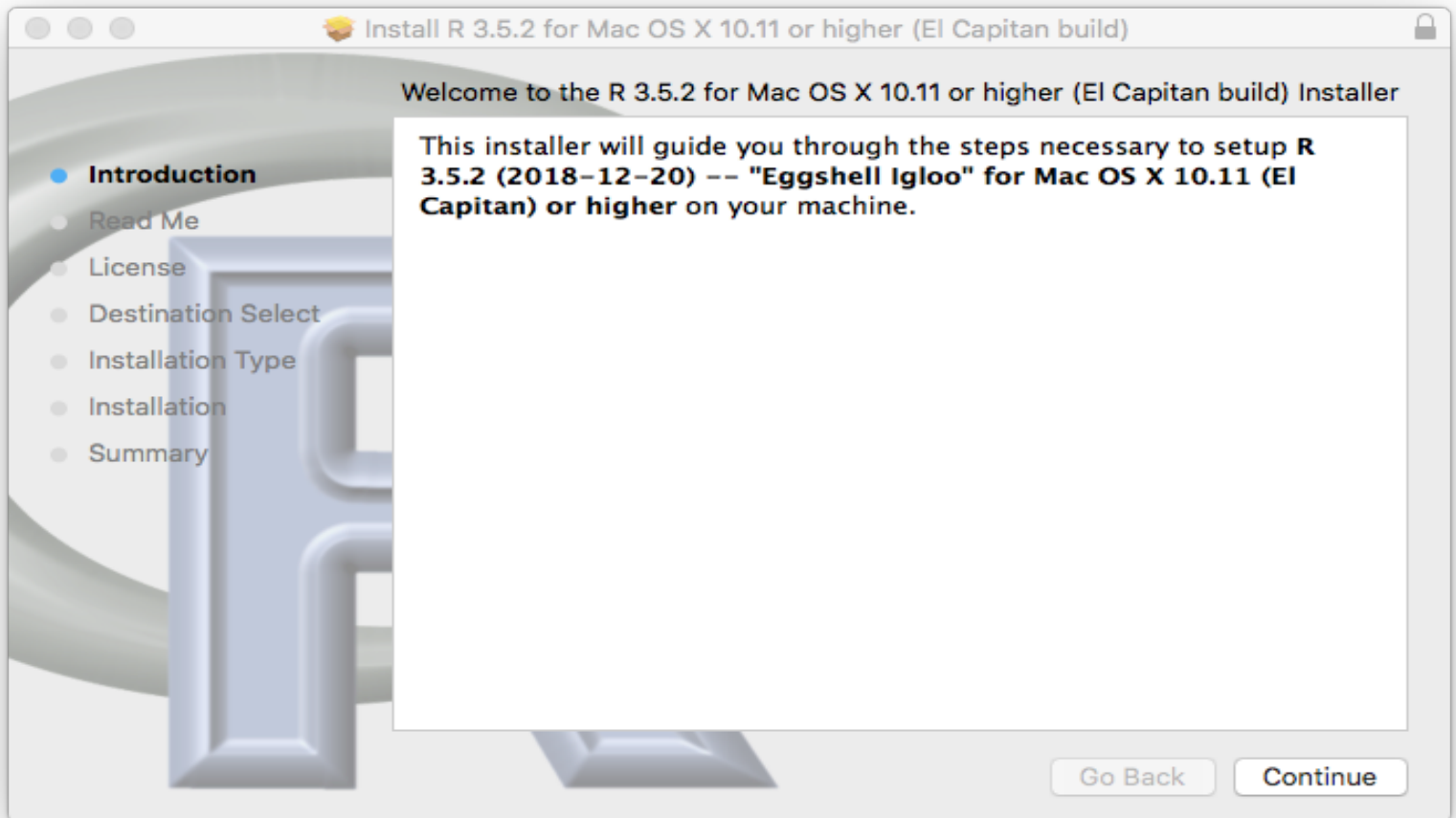
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2018-12-20, Eggshell Igloo) [R-3.5.2.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

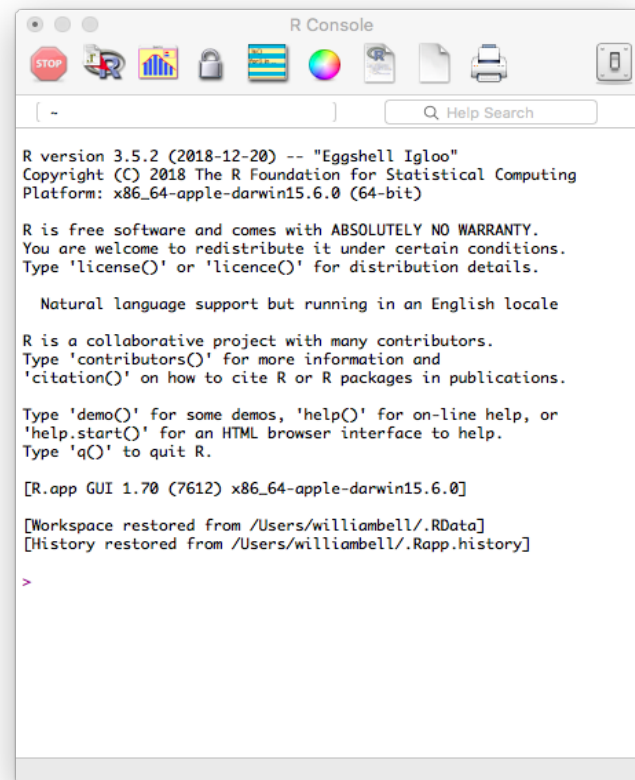
Getting R, cont.

2. Use the installer on a Mac



Getting R, cont.

The default R console looks like this (ugly and annoying to use):



```
R Console

[ - ] [ Help Search ]

R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

[R.app GUI 1.70 (7612) x86_64-apple-darwin15.6.0]

[Workspace restored from /Users/williambell/.RData]
[History restored from /Users/williambell/.Rapp.history]

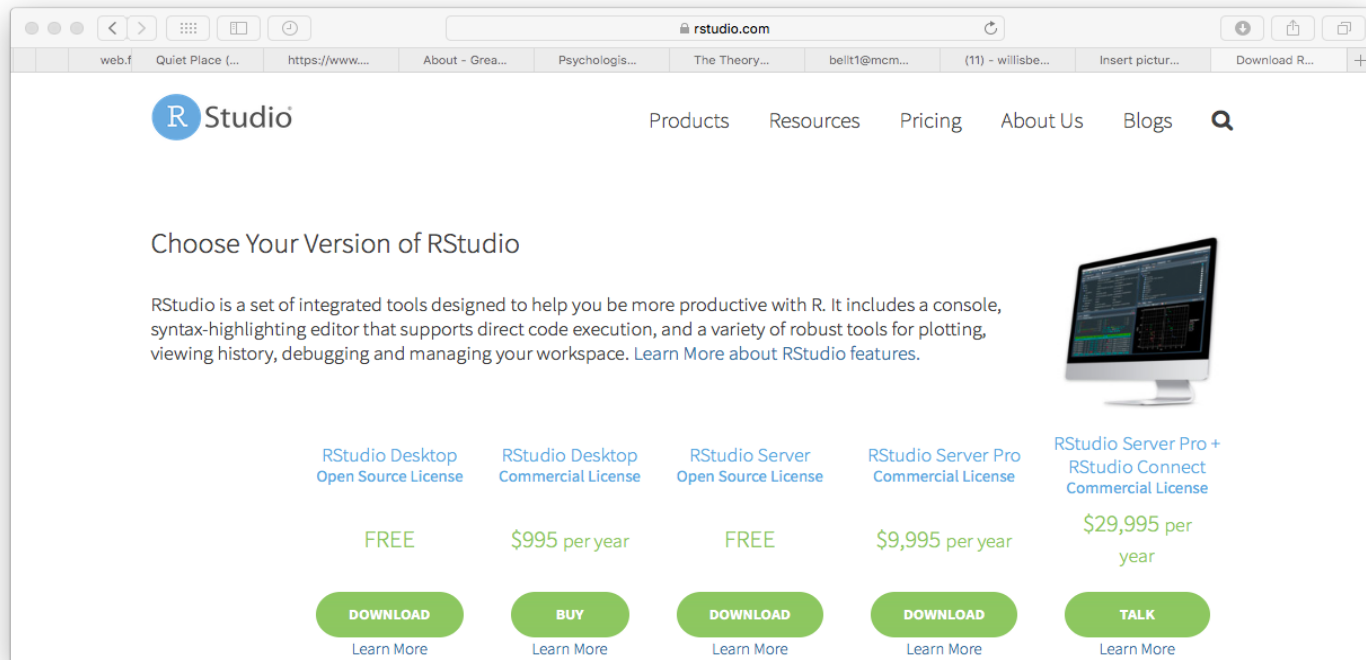
>
```

Getting R, cont.

3. Go to R Studio's website:

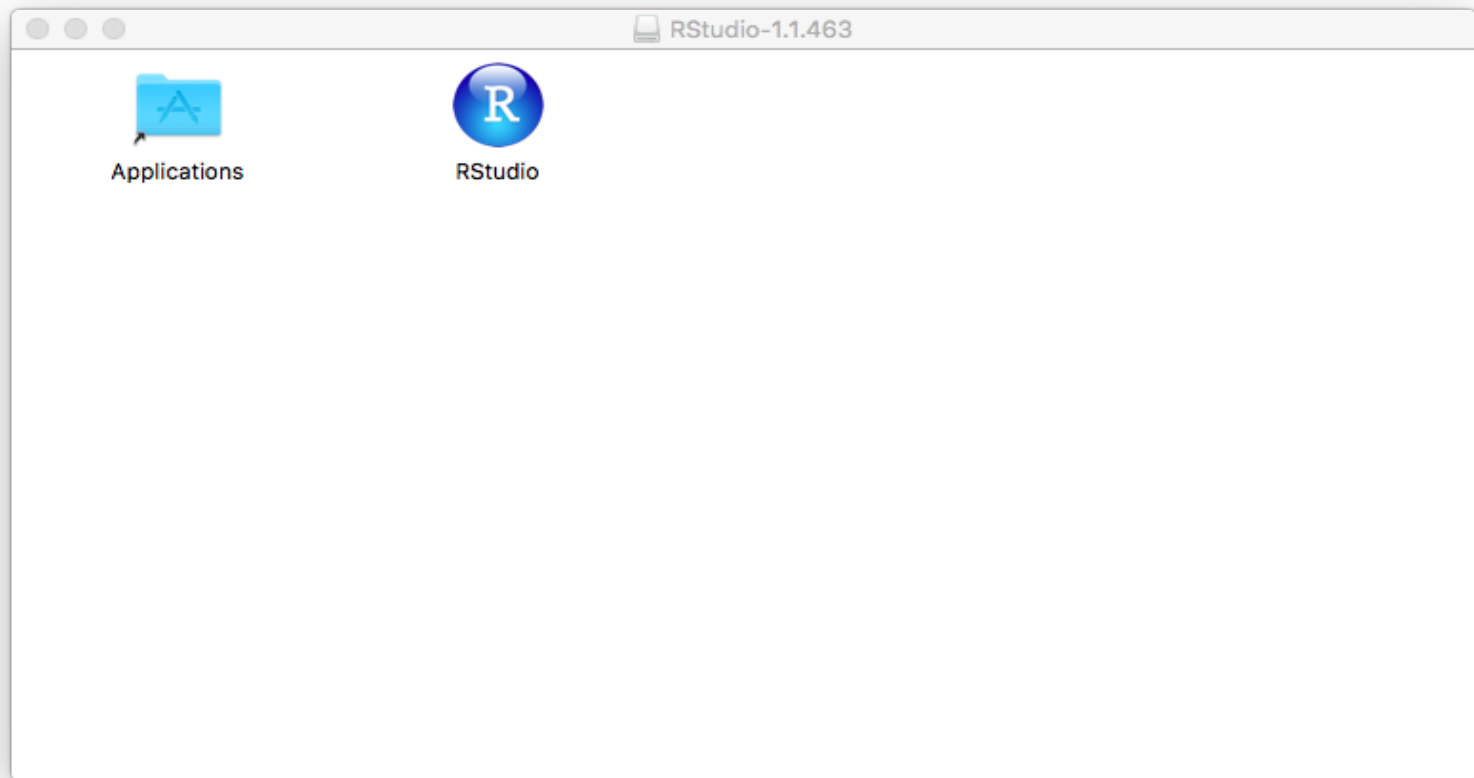
<https://www.rstudio.com/products/rstudio/download/>

Obviously get the free one.



Getting R, cont.

4. On a Mac, just open the file and drag it to the Applications folder

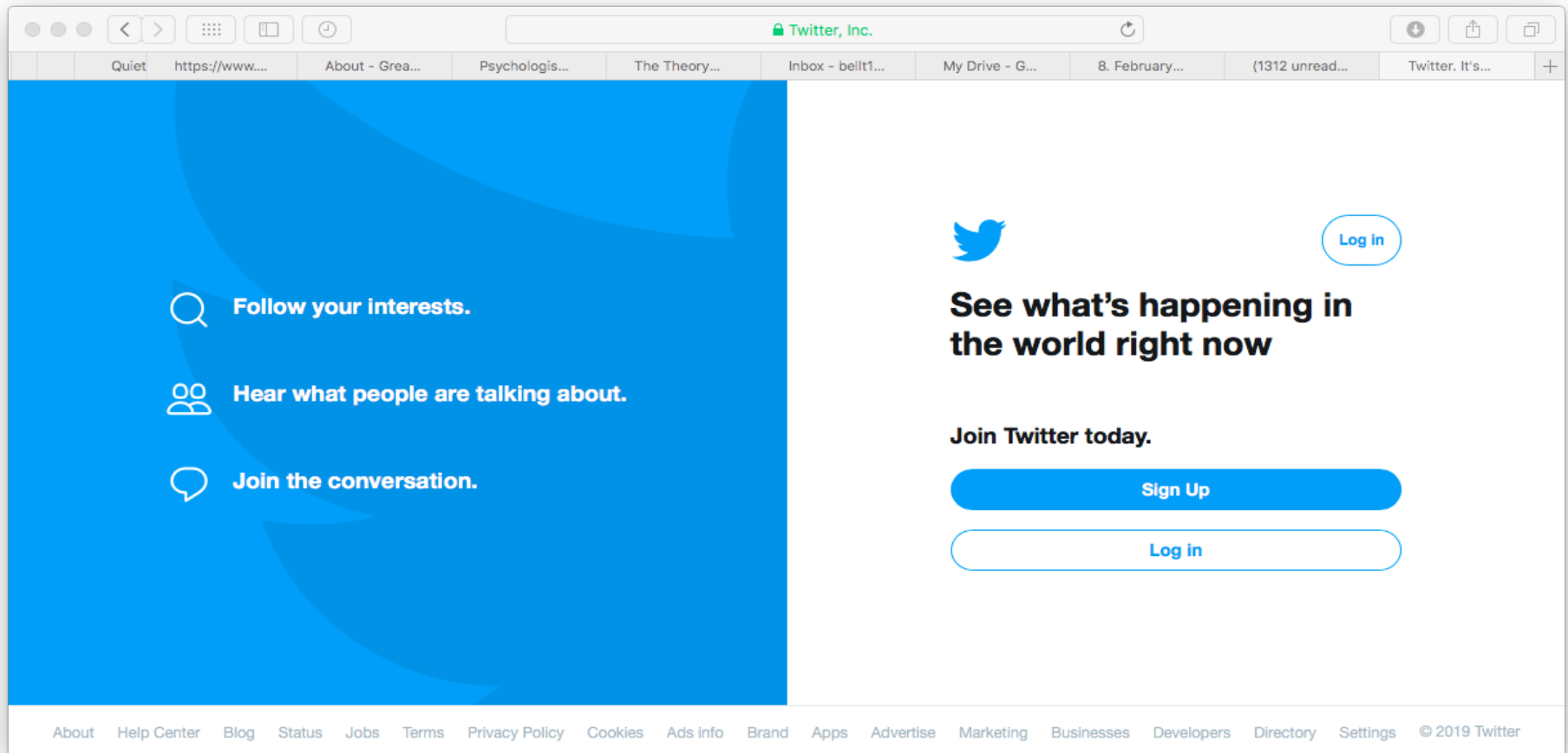




Step one done!

Set up a twitter account

1. Sign up





Next

Create your account

lolcat.lol

10/50

ollyoxenfree1997@yahoo.com

[Use phone instead](#)

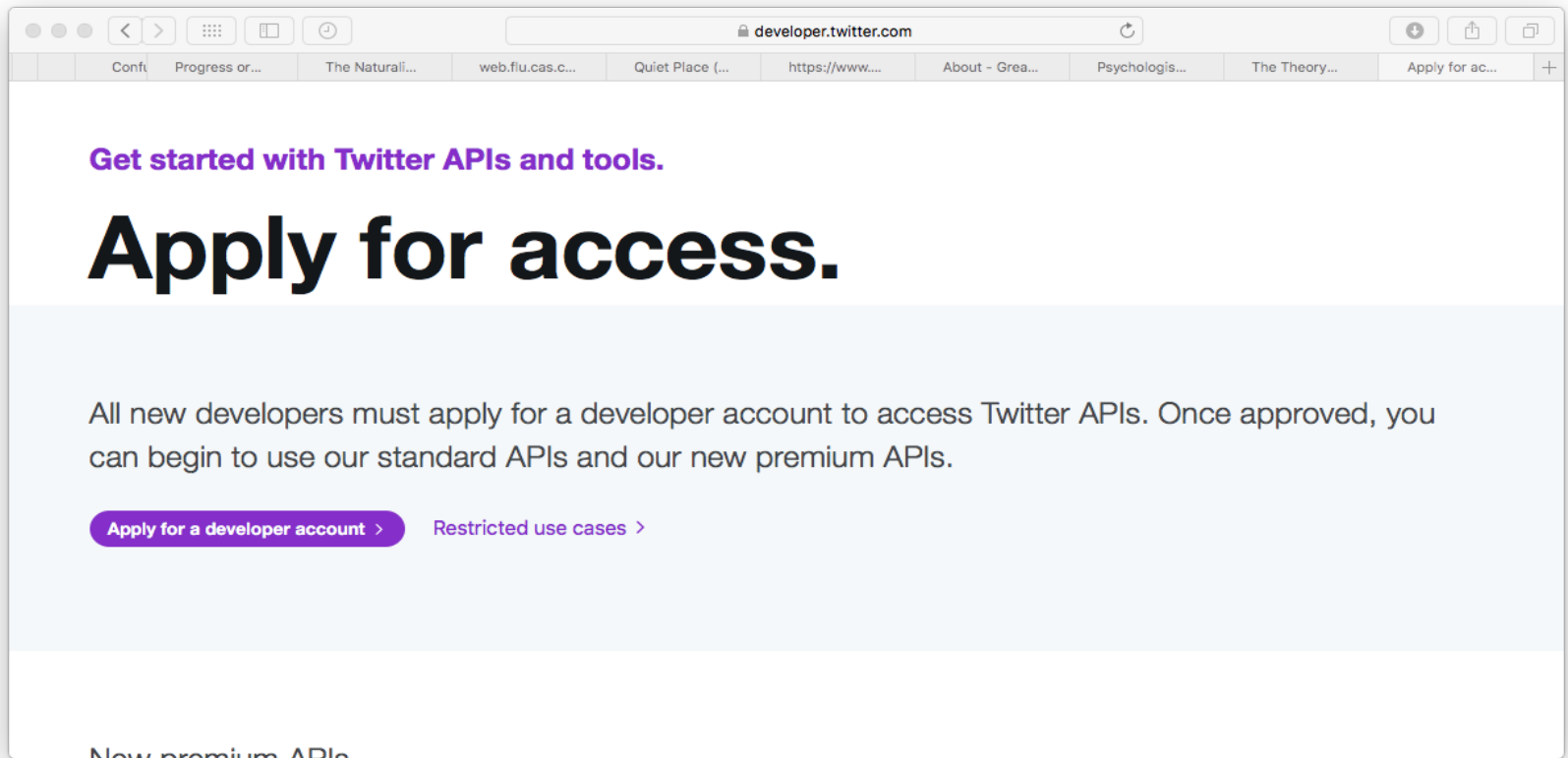


Step two done!

Get Developer Permissions

1. Go to

<https://developer.twitter.com/en/apply-for-access>



Get Developer Permissions, cont.

2. Apply for permission

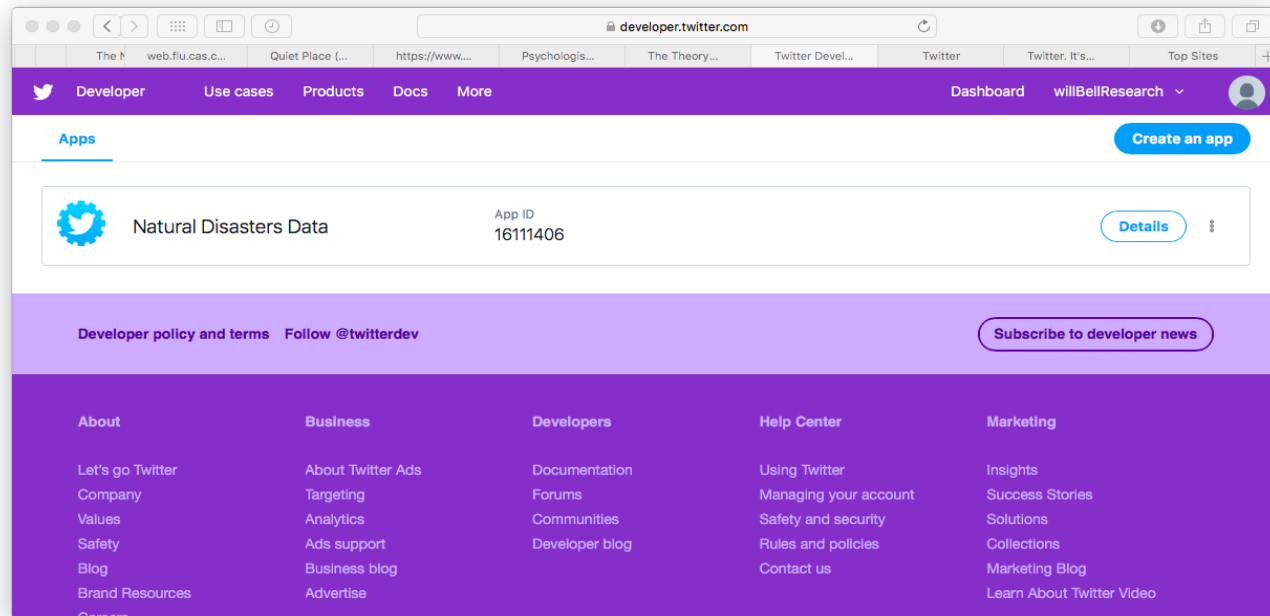
- Requires an explanation of your intent, etc.
- Application is processed automatically.



Step three done!

Say you're building an 'app'

1. Go to the Twitter API website (you're already on it when you applied for access), hover over your username, and choose 'Apps' and then create App

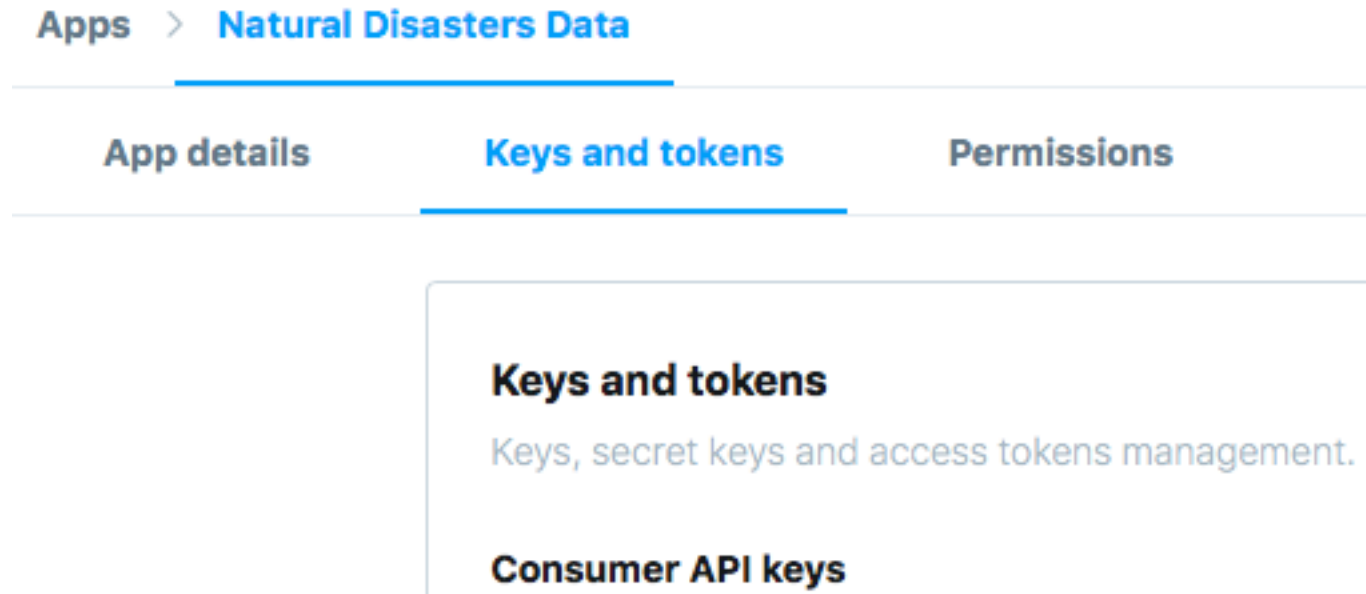


Say you're building an 'app', cont.

2. More questions that are processed automatically

Say you're building an 'app', cont.

3. Go into your app and click on the tab 'Keys and Access Tokens'



Say you're building an 'app', cont.

4. Generate the access tokens

Access token & access token secret

1050200743246991360-H48qUELYp2eM

qAlqm9Ciwg0zak6wofB38SY3T7S61JqBl

Read and write (Access level)

Revoke

Regenerate

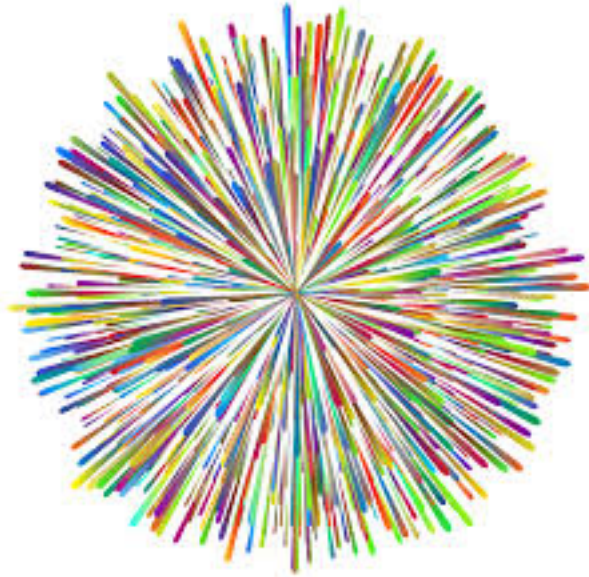


Step four done!

Install TwitterR

Open R Studio and in the console, type the following line:

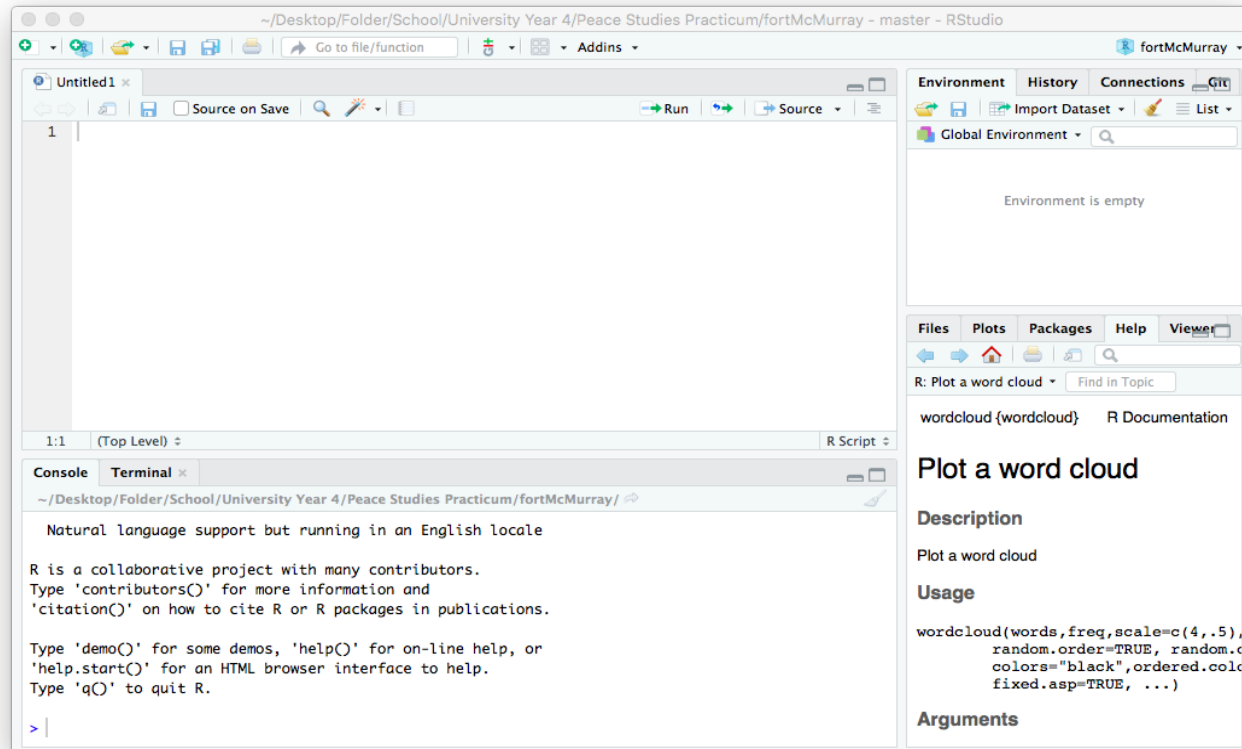
```
> install.packages("twitterR")
```



Step five done!

Scrape Data

1. From the file drop down menu, choose 'New File > R Script'



Scrape Data, cont.

2. I will show an example of what an R script to download twitter data looks like (on the board if necessary)

```
library(twitteR)

twitterKey <- c("the first key on the twitter api website", "the second key on the twitter
api website", "the third key on the twitter api website", "the fourth key on the twitter api
website") ## Holds all of the information you need to 'log in' to the twitter API and start
downloading data

setup_twitter_oauth(consumer_key = twitterKey[1], consumer_secret = twitterKey[2], access_to
ken = twitterKey[3], access_secret = twitterKey[4]) ## Logs in for you based on the four
keys you need

dogs <- searchTwitter("#Dogs", n = 100, lang = 'en') ## Searches, takes in several argument|

## searchString = 'what you want to look for, must be in quotes/apostrophes'
## n = number of tweets you want to find, will return either n tweets or as many have been
posted in the last 6 to 9 days (whichever is lower)
## lang = defaults to any language, 'en' for english, the codes are called ISO 639-1 codes
(you can look that up to find the codes in other languages)
## since = earliest date you want tweets from, must be 'YYYY-MM-DD' including quotes
## until = latest data you want tweets from, " " " "
## geocode = returns all tweets within a certain radius of a given latitude and longitude
## For more information, type ?searchTwitter into the console

dfdogs <- twListToDF(dogs) ## Converts the previous information into a more useful format

# head(dfdogs) ## Prints the first few rows of dfdogs so you can look at them
# write.csv(dfdogs, "dfdogs.csv") ## Outputs the data into a csv file
```

