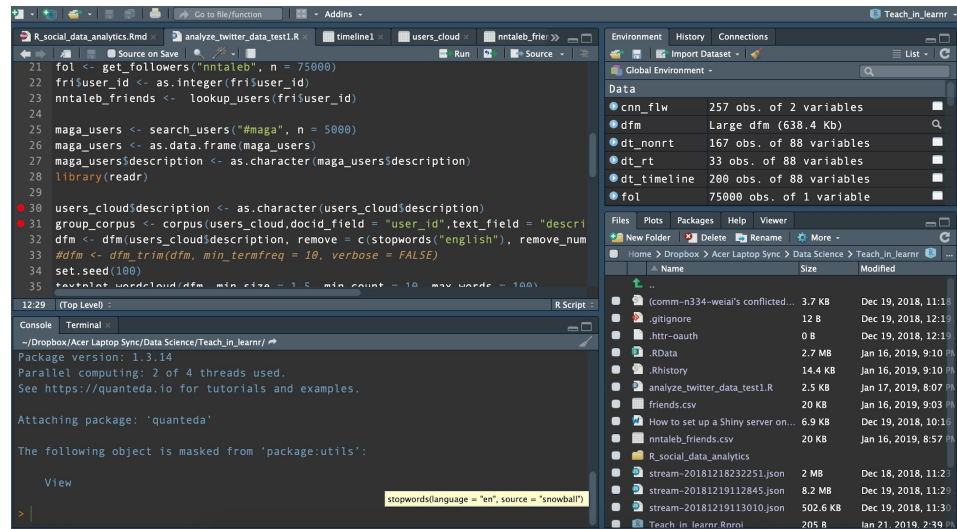


# Set up R and RStudio, install libraries and explore R data frames

Weiai Xu (Wayne), PhD  
Assistant Professor  
Department of Communication, UMass-Amherst  
Email: [weiaixu@umass.edu](mailto:weiaixu@umass.edu)  
[curiositybits.cc](http://curiositybits.cc)

# Overview

1. Install R and RStudio
2. R interface
3. Create a R *project*
4. Introducing R *libraries*
5. What is *working directory*?



# Install R in your machine



Download and install R:

<https://www.r-project.org/>

Download and install RStudio:

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



# Install R and RStudio

1. Install R language:  
<https://cloud.r-project.org/>



[CRAN](#)  
[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

*About R*  
[R Homepage](#)  
[The R Journal](#)

*Software*  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

*Documentation*  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

## 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

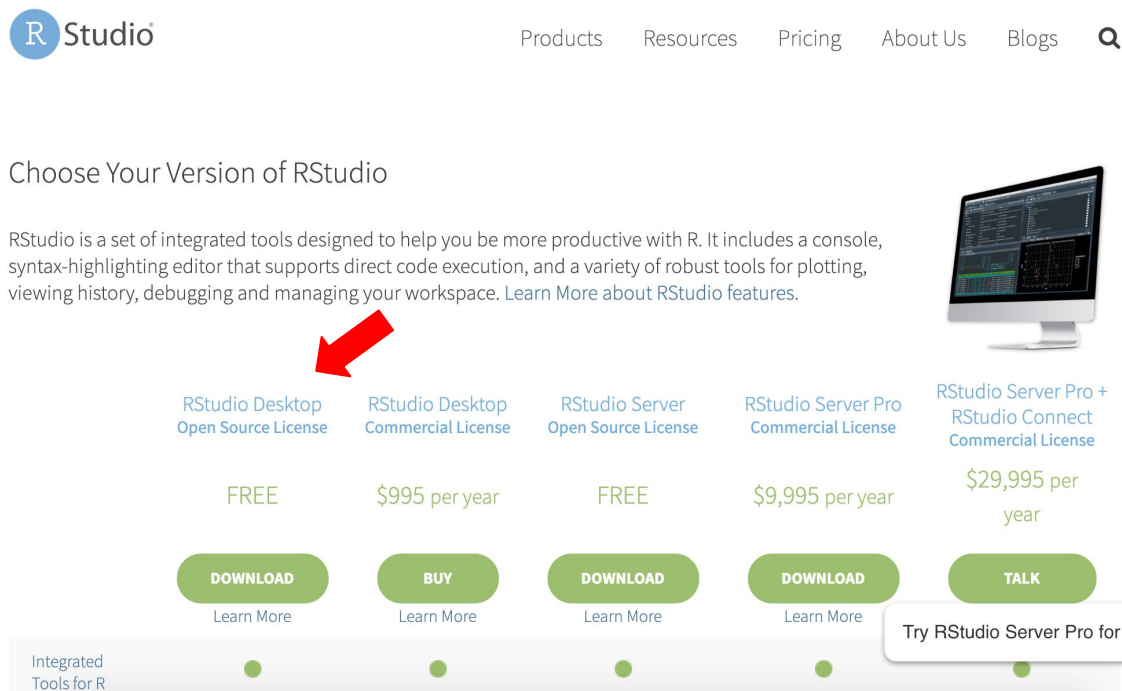
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- 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](#).

# Install R and RStudio

## 1. Install RStudio:

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



The screenshot shows the RStudio website's product selection page. At the top is the RStudio logo and a navigation menu with links for Products, Resources, Pricing, About Us, and Blogs, along with a search icon. Below the navigation is the heading 'Choose Your Version of RStudio'. A descriptive paragraph states that RStudio is a set of integrated tools for R, including a console, editor, and plotting tools, with a link to 'Learn More about RStudio features.' To the right of the text is an image of a computer monitor displaying the RStudio interface. Below the text and image is a table of product options. A red arrow points to the 'RStudio Desktop Open Source License' option. The table lists five options: RStudio Desktop Open Source License (FREE), RStudio Desktop Commercial License (\$995 per year), RStudio Server Open Source License (FREE), RStudio Server Pro Commercial License (\$9,995 per year), and RStudio Server Pro + RStudio Connect Commercial License (\$29,995 per year). Each option has a corresponding button: 'DOWNLOAD' for the first four and 'TALK' for the last one. Below the buttons are links to 'Learn More' for each product. At the bottom left, there is a section titled 'Integrated Tools for R' with five green dots. A callout box at the bottom right says 'Try RStudio Server Pro for free!'.

RStudio

Products Resources Pricing About Us Blogs

### Choose Your Version of RStudio

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. [Learn More about RStudio features.](#)

RStudio Desktop Open Source License	RStudio Desktop Commercial License	RStudio Server Open Source License	RStudio Server Pro Commercial License	RStudio Server Pro + RStudio Connect Commercial License
FREE	\$995 per year	FREE	\$9,995 per year	\$29,995 per year
<a href="#">DOWNLOAD</a>	<a href="#">BUY</a>	<a href="#">DOWNLOAD</a>	<a href="#">DOWNLOAD</a>	<a href="#">TALK</a>
<a href="#">Learn More</a>	<a href="#">Learn More</a>	<a href="#">Learn More</a>	<a href="#">Learn More</a>	

Integrated Tools for R

Try RStudio Server Pro for free!

# R Interface

The image shows the RStudio IDE interface with several annotations highlighting key components:

- Source windows:** A yellow box highlights the script editor area where R code is written and edited. The code visible includes functions for fetching followers, searching users, and processing text data.
- Environment:** A yellow box highlights the Environment pane on the right, which lists the objects in the current R session. The objects listed are: `cnn_flw`, `dfm` (Large dfm (638.4 Kb)), `dt_nonrt` (167 obs. of 88 variables), `dt_rt` (33 obs. of 88 variables), `dt_timeline` (200 obs. of 88 variables), and `fol` (75000 obs. of 1 variable).
- Console Tab:** A yellow box highlights the Console pane at the bottom, which shows the output of the R code. The output includes package version information, parallel computing status, and the installation of the 'quanteda' package.

The R code in the source window is as follows:

```
21 fol <- get_followers("nntaleb", n = 75000)
22 fri$user_id <- as.integer(fri$user_id)
23 nntaleb_friends <- lookup_users(fri$user_id)
24
25 maga_users <- search_users("#maga", n = 100)
26 maga_users <- as.data.frame(maga_users)
27 maga_users$description <- as.character(maga_users$description)
28 library(readr)
29
30 users_cloud$description <- as.character(users_cloud$description)
31 group_corpus <- corpus(users_cloud, docid_field = "user_id", text_field = "description")
32 dfm <- dfm(users_cloud$description, remove = c(stopwords("english"), remove_numbers))
33 #dfm <- dfm_trim(dfm, min_termfreq = 10, verbose = FALSE)
34 set.seed(100)
35 textstat_wordcloud(dfm, min_size = 1.5, min_count = 10, max_words = 100)
```

The console output shows:

```
Package version: 1.3.14
Parallel computing: 2 of 4 threads used.
See https://quanteda.io for tutorials and examples.
Attaching package: 'quanteda'
The following object is masked from 'package:utils':
  View
stopwords(language = "en", source = "snowball")
```


# Create a R project

*File--New Project...*

New Project

**Back**

**Create New Project**



Directory name:  **Name of your R project**

Create project as subdirectory of:  **Browse...**

☐ Create a git repository

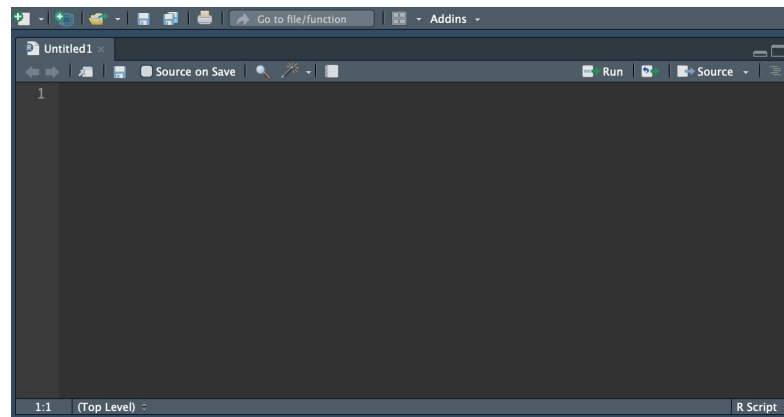
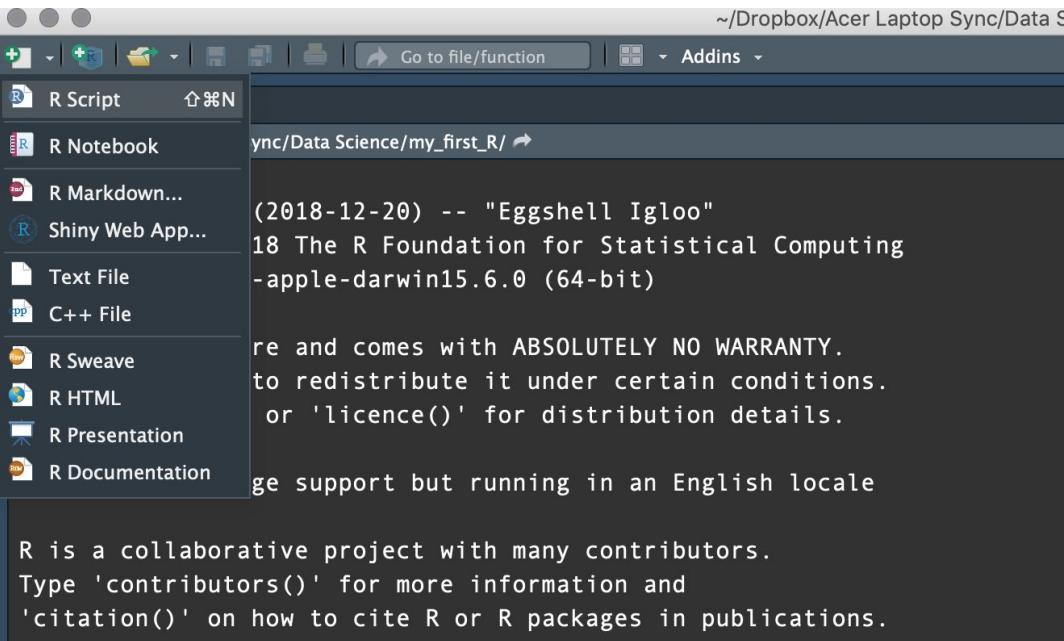
☐ Use packrat with this project **The working directory of this R project to be created**

☐ Open in new session

**Create Project** **Cancel**

# Create a R project

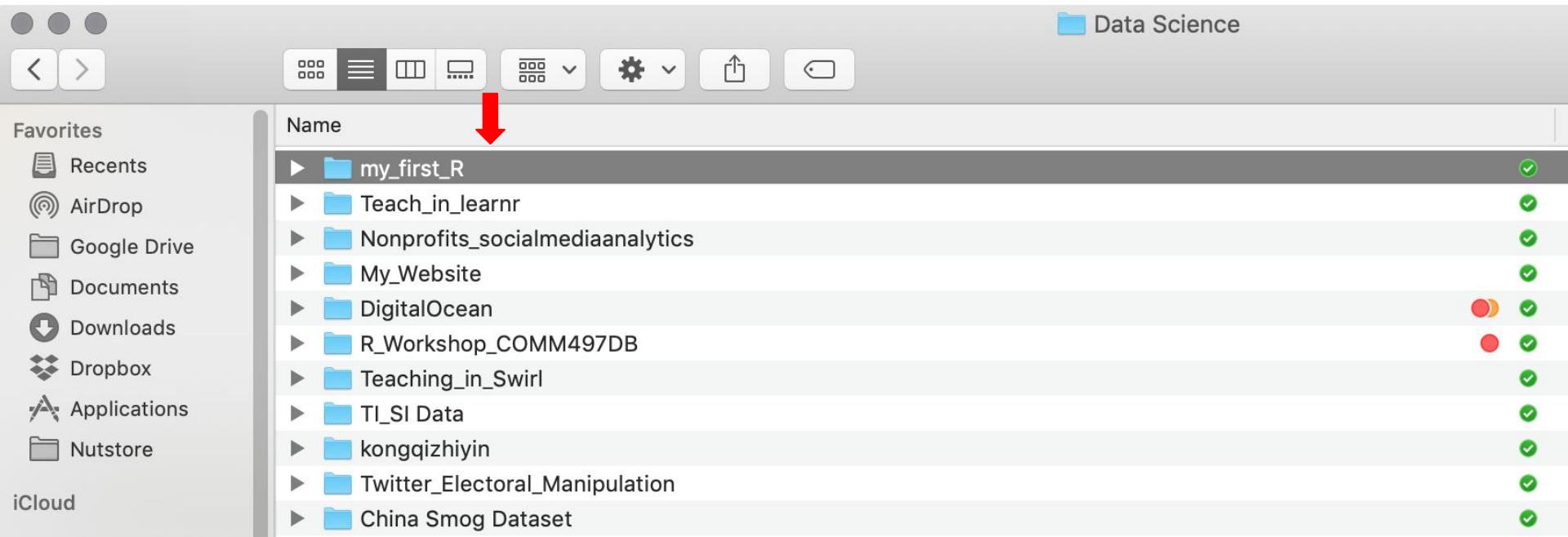
Add a R script for the newly created project





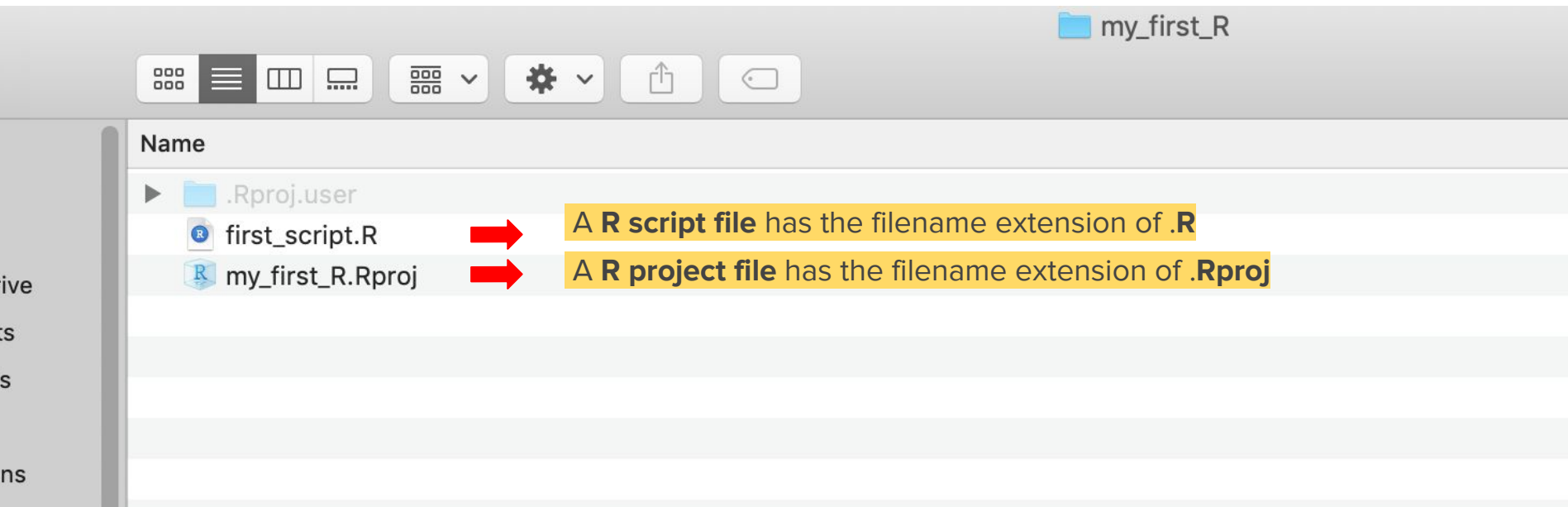
# Create a R project

Explore the working directory for this R project



# Create a R project

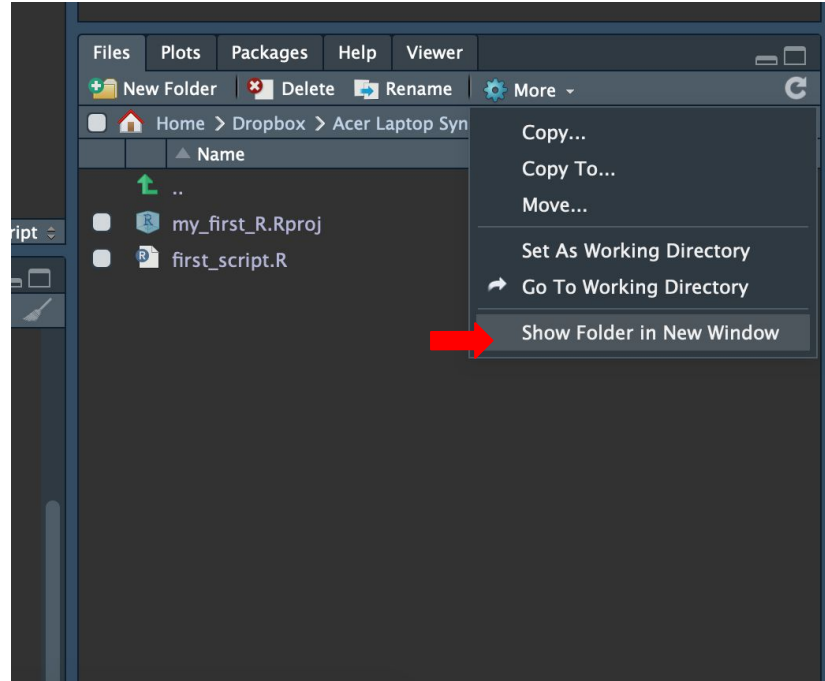
Explore the working directory for this R project



# Create a R project

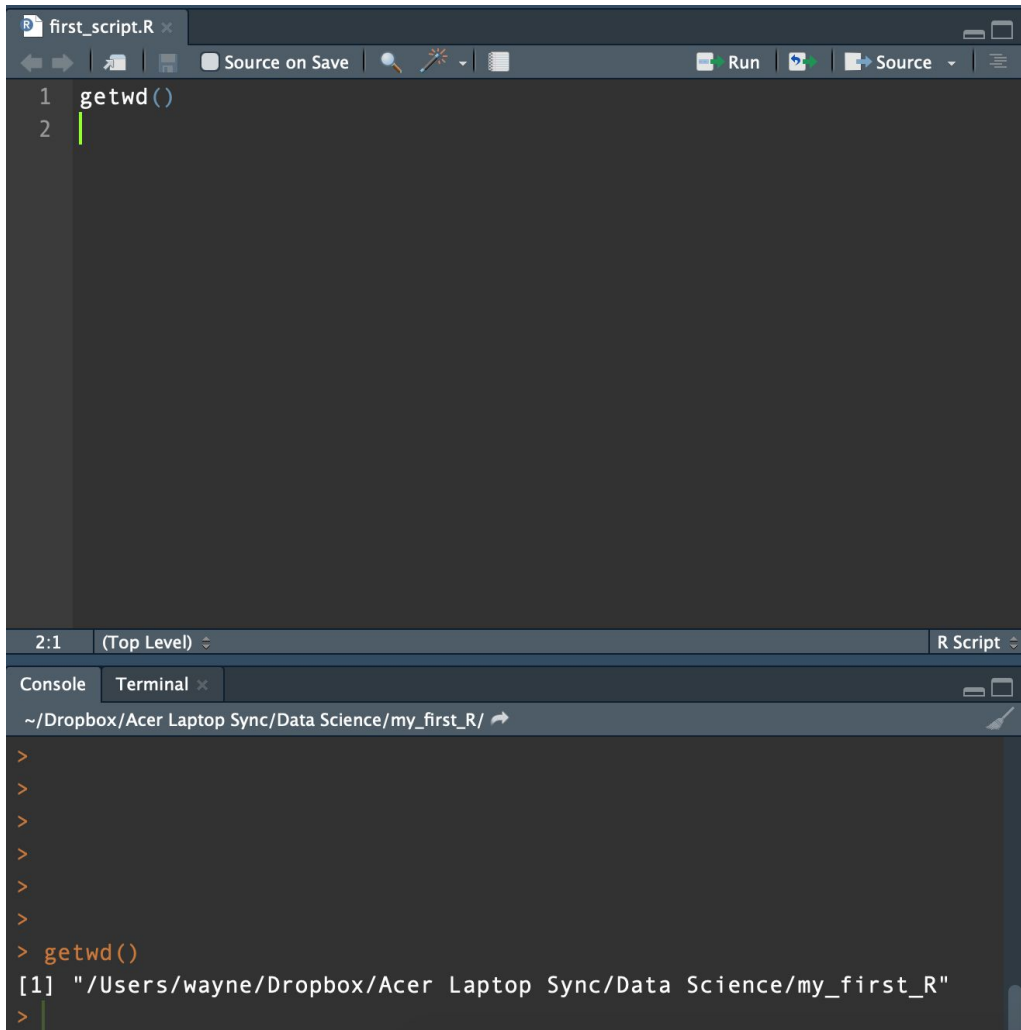
Open the directory folder from RStudio

Each R project has a designated folder to store R scripts and data (the folder is referred to as the working directory)



# Create a R project

Find out your current working directory

The image shows a screenshot of the RStudio application. The top pane is the 'Script Editor' with a file named 'first\_script.R'. It contains two lines of R code: '1 getwd()' and '2'. The bottom pane is the 'Console', which shows the output of the 'getwd()' command: '[1] "/Users/wayne/Dropbox/Acer Laptop Sync/Data Science/my\_first\_R/"'. The console prompt is '>'. The status bar at the bottom indicates the current working directory is '~ /Dropbox/Acer Laptop Sync/Data Science/my\_first\_R/'.

```
first_script.R
1 getwd()
2

2:1 (Top Level) R Script
Console Terminal x
~/Dropbox/Acer Laptop Sync/Data Science/my_first_R/
>
>
>
>
>
>
> getwd()
[1] "/Users/wayne/Dropbox/Acer Laptop Sync/Data Science/my_first_R/"
>
```

# Create a R project

Let's try some simple R codes

Use [**COMMAND**]+[**ENTER**] key on Mac, and [**CONTROL**]+[**ENTER**] key on PC to run a R script one line at a time.

To run multiple lines, just select them and run.

```
getwd()
```

```
a <- "this is my first line of R code"
```

```
typeof(a)
```

## Install a R library

What is a library/package? Think of R as an operating system (e.g., iOS, Windows) and a library/package as an app running on the system. Each library is designed to accomplish specific tasks.

Use **installed.packages()** to install libraries. Use **library()**, or **require()** to load an installed library.

Please follow the tutorial at:

[https://curiositybits.shinyapps.io/R\\_social\\_data\\_analytics/#section-librariespackages](https://curiositybits.shinyapps.io/R_social_data_analytics/#section-librariespackages)

# Data frame

A R data frame is a two-dimensional table with rows and columns. Put simply, a R data frame is a spreadsheet.

Please follow the tutorial at:

[https://curiositybits.shinyapps.io/R\\_social\\_data\\_analytics/#section-data-frames](https://curiositybits.shinyapps.io/R_social_data_analytics/#section-data-frames)

And use the practice script available on Moodle  
(W1\_install\_libraries\_and\_explore\_data\_frames.R)