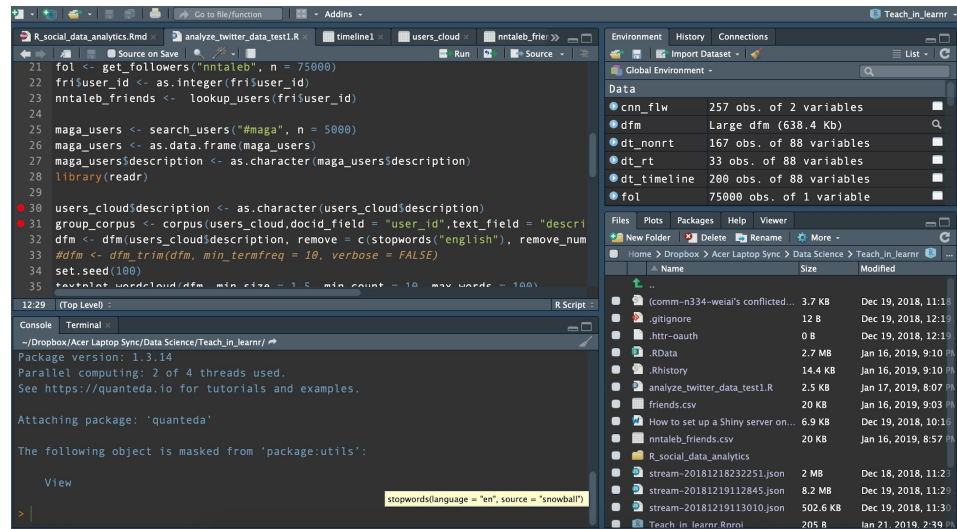


Set up R and RStudio, install libraries and explore R data frames

Weiai Xu (Wayne), PhD
Assistant Professor
Department of Communication, UMass-Amherst
Email: weiaixu@umass.edu
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

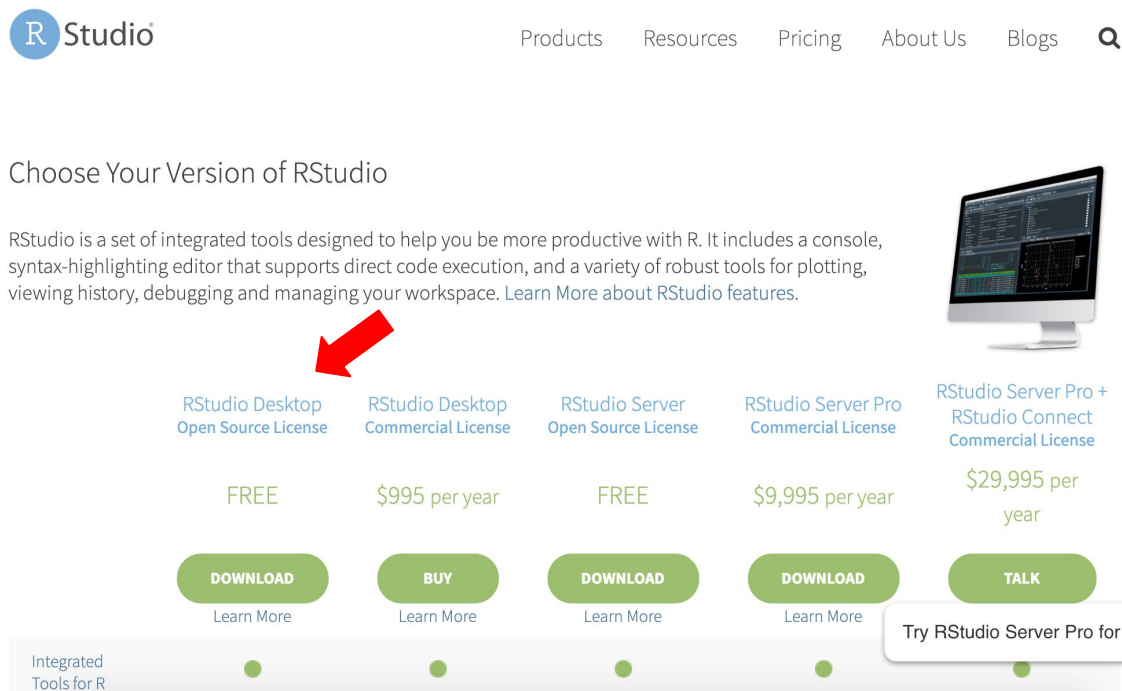
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- 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 five product options. A red arrow points to the 'RStudio Desktop Open Source License' option. Each option includes a license type, a price, a button to download or buy, and a 'Learn More' link. At the bottom, a row of green dots indicates which features are included in each version. A callout box for the 'RStudio Server Pro + RStudio Connect Commercial License' option says 'Try RStudio Server Pro for free!'.

Product	License	Price	Action	Learn More
RStudio Desktop	Open Source License	FREE	DOWNLOAD	Learn More
RStudio Desktop	Commercial License	\$995 per year	BUY	Learn More
RStudio Server	Open Source License	FREE	DOWNLOAD	Learn More
RStudio Server Pro	Commercial License	\$9,995 per year	DOWNLOAD	Learn More
RStudio Server Pro + RStudio Connect	Commercial License	\$29,995 per year	TALK	Learn More

Try RStudio Server Pro for free!

R Interface

The image shows the RStudio interface with several windows and tabs. The main window displays R code for analyzing Twitter data. The Environment pane on the right lists the objects in the workspace. The Files pane at the bottom right shows the project's file structure. The Console pane at the bottom left shows the output of the code.

Source windows: view/edit your R scripts

Environment: a list of *things* in your R project

Console Tab: show progress in data processing and output

```
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_num = TRUE))
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)
```

12:29 (Top Level) R Script

Console Terminal x

~/Dropbox/Acer Laptop Sync/Data Science/Teach_in_learnr/

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")

Environment History Connections

Global Environment

Data

- 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
- fol 75000 obs. of 1 variable

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home > Dropbox > Acer Laptop Sync > Data Science > Teach_in_learnr

Name	Size	Modified
..		
(comm-n334-weiai's conflicted...	3.7 KB	Dec 19, 2018, 11:18
.gitignore	12 B	Dec 19, 2018, 12:19
.httr-oauth	0 B	Dec 19, 2018, 12:19
.RData	2.7 MB	Jan 16, 2019, 9:10 PM
.Rhistory	14.4 KB	Jan 16, 2019, 9:10 PM
analyze_twitter_data_test1.R	2.5 KB	Jan 17, 2019, 8:07 PM
friends.csv	20 KB	Jan 16, 2019, 9:03 PM
How to set up a Shiny server on...	6.9 KB	Dec 19, 2018, 10:16
nntaleb_friends.csv	20 KB	Jan 16, 2019, 8:57 PM
R_social_data_analytics		
stream-20181218232251.json	2 MB	Dec 18, 2018, 11:23
stream-20181219112845.json	8.2 MB	Dec 19, 2018, 11:29
stream-20181219113010.json	502.6 KB	Dec 19, 2018, 11:30
Teach_in_learnr_Data		


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

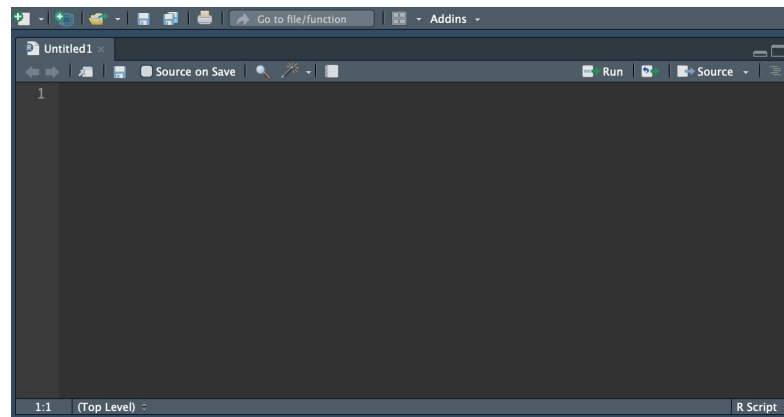
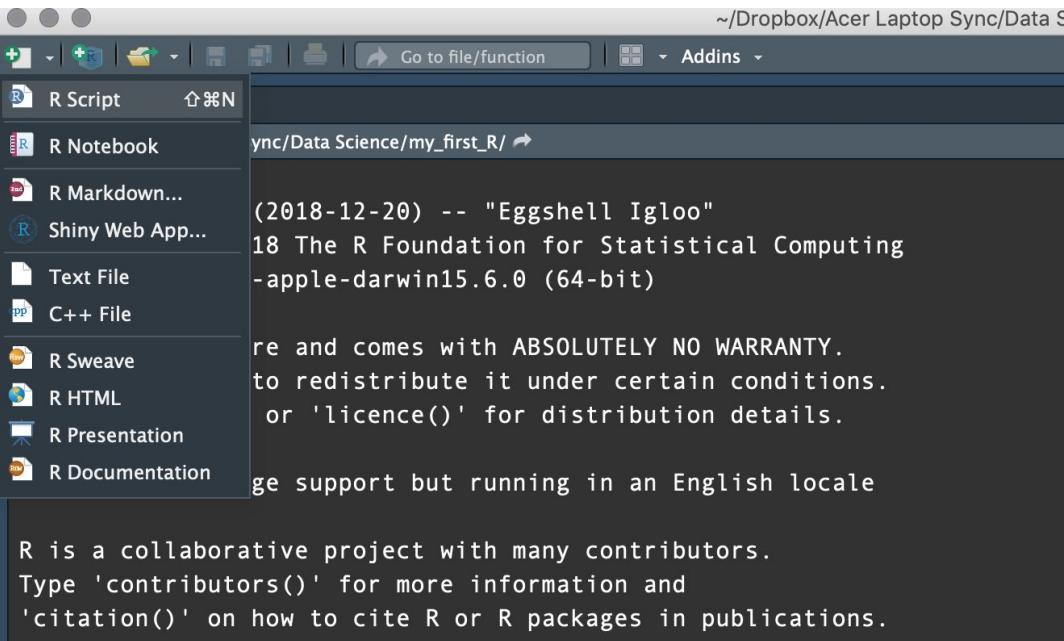
☐ Use packrat with this project **The working directory of this R project to be created**

☐ Open in new session

Create Project **Cancel**

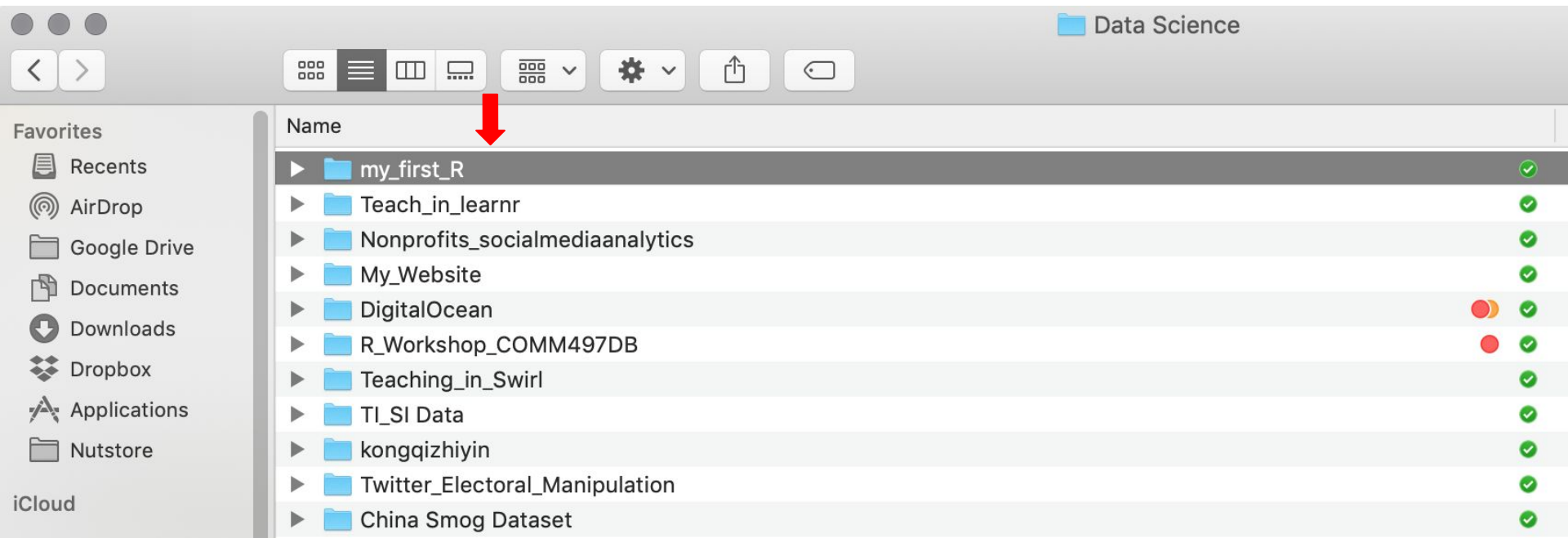
Create a R project

Add a R script for the newly created project



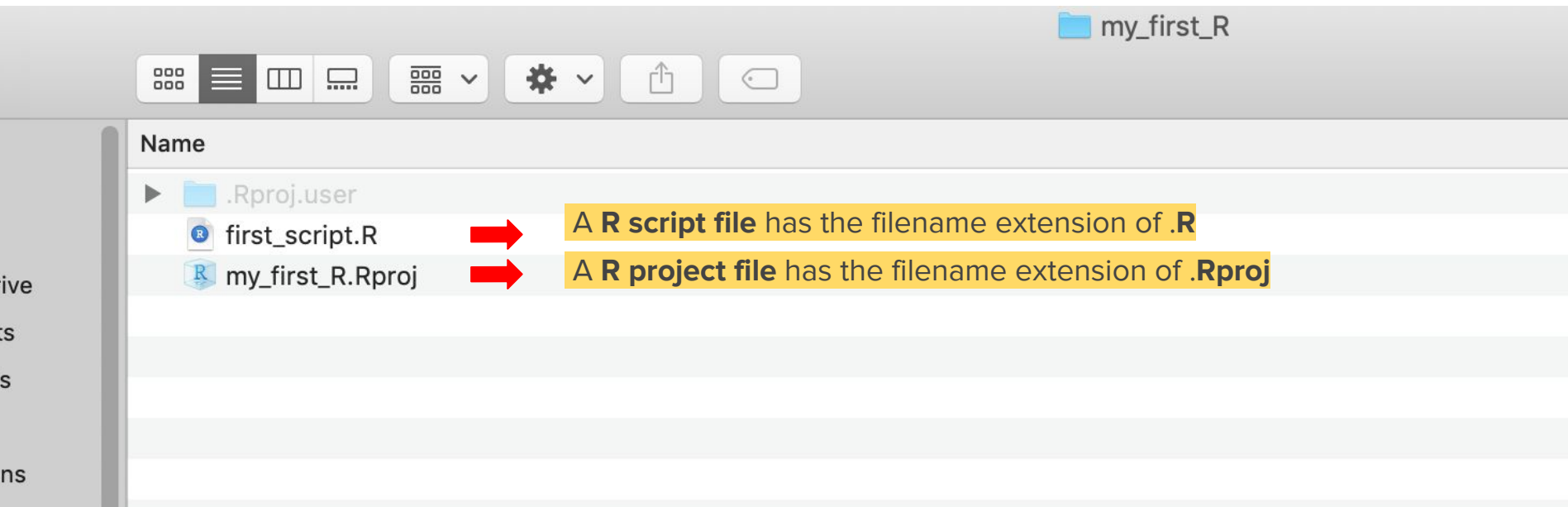
Create a R project

Explore the working directory for this R project



Create a R project

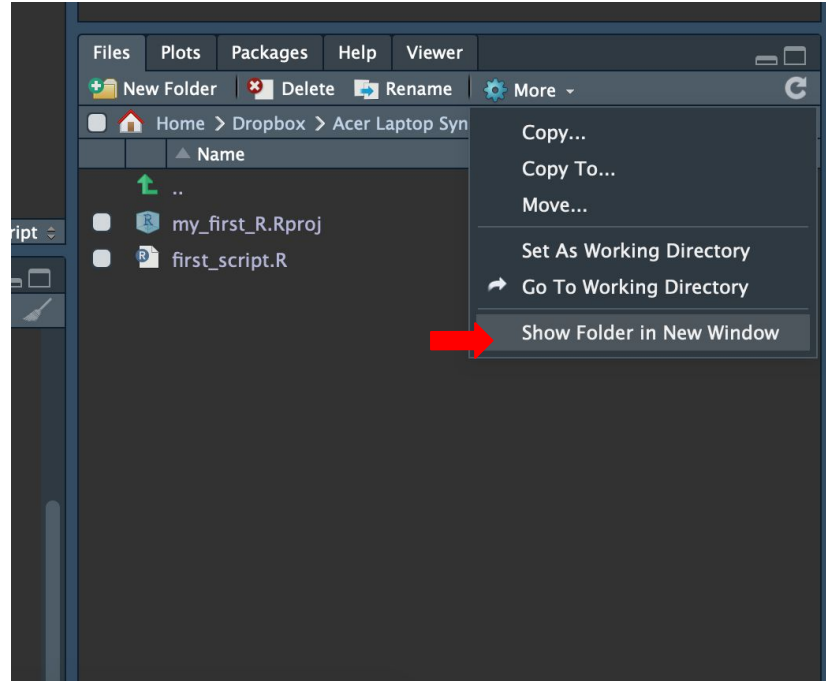
Explore the working directory for this R project



Create a R project

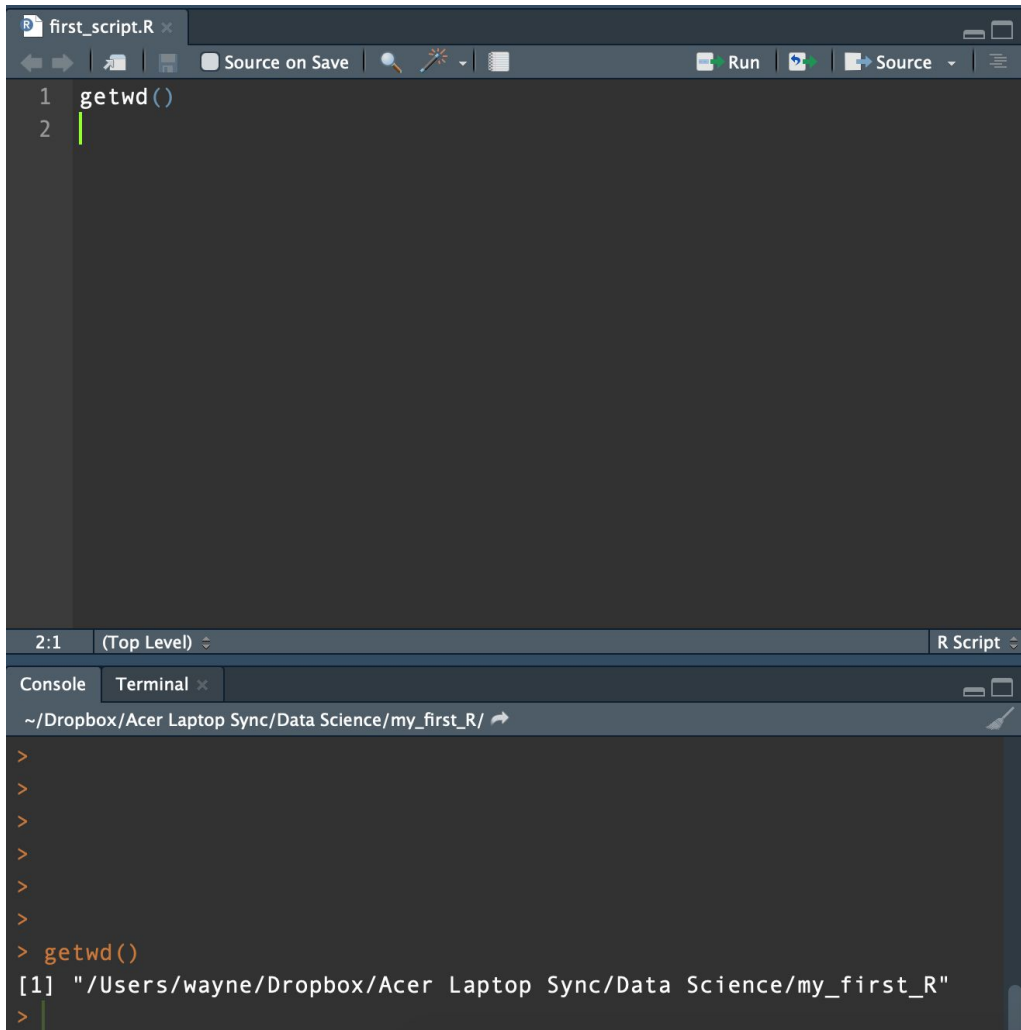
Open the directory folder from RStudio

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 screenshot shows the RStudio environment. The top pane is the script editor, displaying a file named 'first_script.R' with two lines of code: `1 getwd()` and `2` (with a cursor). The bottom pane is the console, showing the current working directory as `~/Dropbox/Acer Laptop Sync/Data Science/my_first_R/`. The console output shows the result of the `getwd()` command: `[1] "/Users/wayne/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 **install.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

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

And use the practice script available on Moodle
(W1_install_libraries_and_explore_data_frames.R)