

1201 Data Science Laptop Setup

Tim Raiswell

10/11/2018

1. Install R

R is a programming language that runs on Windows, Mac and Linux. It was designed for the purpose of statistical computing. As a result, it already has many data science and statistics features baked into it.

1. Go to the R website here: **R Link**
2. Select the right version of R for your machine.
3. Select the 'base' version.
4. Download and install.

You will get a message that because you do not have administrator rights you cannot associate the .Rdata file type. We can live with that.

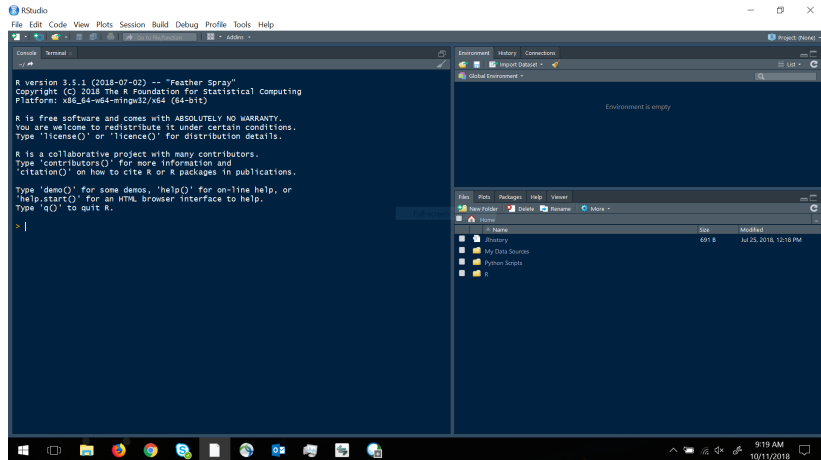
2. Install R Studio

R Studio is an open source R integrated development environment (IDE). It is designed to make data science and statistical experimentation easier and reproducible.

1. Go to the Rstudio download page here: **R Studio Link**
2. Download the installer for your machine.
3. Install the software.

3. Open R Studio

You should be presented with an interface that looks something like this:



Tweak the R Studio Interface

You will be using the interface a lot so you should change it to a format you enjoy. Navigate to:

1. Tools
2. Global Options
3. Appearance

Play around with the settings until you find a version you can stare at for hours on end.

Install Your First Package

You should already be in the R console (big panel on the left).
Type the code I have included below and hit 'Enter'.

```
install.packages("pacman") # this command tells R  
# to install the package named inside quotation marks.  
# Don't forget the quotation marks!  
# R lets you use single or double quotes.
```

Code Markup Convention

Writing notes to yourself or others that follow so that your code is readable and relatable is important.

```
# Note that when I type text after a hashtag,  
# it tells R that I'm just adding notes, not writing code
```

```
1+6 # this is code. It has no hashtag before it.
```

```
## [1] 7
```

```
# So, R calculates the output above.  
# 1+6 - this is not code because of the hash,  
# so no output is given.
```

Install Several Packages at Once

Now we can install and load a few packages we'll use for data science related work. Type this code in the console (minus my notes to you). The double colon tells R to look inside the pacman package for the 'p_load' function. It's a good way to find functions in the packages we load into R.

```
pacman::p_load(tidyverse, psych, readr) # we're using  
# the pacman packages  
# we just installed to install and  
# load three new packages.
```


Setting up an R Project

R's project and file system takes a little getting used to. We'll eventually cover an entire data science workflow so that you can maintain projects for reproducibility, but for now we'll keep it simple.

1. In the top right of the interface, select the dropdown box next to **Project (none)**.
2. Select 'New Project'.
3. Click on 'New Directory' and then 'New Project' at the top of the menu.
4. Now name the project directory '1201' in directory name field.
5. And navigate to your Documents folder using the Browse button.

Create the project!

Downloading the Data We Need for Monday

Now we'll load the data we'll use in our first meeting on **Monday, October 15**. The data is related to Facebook posts published during the year of 2014 on the page of a renowned cosmetics brand. This dataset contains 500 of the 790 rows and part of the features analyzed by Moro et al. (2016). The remaining instances were omitted due to confidentiality issues.

```
library(readr) # load the readr package
facebook_data_raw <- read_csv2('https://bit.ly/2NB7vKp')

# this function tells R to grab the data from a cloud
# folder and drop it into your laptop's project folder.
```

Check that the Data Made it Over

If not, don't worry. Send me a note on Slack or we'll solve the problem on Monday.

```
dim(facebook_data_raw)  # we'll use the 'dim' function.
```

```
## [1] 500  19
```

```
# It gives us the dimensions of the data frame object:  
# 500 instances of 19 variables!
```

Understanding the Shape of the Data

If you have time, try these lines of code one at a time in the console and review the output they produce.

```
head(facebook_data_raw)
head(facebook_data_raw, 10)
tail(facebook_data_raw)
names(facebook_data_raw)
str(facebook_data_raw)
glimpse(facebook_data_raw) # from the 'tibble' package
describe(facebook_data_raw) # from the 'psych' package
View(facebook_data_raw)
```

Closing Down

When you've finished, close R. It will ask you if you want to save the workspace image. Select 'Save'.

Other Useful Information

1. **Link to the paper that the Facebook data comes from.**
2. **Link to the Tidyverse homepage.** The Tidyverse is an ecosystem of R packages for data engineering, analysis and machine learning.
3. Link to the Psych package **manual and vignettes**. Many well-maintained R and Python packages have their own online vignettes, where the major tools of the package are showcased.