

Tidyverse Mini-Bootcamp

Pritam Dalal

May 4, 2019

Thank You!

Let me start with a **big thanks** to the sponsors, organizers, and volunteers that have made this event possible:

- ▶ Social Data Science
 - ▶ John Houge
 - ▶ Sona Maniyan
 - ▶ Abishek Roy
- ▶ The University of St Thomas
- ▶ The volunteers that will be helping through out the sessions

A Bit About Pritam

- ▶ I've worked in quantitative finance for 10 years - trading and analysis roles. I gravitated towards data-centric work.
 - ▶ Cargill
 - ▶ Wolverine Trading
 - ▶ Allianz
- ▶ I currently do independent consulting in quantitative finance and data science.
- ▶ I am an instructor at the University of Minnesota
 - ▶ Master of Financial Mathematics Program
 - ▶ R and Python
- ▶ Feel free to stay in touch: dalal030@umn.edu

Delta-Neutral

www.historicaloptiondata.com



- ▶ All the option data from this course was donated by Delta-Neutral.
- ▶ Professional quality EOD equity option data.
- ▶ Over 4000 underlyings covered.
- ▶ <http://www.deltaneutral.com/>
- ▶ support@deltaneutral.com

A Bit About the Mini-Bootcamp

- ▶ An introduction to basic data wrangling in R.
- ▶ I am assuming no previous background in R or programming.
- ▶ We will be using a suite of packages called the **tidyverse** which has become the de facto standard in for data analysis in R.
- ▶ What is wrangling? Why wrangling?
 - ▶ manipulation of data: filtering, subsetting, performing aggregation calculations on subsets of data
 - ▶ the best place to start

Schedule and Format

- ▶ Schedule
 - ▶ 9:00a - 10:45a - tutorials
 - ▶ 10:45a - 11:00a - break
 - ▶ 11:00a - 12:30p - tutorials
- ▶ Our time will be spent working through tutorials and a capstone project.
 - ▶ I will live-code all the examples
 - ▶ I **highly** encourage you to type along with me
- ▶ Lots of intermittent coding challenges.
- ▶ Please ask questions
 - ▶ interrupting is encouraged
 - ▶ be brave

Course Materials

Let's take a moment and download the course materials.

1. https://pritamdalal.github.io/rff_course/
2. Go to **Course Content** page
3. Under *Basic Wrangling* click the blue link: Download Course Material
4. Move that folder to your Desktop

Contents of bw_material

The folder you downloaded contains several subfolders:

- ▶ bw_capstone - the files for the capstone project (we will use this later)
- ▶ bw_slides - slide presentations
- ▶ bw_tutorial - contains all the tutorials we will work through

Launching R Studio (1 of 2)

To launch R Studio double click on the following file:

- ▶ bw_materials
 - ▶ bw_tutorial
 - ▶ **bw_tutorial.Rproj**

Double clicking on the .Rproj file will automatically launch R Studio.

Launching R Studio (2 of 2)

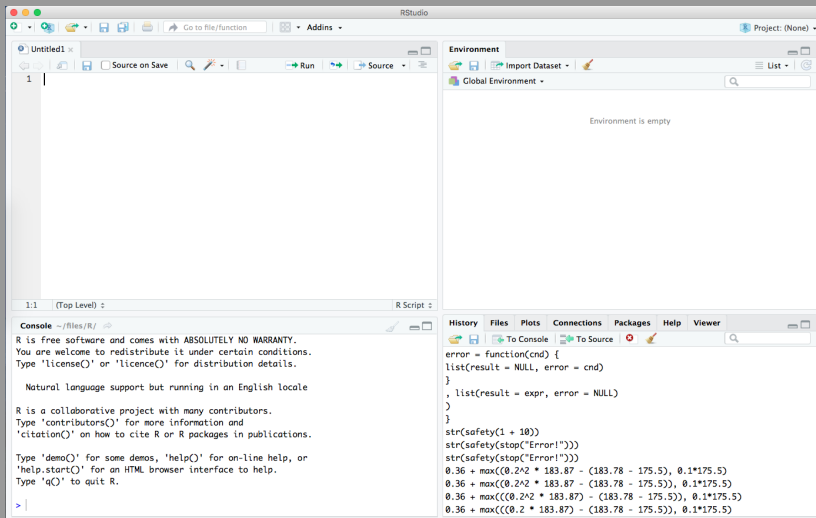


Figure 1:

Setting the Dark Theme (Optional) - (1 of 2)

I find that students often want the same dark theme as me. . .

1. In the menu bar, navigate as follows:

- ▶ Tools > Global Options

2. In the resulting dialogue box:

- ▶ click *Appearance* on left-hand side

3. Set *Editor theme: Cobalt*

Setting the Dark Theme (Optional) - (2 of 2)

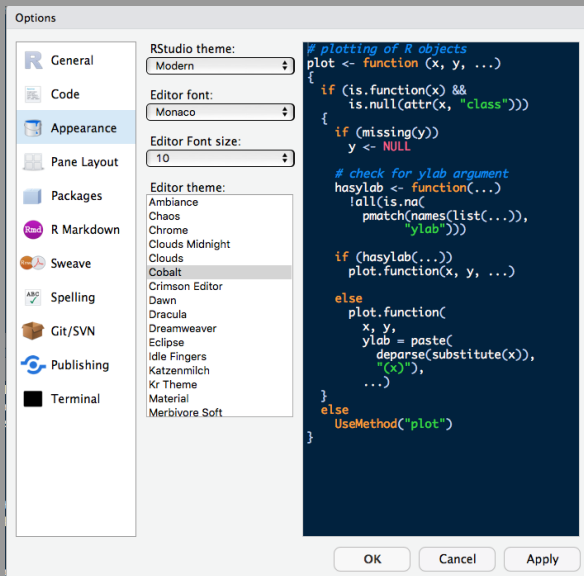


Figure 2:

Setting Up Tutorial 01 (1 of 3)

1. In Rstudio, you will have a file open in source window called `follow_tutorial_01.Rmd`.
 - ▶ this is an Rmarkdown file
 - ▶ later we will use scripts
2. press **control + shift + 1**
 - ▶ this will change R Studio so you **only** see the source window
3. Reshape the R Studio window to only take up half of your screen.

Setting Up Tutorial 01 (2 of 3)

4. In your file system, double click on:
 - ▶ rff_materials
 - ▶ bw_tutorial
 - ▶ **01_intrduction.html**
5. The first tutorial will launch in your web browser.
6. Get your screen to look like mine:
 - ▶ left-half R Studio
 - ▶ right-half tutorial

Setting Up Tutorial 01 (3 of 3)

7. Finally, turn on: Chunk Output Inline

The image shows two side-by-side windows. The left window is RStudio, and the right window is a web browser.

RStudio Window:

- File: `follow_01.Rmd`
- Menu: `Knit` > `Insert` > `Chunk Output Inline` (checked)
- Code editor shows R code with chunk markers: `1 {r setup, include=FALSE}`, `2 knitr::opts_chunk$set(paged,`, `3 {r}`, `4 {r}`, `5 {r}`, `6 {r}`, `7 {r}`, `8 {r}`, `9 {r}`, `10 {r}`, `11 {r}`, `12 {r}`, `13 {r}`, `14 {r}`, `15 {r}`, `16 {r}`, `17 {r}`, `18 {r}`, `19 {r}`, `20 {r}`, `21 {r}`, `22 {r}`, `23 {r}`, `24 {r}`, `25 {r}`, `26 {r}`, `27 {r}`, `28 {r}`, `29 {r}`, `30 {r}`, `31 {r}`, `32 {r}`, `33 {r}`, `34 {r}`
- Status bar: `1:1 (Top Level)` | `R Markdown`

Web Browser Window:

- Address bar: `File | /Users/Pritam/files/ods/rff_course/rff_b...`
- Page title: **Tutorial 1**
- Section: *A Brief Introduction*
- Text:

The purpose of this tutorial is to have you type some R code, and to give a brief tour of R.

The code that we write will accomplish two finance related tasks:

 1. download historical SPY price data from Yahoo finance
 2. graph the time series of the prices

Don't worry if the code you are typing seems foreign. As long as you are running code and getting output, you are making progress.
- Section: **Let's Get Typing**
- Text:

Type the following and then press `ctrl + shift + enter`.
- Code block:

```
rnorm(5)
[1] 1.1418555 2.0916099 0.3110570 -0.3971084 -0.4768715
```
- Text:

What did we do?

 - we typed code that called the function `rnorm()` with the input `5`.
 - we ran the code by pressing `ctrl + shift + enter`.
- Text:

What happened?

 - 5 normal random numbers were generated
 - these numbers were printed to the screen
- Code Challenge box:

Code Challenge: Generate a set of 10 random numbers and print them to the screen.

Figure 3:

Setting Up Tutorial 11 (1 of 2)

1. File > Open > 11_script.R
2. Press **ctrl + shift + 0**
 - ▶ your RStudio window now have four seaparte quandrants
3. Rearrange your window panes to look like mine.
 - ▶ Tools > Global Options...
 - ▶ Pane Layout
4. I recommend minimizng the Enviroment and the Help quadrants

Setting Up Tutorial 11 (2 of 2)

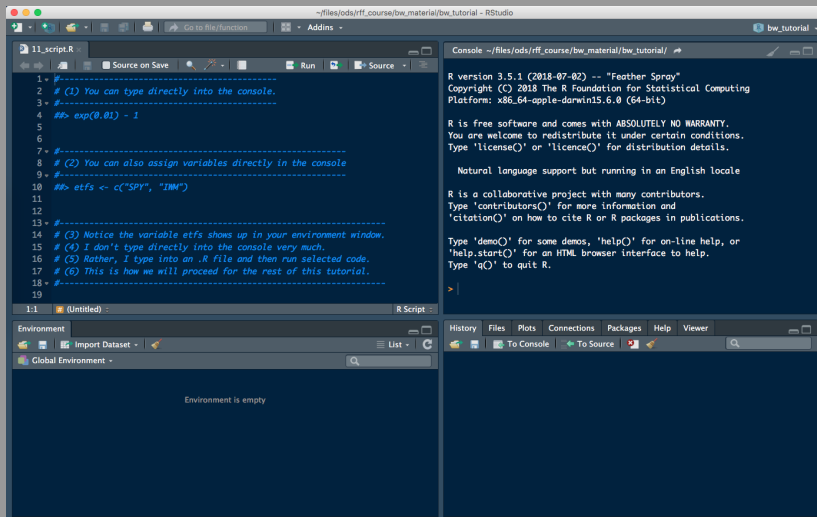


Figure 4:

Launching the Capstone Project

1. Quit out of RStudio
2. Launch a new session of RStudio
3. File > Open Project...
 - ▶ bw_materials
 - ▶ bw_capstone
 - ▶ **bw_capstone.Rproj**

Contents of **bw_capstone** Folder

I find that my analysis projects tend to have three components. These are reflected in the subfolders that I have created in the `bw_capstone` project folder.

- ▶ `01_wrangling` - .R scripts that perform wrangling on data.
- ▶ `02_data_output` - the wrangling produces intermediate data files.
- ▶ `03_analysis` - once I have all the data files I need, I summarize my findings in on or more .Rmd documents.

Instructions for Project

1. You have the core tidyverse knowlege for this project, but it is also intended to push your comfort zone and teach you a few new functions/techniques along the way.
2. We will work through the project together, step-by-step.
3. In your file system, double click the the following file and it will open in your browser:
 - ▶ bw_materials
 - ▶ bw_capstone
 - ▶ **bw_capstone_instructions.html**