

# R Resources

Rafael M. Batista\*

Last edited: **20 April 2021**

## Contents

<b>Intro</b>	<b>2</b>
<b>Dipping your toes</b>	<b>2</b>
<b>Programming &amp; Statistics</b>	<b>2</b>
<b>Data Visualization</b>	<b>2</b>
<b>Text Mining and Analysis</b>	<b>2</b>
<b>Machine Learning</b>	<b>3</b>
<b>File Management and Sharing</b>	<b>3</b>
<b>Miscellaneous</b>	<b>3</b>

---

\*University of Chicago Booth School of Business, Chicago, IL 60637. [rafael.batista@chicagobooth.edu](mailto:rafael.batista@chicagobooth.edu).

## Intro

A list of resources I have found helpful over the years. I use most of the items here regularly and have each of them bookmarked.

## Dipping your toes

When learning something new a bit of hand-holding can go a long way. For example, when I get to a new city with an unfamiliar bus or train system, I will often avoid using that mode of transportation until I've gone on it at least once with a local, just to see how it works - do you pay on the train or before your board? Is change provided? Do you tap in *and* out or just in or just out? Typically all I need is one guided trip.

If this resonates at all with you, here are two particularly good resources to show you the ropes:

- **Swirl Stats** (link) gets you started with R *within* R. The lessons are quick to get through and will help flatten the learning curve
- **Markdown Tutorial** (link) does something similar to teach you Markdown. This isn't R per se, but within 10 minutes you'll learn the basics of Markdown syntax which can then be used together with R to create clean, reproducible documents.

## Programming & Statistics

- ***Discovering Statistics Using R* by Andy Field, Jeremy Miles, and Zoe Field** (Amazon.com link) has been my go-to book since I started using R in 2014 and I continue to refer back to it to this day both for basic. One downside is that some of the packages are a bit outdated (note a newer version is set to come out soon, the beta version of the accompanying R tutorial is currently here)
- ***R for Data Science* by Garrett Golemund and Hadley Wickham** (link) is worth reading basically end-to-end for anyone using R for data analysis. This book goes through the basics of importing, cleaning, transforming, and visualizing data.
- **\*Causal Inference: The Mixtape** by Scott Cunningham\*\* (<https://mixtape.scunning.com/index.html>) this is a relatively new book which not only teaches causal inference techniques but also provides corresponding code for Stata, R, and Python

## Data Visualization

- ***R Graphics Cookbook, 2nd Edition* by Winston Chang** (<https://r-graphics.org/>) - This book has over 150 'recipes' for the most common graphs. Consider this a menu of options; for example, if you know you want a histogram or a box plot, just jump to the relevant chapter and you'll find a template for it.
- **ggplot2 reference** (<https://ggplot2.tidyverse.org/reference/>) - **ggplot2** is a package used for building graphs in R. It works by layering different components on top of each other. This website provides a reference list for most components and examples of how to use each. I usually use it when I have something in mind that I want to do, such as changing the scales on the y-axis.
- ***ggplot2: Elegant Graphics for Data Analysis* by Hadley Wickham** (<https://ggplot2-book.org/>) - this is a shorter book, that basically gets you started with **ggplot2**. If you're new to R or new to **ggplot2**, I'd skim through this.
- ***Data Visualization: A practical Introduction* by Kieran Healy** (<https://socviz.co/>) - this book I haven't used very often but I've seen others referencing it and provides a series of considerations

## Text Mining and Analysis

- ***Text mining with tidy data principles* by Julia Silge** (<https://juliasilge.shinyapps.io/learntidytext/#section-introduction>) offers a straightforward tutorial for using **tidytext**, a package in R designed

to make text mining easier and more effective

- ***Text Mining with R: A tidy Approach*** by Julia Silge and David Robinson (<https://www.tidytextmining.com/index.html>) goes a bit deeper than the tutorial but follows the same principles
- ***Supervised Machine Learning for Text Analysis in R*** by Emil Hvitfeldt and Julia Silge (<https://smltar.com/>) focuses on using text for prediction (ie. text + machine learning)

## Machine Learning

### File Management and Sharing

- ***Happy Git and Github for the userR*** by Jenny Bryan, the STAT 545 TAs, and Jim Hester (<https://happygitwithr.com/>) - I used this to get setup with Git and Github. Neither are very intuitive to use, in my opinion, but incredibly valuable for organizing and storing files. It takes a couple hours to get through and set things up but you will thank yourself for doing this sooner rather than later.

### Miscellaneous

- **RStudio Cheatsheets** (<https://rstudio.com/resources/cheatsheets/>) - There are literally dozens of cheatsheets here. I've downloaded many of them and open at least one of them up every time I use R. If you believe in the 80/20 rule, these cheatsheets will cover the 80% of things you need to know under each topic.
- **stargazer** (<https://www.jakeruss.com/cheatsheets/stargazer/>) - I've been using the **stargazer** package to generate tables in R, specifically regression tables. This is a fairly extensive resource for this package.