Resources that have been useful

originally compiled by Tyler Clark, Class of 2020

Become a better R programmer:

* [TidyTuesday](https://www.youtube.com/user/safe4democracy) – Weekly YouTube series by David Robinson. Incredibly proficient R programmer spends one hour analyzing a dataset he has never seen before. **Thoroughly explains every decision he makes while conducting the analysis along with the purpose of each function and argument he uses.** Very high level exploratory data analysis on quantitative and text datasets to create insightful visualizations, mainly using dplyr and ggplot2, that is incredibly easy to understand and learn from. His channel has very low view counts, often below 2000 views per video but don’t let that deter you. If you struggle with EDA or using the Tidyverse, I really can’t recommend this enough. Even if you are confident in your programming skills, I guarantee he will teach you something useful in any given video you click on and you can see the thought process of a very successful data scientist.

David Robinson – “I’m a data scientist at [Heap](https://heap.io/). My interests include statistics, data analysis, education, and programming in R.

I’m the co-author with Julia Silge of the tidytext package and the O’Reilly book Text Mining with R. I’m also the author of the [broom](http://github.com/dgrtwo/broom) and [fuzzyjoin](http://github.com/dgrtwo/fuzzyjoin) packages, and of the e-book [Introduction to Empirical Bayes](http://varianceexplained.org/r/empirical-bayes-book/).

I previously worked as Chief Data Scientist at [DataCamp](http://datacamp.com/) and as a data scientist at Stack Overflow, and received a PhD in [Quantitative and Computational Biology](http://www.princeton.edu/qcbgrad/) from Princeton University.”

Retrieved from <http://varianceexplained.org/about/>.

* [Advanced R Programming](https://adv-r.hadley.nz/index.html) – Free textbook that discusses how R works as a programming language. **Not at all like reading a traditional textbook, very minimalistic, understandable, applicable and gets the point across in a few sentences rather than pages of filler information.** Short sections and chapters with simple examples that make it very easy to understand the sections. Sections are often as short as a paragraph with an example and you can finish chapters in 15 to 20 minutes with ease. Great resource to casually read through when you have a few free minutes. You can get by without ever looking at this book but if you casually read it, you will have a better understanding of how R works as a programming language. You will **absolutely** get those 20 minutes back in the future when you are trying to fix a frustrating error message and have a better understanding of what is causing it, or you avoid the bug entirely because you programmed a tricky problem correctly the first time.

Make high quality graphs for presentations

* [R Graphics Cookbook](https://r-graphics.org/) – Free textbook for using ggplot2. Great book to keep saved in your favorites and reference when you are trying to figure something out.
* [The R Graph Gallery](https://www.r-graph-gallery.com/index.html) – If you can picture what you want your graph to look like in your mind but can’t code it, The R Graph Gallery shows you all the different combinations possible and then shows you what functions to use when you find the one you want.
* [ggplot2 Quick Reference: shape](http://sape.inf.usi.ch/quick-reference/ggplot2/shape) – Cheat sheet for when you need to manually decide what shape you want your dots to be on a graph.
* [ggplot2: Quick Reference: color (and fill)](http://sape.inf.usi.ch/quick-reference/ggplot2/colour) – Cheat sheet for when you want to add some color to your graphs. But also keep in mind you aren’t restricted to the pre-named colors listed, you can also specify your own desired colors using RGB or HTML (Hex) color codes. Perfect for when you are presenting on a company and want to match their colors exactly.

Find out what is going on in the Data Science community

* Podcasts – All hosted by data scientists holding high level positions in large tech companies.
  + Linear Digressions
  + Not So Standard Deviations
  + Data Skeptic

Those are my personal favorites, but there are many others <https://towardsdatascience.com/top-20-podcasts-for-data-science-83dc9e07448e>.

* [R-bloggers](https://www.r-bloggers.com/) – R news and tutorials contributed by hundreds of R bloggers. Interesting to scroll through occasionally. Very wide range in levels of complexity and understanding so there is something for everyone.

Visualizations of probability and statistical theorems discussed in classes

* [Seeing Theory](https://seeing-theory.brown.edu/) – Interactive web page that lets you easily adjust individual parameters to visualize and understand how distributions and probabilities change with the parameters along with how the parameters themselves interact.

Practice coding and developing new skills

* [Kaggle](https://www.kaggle.com/competitions) – Website owned by Google that hosts data science competitions where you can compete against other teams and win real money if you place high enough. The more projects you have completed and can discuss with potential employers, the better.