Statistics I:
Section 0- R Resources

Alec Tripp
alexander.j.tripp@vanderbilt.edu

August 30, 2022

As an open-source programming language, R has a huge community with great resources for resolving issues and diving deeper. This document is meant to provide you with a sampling of these resources that may 1) help you solve your problems or 2) direct you to other useful sources. I aggregate tutorials for different aspects of R, as well as individuals within the department, that you can consult for more advanced or lesser known empirical strategies.

General Resources

This section of resources is better for general reference.

- While slightly dated, this is a great tool for learning the basic usage of R: `https://bookdown.org/ndphillips/YaRrr/`. Note that most of this resource uses base R, which I do not recommend. This is a great starting point for basics, but I would then suggest you invest your time into investing into ggplot2 and tidyverse.

- This should be your go-to resource for all things tidyverse: `https://r4ds.had.co.nz/`. It is my favorite R reference book and does a great job of walking you through all that the tidyverse has to offer.

- If you want to become fluent in R–read, write functions and create new packages–this is the book for you: `https://adv-r.hadley.nz/index.html`. It is very advanced and very technical.

- This is a cheatsheet for the R package data.table, which is helpful for cleaning and subsetting data: `https://raw.githubusercontent.com/rstudio/cheatsheets/main/datatable.pdf`

- Rather than a book, this is a collection of short courses that teach basic and advanced R skills in an approachable way: `https://rpubs.com/timothyfraser`.

- You will use this book next year in Brenton's Stats II class: `https://bkenkel.com/pdaps/`. It mostly involves statistical theory, but the appendix of each chapter has examples in R.

- Lastly, here is a textbook that employs R to analyze econometrics: `https://www.econometrics-with-r.org/1.2-a-very-short-introduction-to-r-and-rstudio.html`. I have not really used it, but I have heard good things.

Plotting

- While a bit buggy, STHDA is a repository of tutorials for plotting and other aspects of data visualization: `http://www.sthda.com/english/wiki/ggplot2-essentials`.

- Need inspiration? Take a look at this page, which has examples of all the graphs you could imagine (and more): `https://r-graph-gallery.com/`.

- This is a page that describes the sjplot package in R: `https://strengejacke.github.io/sjPlot/reference/index.html`. It is useful for plotting the results of regression tables and other statistical tests.

- Have something against ggplot2 and base R graphics? Look no further! Plotly is a cool alternative to traditional graphics in R: `https://plotly.com/r/` (just...don't waste your time learning plotly unless you have some good reason for it).

Survey Analysis

- This resource explains survey data analysis in R: `https://stats.oarc.ucla.edu/r/seminars/survey-data-analysis-with-r/`. It also compares R and Stata coding, which is helpful for all you Stata users.

- R's survey package is a great resource for survey analysis, and this page explains what the package has to offer: `http://r-survey.r-forge.r-project.org/survey/`.

- This is a quick and easy way to conduct a statistical power analysis: `https://egap.org/resource/script-power-analysis-simulations-in-r/`. A power analysis simulates the sample size needed to find significant results given some baseline results.

- Here is a resource explaining how to conduct a principal components analysis in R: `https://aedin.github.io/PCAworkshop/articles/b_PCA.html`.

- Here is a resource explaining how to conduct a factor analysis in R: `https://www.geo.fu-berlin.de/en/v/soga/Geodata-analysis/factor-analysis/`

`A-simple-example-of-FA/index.html#:~:text=In%20the%20R%20software%20factor,specified%20by%20the%20argument%20factors%20..`

- Kristopher Preacher, a professor in Psychology at Vanderbilt, offers a great class on Multilevel Modeling. He previously had course resources online, so maybe you can find something on his website: `http://quantpsy.org/`.

- Finally, here are a few examples of how to use factor analysis and principal components analysis in R: `http://factominer.free.fr/` and `http://www.sthda.com/english/wiki/wiki.php?id_contents=7851#principal-component-analysis`. If you can't tell, I struggled finding acceptable explanations of factor analysis and principal components analysis in R, so I am sharing my findings here.

Networks and Data Science

- Unfortunately, I am not as familiar with network analysis. However, Jenn Larson and Cassy Dorff in our department are. Here are their websites: `https://www.jmlarson.com/home` and `http://www.cassydorff.com/`.

- Fortunately, our department has a good representation of data science. You can check out the links above for Jenn and Cassy's work in data science. Additionally, you can check out the following websites of data scientists Josh Clinton and James Bisbee within the department: `https://www.joshclinton.com/` and `https://www.jamesbisbee.com/`.

Machine Learning

- For an in-depth overview of machine learning methods, I highly recommend this book: `https://bradleyboehmke.github.io/HOML/`. It is a fun read and has replicable coding examples for you to follow along.

- This is an explanation of the caret package in R, which is essential for implementing machine learning methods in R: `https://topepo.github.io/caret/index.html`.

- Tidymodels is a standardization of empirical modeling in R using a variety of approaches: `https://www.tidymodels.org/start/`. If you delve into machine learning using the above resources, you will see how a standardized approach can greatly improve your workflow.

Text as Data

- Here is a highly recommended short course on how to use text as data as an empirical approach: `https://joeornstein.github.io/text-as-data/`.

- This is a cheatsheet that explains the basics of regular expressions ("regex"), which are necessary for searching within bodies of text: `https://raw.githubusercontent.com/rstudio/cheatsheets/main/regex.pdf`.

- With this website, you can test your regular expressions before you try to implement them in R functions: `https://regex101.com/`.

Causal Inference

- Now, you won't be needing this until Statistics III (Causal Inference), but this was my favorite resource from the class: `https://mixtape.scunning.com/`. It has good explanations of complex models, and it provides examples of its models using R, Stata, and SPSS.

- Here is a course in causal inference taught by one of the greats, Kosuke Imai: `https://imai.fas.harvard.edu/teaching/cause.html`.