Data Analysis Notes

Jeffrey B. Arnold

2017-01-03

Contents

Chapter 1

Introduction

Notes used when teaching "POLS/CS&SS 501: Advanced Political Research Design and Analysis" and "POLS/CS&SS 503: Advanced Quantitative Political Methodology" at the University of Washington.

Chapter 2

Writing Resources

2.1 Writing and Organizing Papers

Here are a few useful resources for writing papers:

- Chris Adolph. Writing Empirical Papers: 6 Rules & 12 Recommendations
- Barry R. Weingast. 2015. Caltech Rules for Writing Papers: How to Structure Your Paper and Write an Introduction
- The Science of Scientific Writing *American Scientist*
- Deidre McCloskey. Economical Writing
- William Thompson. A Guide for the Young Economist. "Chapter 2: Writing Papers."
- Stephen Van Evera. Guide to Methods for Students of Political Science. Appendix.
- Joseph M. Williams and Joseph Bizup. Style: Lessons in Clarity and Grace
- Strunk and White. *The Elements of Style*
- Chicago Manual of Style and APSA Style Manual for Political Science for editorial and style issues.
- How to construct a Nature summary paragraph. Though specifi to *Nature* is good advice for structuring abstracts or introductions.
- Ezra Klein. How researchers are terrible communications, and how they can do better.
- The advice in the *AJPS* Instructions for Submitting Authors is a concise description of how to write an abstract:

  > The abstract should provide a very concise descriptive summary of the research stream to which the manuscript contributes, the specific research topic it addresses, the research strategy employed for the analysis, the results obtained from the analysis, and the implications of the findings.

- Concrete Advice for Writing Informative Abstracts and pHow to Carefully Choose Useless Titles for Academic Writing](http://www.socialsciencespace.com/2014/03/how-to-carefully-choose-useless-titles-for-academic-w

2.2 Finding Research Ideas

Paul Krugman How I Work. His basic rules are

1. Listen to the Gentiles

2. Question the question
3. Dare to be silly
4. Simplify, simplify

See Hal Varian. How to build an Economic Model in your spare time

> The first step is to get an idea. This is not all that hard to do. The tricky part is to get a good idea. The way you do this is to come up with lots and lots of ideas and throw out all the ones that aren't good.
>
> But where to get ideas, that's the question. Most graduate students are convinced that the way you get ideas is to read journal articles. But in my experience journals really aren't a very good source of original ideas. You can get lots of things from journal articles—technique, insight, even truth. But most of the time you will only get someone else's ideas. …
>
> My suggestion is rather different: I think that you should look for your ideas outside the academic journals—in newspapers, in magazines, in conversations, and in TV and radio programs. …
>
> Conversations, especially with people in business, are often very fruitful. Commerce is conducted in many ways, and most of them have never been subjected to a serious economic analysis. …
>
> In many cases your ideas can come from your own life and experiences. …
>
> Before you start trying to decide whether your idea is correct, you should stop to ask whether it is interesting. If it isn't interesting, no one will care whether it is correct or not. So try it out on a few people—see if they think that it is worth pursuing. What would follow from this idea if it is correct? …
>
> The first thing that most graduate students do is they rush to the literature to see if someone else had this idea already. However, my advice is to wait a bit before you look at the literature. Eventually you should do a thorough literature review, of course, but I think that you will do much better if you work on your idea for a few weeks before doing a systematic literature search….

Greg Mankiw, My Rules of Thumb:

> Coming up with ideas is the hardest and least controllable part of the research process. It is somewhat easier if you have broad interests. Most obviously, broad interests give you more opportunities for success. A miner is more likely to strike gold if he looks over a large field than over the same field over and over again

Also the links in this Greg Mankiw Advice for Grad Students

2.3 Replications

Gary King has advice on how to turn a replication into a publishable paper:

- Gary King How to Write a Publishable Paper as a Class Project
- Gary King. 2006. "Publication, Publication." *PS: Political Science and Politics*.
- Political Science Should Not Stop Young Researchers from Replicating from the Political Science Replication blog.

And see the examples of students replications from his Harvard course at https://politicalsciencereplication.wordpress.com/

Famous replications.

- David Broockman, Joahua Kalla, and Peter Aronow. 2015. Irregularities in LaCour (2014).
- Homas Herndon, Michael Ash & Robert Pollin (2013). Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff. Working Paper Series 322. Political Economy Research Institute. [URL]

However, although those replications are famous for finding fraud or obvious errors in the analysis, replications can lead to extensions and generate new ideas. This was the intent of Brookman, Kalla, and Aronow when starting the replication.

Chapter 3

Typesetting and Word Processing Programs

3.1 LaTeX

LaTeX is a document markup language (think something like HTML) that is widely used in academia.[1] Its primary advantages over Word (and word processors) are the separation of content and presentation and its formatting of mathematical equations. In addition to papers, it is often used for academic slides; many talk slides are prepared with beamer.

3.1.1 Learning LaTeX

Here are some links to get started learning LaTeX:

- Overleaf Free & Interactive Online Introduction to LaTeX
- LaTeX Tutorial has interactive lessons
- ShareLaTeX Documentation
- Overleaf Example Templates has many different examples of LaTeX documents.
- LaTeX Wikibook
- Not So Short Introduction to LaTeX is a classic, but not as as newby friendly as the others.

3.1.2 Using LaTeX

- Use an online service such as Overleaf or ShareTeX. These are great for collaboration, but become inflexible when you want to customize your workflow.
- Write it with a specialized editor such as TeXmaker, TeXStudio, or TeXshop. These generally have built ways to insert text, and also live preview. I would stay away from editors such as LyX that are WYSWYG.
- Write it with an general purpose editor such as Atom or Sublime Text.[2] Most editors have a plugin to make writing LaTeX easier. For Atom there is LaTeXTools, and for Sublime Text, LaTeXTools

[1] TeX is pronounced as "teck" because the X is a Greek chi. The pronunciation of of LaTeX is thus lah-teck or lay-teck. It is not pronounced like the rubber compund. See this StackExchange question on the pronunciation of LaTeX.

[2] And of course Vim or Emacs.

3.1.3 LaTeX with R

This is pretty easy. Rnw, also called Sweave, documents allow you to mix R chunks with LaTeX. This is similar to R markdown, but with LaTeX instead of markdown.[3] knitr.

Many packages, such as xtable](https://cran.r-project.org/package=xtable), stargazer, or texreg produce formatted output in LaTeX. When you use these programs, do not copy and paste the output. Instead, save it to a file, and use `\input{}` to include the contents in your document.

3.1.4 My workflow

At the moment, I use emacs with AucTeX for writing pure LaTeX documents. However, I've started using Atom more and more as a general text editor. I use RStudio for writing Rnw files. For papers, I haven't moved to using (R) markdown and pandoc yet, because I find the layout and reference support still too limiting; I end up writing enough LaTeX that I figure I might as well write the whole thing in LaTeX. I've been using Overleaf for coauthoring in a couple of cases. I often build my document using Makefiles.

3.2 Word

While I use LaTeX in my own work, Microsoft Word is powerful piece of software, and many of the complaints against Word come down to not being aware of its features. There are many tools you can use to build your research paper; whatever tool you use, learn how to use it proficiently.

3.2.1 General Advice

This guide on using Microsoft Word for Dissertations covers everything and more that I would have. Also see this

- separate presentation and content using styles
- Automatically number figures and tables
- Use a reference manager like Mendeley, Zotero, Colwiz, or Papers. They have plugins for citations in Word.
- When exporting figures for Word, if you must use a raster graphic use PNG files (not JPEG). For publication, use a high DPI (600) with PNG graphics.
- Learn to use *Fields*. You can insert figures from files that you can update using `Insert > Field > Links and References > IncludePicture`. This is useful for programmatically generating figures to insert into your document. Likewise, you can insert text from files that you can update using `Insert > Field > Links and References > IncludeText`.

3.2.2 Using R with Word

For a dynamic reports you can use R Markdown and export to a word document. When doing this, use a reference document to set the the styles that you will use. See Happy collaboration with Rmd to docx for more advice on using R Markdown with Word.

When using functions from packages such as xtable, stargazer, or texreg output HTML, which can be copy and pasted into word.

Finally, the [ReporteR][Create Word documents from R](http://davidgohel.github.io/ReporteRs/word.html) package is an alternative method to generate Word Documents from R.

[3] And Sweave files preceded R markdown and knitr by many years.

Chapter 4

Coming to R from other Languages

4.1 Stata

There are not as many of these as I would have expected. These are the only useful ones that I found.

- Matthieu Gomez "R for Stata Users". http://www.princeton.edu/~mattg/statar/
- EconometricsBySimulation. "Dictionary: Stata to R" https://github.com/EconometricsBySimulation/RStata/wiki/Dictionary:-Stata-to-R
- DataCamp course *R for SAS, SPSS and Stata Users". https://www.datacamp.com/courses/r-for-sas-spss-and-stata-users-r-tutorial
- Muenchen, Robert A. and Joseph Hilbe. 2010. *R for Stata Users* http://link.springer.com/book/10.1007%2F978-1-4419-1318-0. Useful, although R has changed a surprisingly large amount in the last four years. For example, this does not include RStudio. *Note that if you're unaware, almost all Springer books are available through the library*
- Oscar Torres-Reyna, "Getting Started in R~Stata Notes on Exploring Data" http://www.princeton.edu/~otorres/RStata.pdf.