

Knit APA Style with \LaTeX , knitr and R

Bodong Chen

University of Toronto

Author Note

Word count: 2000. Correspondence concerning this article should be addressed to Bodong Chen, bodong.chen@utoronto.ca, 252 Bloor St W, Institute for Knowledge Innovation and Technology, OISE/University of Toronto.

Abstract

Starting my journey of doctoral thesis writing, I decided to rethink my academic writing environment which for years has been mainly composed of Microsoft Word and Google Documents. After some research, I finally decided to do something different for my thesis writing. In a nutshell, I am going to do everything (i.e., data analysis and actual writing) in LaTeX and R within the RStudio environment, with *knitr* as a wonderful bridge to “knit” them together. I am pretty happy with this setup, and believe it is much superior to my old writing environment especially for a bigger writing project like a thesis. This article, together with its code hosted on Github¹, aims to demonstrate the benefits and easiness of writing APA manuscripts in LaTeX and R. More specifically, this article briefly introduces key issues including environment setup, APA formatting with LaTeX, data analysis with R, and results integration with knitr. At the end, it will also describe ways to export produced documents to various formats to support collaboration with colleagues using Word-like software. My little ambition is anyone with no prior experience with LaTeX or R should feel prepared to play with this environment after reading this article.

Keywords: latex, apa, academic writing, r, reproducible research

¹<https://github.com/dirkchen/knit-r-apa6>

Knit APA Style with L^AT_EX, knitr and R**Why LaTeX and R**

As far as I know, only two colleagues in my department are using R², and I am probably the only person who uses LaTeX³. This situation is not surprising at all because working with LaTeX and R is not common in the field of education, or even social sciences in general. Without the need to deal with mathematic equations like our science or engineering colleagues, why should we suffer the pains of learning “alien languages” like LaTeX and R which are created for “geeks”? However, another kind of suffering might be widely experience by colleagues in my field. We constanly find ourselves typesetting our manuscripts for all kinds of creative formatting requirements we encounter in our adademic lives. Even people in my field usually follows an established standard, i.e., APA Style, dealing with headings, numbering of pages and headings is still a pain. Moreover, I feel increasingly troublesome working in Microsoft Word and SPSS especially for big writing projects like a thesis, finding myself constantly switching between pieces of software, tweaking data analysis results and inserting them as figures and tables into the manuscript. At the corner of launching my PhD dissertation writing project, which is expected to go beyond 200 pages based on my impression of my field, I am firmly convinced I should try something else that is less painful.

Luckily, I got to know LaTeX around two year ago and then R eight months ago, which now turn to my Swiss Army knife for academic writing (and data analysis). My main experience with Latex was writing my comprehensive exam last year⁴. And my R knowledge is also limited as I still constantly find myself doing basic stuff of R in stupid ways. Even though I am not an expert in either, I had a lot of fun with them (especially R⁵). However, LaTeX and R have been thought to live in two separate worlds—one for

²<http://www.r-project.org/>

³<http://www.latex-project.org/>

⁴My blog post about writing comps: <http://bodongchen.com/blog/?p=200>

⁵My blog posts related to R: <http://bodongchen.com/blog/?cat=56>

typesetting and the other for statistical data analysis—until one day I got to know *knitr*⁶. Knitr is an R package for dynamic report generation. It builds on a package called Sweave⁷, combining features of Sweave and its other add-ons into one single package. This work is also closely linked to the notion of reproducible research (Peng, Dominici, & Zeger, 2006). With knitr, people can combine data and analysis into final reports and make research fully reproducible, for later reproduction by both themselves and their interested readers. Knitr became the final piece that convinced me an integrated writing environment (IWE) is possible. The main idea is I could write manuscripts in LaTeX, embed data analysis as chunks of R code into LaTeX with knitr/Sweave, and compile everything into a PDF for sharing or publication.

Inspired by Zahn's (2006) effort to document "sweaving" in APA Style, I decided to document my endeavor of "knitting" in APA Style. Because his document was written almost five years ago, hopefully my demo could show a few improved techniques. More importantly, rather than writing and sharing a PDF manual, I am also sharing the actual code of the article I am writing right now so it becomes a self-contained learning material for colleagues who may wish to try it.

Environment Setup

Required Software Packages

This environment requires a number of software packages.

- LaTeX
- R (base)
- RStudio
- R packages: knitr, apa6, xtable, ggplot2
- git

⁶<http://yihui.name/knitr/>

⁷<http://www.stat.uni-muenchen.de/~leisch/Sweave/>

- Reference management tool such as Mendeley and RefJab

Workflow in RStudio

<http://www.rstudio.com/ide/docs/>

Working with Chapters or Sections

Managing References

bibtex, Mendeley, ...

Version Controlling with Git

Formatting an APA Document

<http://dw.tug.org/pracjourn/2012-1/beitzel/beitzel.pdf>

Dynamic results integration with knitr

Data analysis in R chunks:

```
data(cars) # load dataset
str(cars)  # structure of data frame

## 'data.frame': 50 obs. of 2 variables:
## $ speed: num  4 4 7 7 8 9 10 10 10 11 ...
## $ dist : num  2 10 4 22 16 10 18 26 34 17 ...

print(cor(cars$speed, cars$dist)) # compute correlation

## [1] 0.8069
```

Integrate results in text:

The average speed was 15.4, with a standard deviation of 5.2876. The average level of distance was 42.98, with a standard deviation of 25.7694. The correlation between income and education was 0.8069.

Embedding R code of data analysis

Including tables

Insert a table containing analysis results with xtable package:

Insert a pure text table with xtable:

Including figures

Insert a figure produced by ggplot2 in R:

Exporting to other formats

pandoc

online converting tool

pdftohtml

sudo apt-get install xpdf sudo apt-get install poppler-utils

pdftohtml Master.pdf Master.html pdftohtml -c Master.pdf Master.html

Conclusions

To appear...

References

- Peng, R. D., Dominici, F., & Zeger, S. L. (2006, May). Reproducible epidemiologic research. *American journal of epidemiology*, *163*(9), 783–9. doi:10.1093/aje/kwj093
- Zahn, I. (2006). Learning to sweave in apa style. *The PracTeX Journal*, *1*.

Table 1

Descriptive Statistics

	Mean	SD	Min	Max
Sepal Length	5.84	0.83	4.30	7.90
Sepal Width	3.06	0.44	2.00	4.40

Table 2

The three-phase research design

	Phase 1	Phase 2	Phase 3
Grade 3 2011	KB	KB	KB
	Pedagogical Intervention	Promisingness Judgments	
	Promisingness Judgments		
Grade 3 2010	KB	KB	KB



Figure 1. A scatterplot of sepal width and length of iris species.



Figure 2. Quality vs. Time