Integrating analysis code and document preparation : A minimal example Rmarkdown + papaja document

Tom Stafford[1] & Able Coauthor[2]

[1] University of Sheffield
[2] University of Bergen

This is an example of a note

R is a statistical programming language which is increasingly popular with psychologists. It can import and process your data, fit statistical models (from simple t-tests to state of the art such as bayesian multilevel model fitting). It also makes nice plots. RStudio is a way of editing R scripts and running R analysis. RMarkdown is a way of using RStudio to produce documents (e.g. as webpages, MS Word or PDF). Another advantage is that you can include R code in your document file - so no more running your analysis in SPSS and copying the results into your document (and making errors / forgetting which version of the analysis you ran etc). This is an example document which integrates all the functions of Rmarkdown - running analysis, formatting references, etc. It uses an add-on for Rmarkdown called papaja which helps us make nicely APA formatted documents

Introduction

This is an example of a reproducible document - it integrates data analysis, visualisation and document preparation. No more copy/paste errors with your statistics! No more wondering which of FINAL_DATA.xls and FINAL_FINAL_DATA.xls is the right data set to produce the version of the graph you liked. Lovely looking preprints!

Rmarkdown is good (Allaire et al., 2020). Need to change reference style? Change one line. Need to submit as PDF rather than .DOC? Just click 'Word' as output rather than 'PDF' (instructions here https://rmarkdown. rstudio.com/articles_docx.html). Need to change to two column style to make a nice pre-print? Again, simple - just change one line! In line 53 'class : "man"' gives you manuscript style; "jou" gives you two column style.

The heart of this is the .rmd document. You make a PDF (or whatever) by opening .rmd file in RStudio and clicking 'knit'

Example Subheading

To get started, let's look at some basics: formatting and referencing. You might want to compare how they look in the .rmd file and in the PDF.

Here are some example references in the following sentence. Two good reviews on speed-accuracy trade-offs are Wickelgren (1977); Heitz (2014). References can also be inside the brackets (e.g. Stafford, Pirrone, Croucher, & Krystalli, 2020). This paper was produced as part of the 'code club' held in the Department of Psychology at the University of Sheffield in 2019, organised by Tom Stafford and Kat Giannadou. Thanks to Frederick Aust for feedback over twitter and help with papaja[1]

Method

Rmarkdown also lets us track figure labels, and updates them automatically. Look! Kittens! Illustrated in Figure 1. And if I add other figures it will sort out the numbering and references automatically

---

[1] And massive kudos for developing the package in the first place. Also this is an example of a footnote.

Figure 1. Example figure caption

Requirements

You should install R, RStudio and tex and papaja (Aust & Barth, 2020). More details here https://crsh.github.io/ papaja_man/introduction.html#getting-started. The contributor roles are autogenerated using the Tenzing package (Holcombe, Kovacs, Aust, & Aczel, 2020). To adjust start by visting https://martonbalazskovacs.shinyapps.io/tenzing/ or edit the yaml frontmatter directly. Use the shinyapp to generate the contributor roles in various different formats (including the yaml we use here).

Results

Now let's integrate some R code to generate/import some data, run and analyse and integrate it into the document:

You can't see it in the PDF, but in between this paragraph and the last we asked R to generate some random data and save it to a CSV file. Now we're going to import the data from the CSV file, as if it was independently created data - from an experiment or similar - and plot a graph.
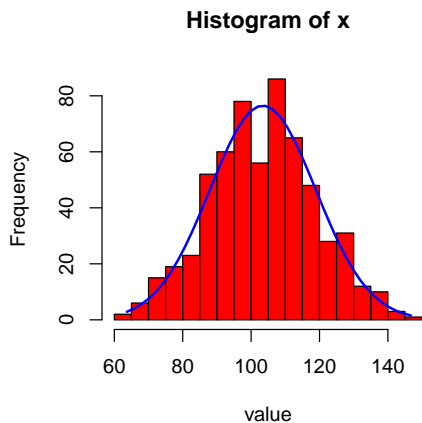

Figure 2. Histogram of all data, grouped

Table 1
Item grouping to subscales

	Hyperactivity	Inattention
ASRS-18	5,6,12,13,14	1,2,3,4,7,8
ASRS-6	5,6	1,2,3,4

Note. This table was created with apa_table().

See Figure 2. Of course we could draw all sorts of things, but this is a proof-of-concept. Finally, let's run a t-test and integrate the results into the text.

We found there was a statistically significant difference between the two groups (t=-6.35 (593.00), p = 0.00). Note how the exact values in the previous sentence change every time we re-make the document (because the document also re-generates the underlying data).

The above paragraph does the integration of statistics longform, because I didn't know a better way (and because I think it is still useful to show). apa_print() exists as a helper function: http://frederikaust.com/papaja_man/reporting.html.

Also, everyone loves tables. See Table 1. I feel obliged to mention that the actual items reported in this table **do not** correctly reflect the actual subscales for the ASRS short and long scales. Note illustration of **bold** and *italics* formatting in this para.

Discussion

To port to your own project just copy across these files:

- example_manuscript.rmd
- references.bib - information on references in bibtex format

And I recommending creating:

- figs folder - where images integrated into the manuscript are kept

(although you can store images anywhere as long as you get path referencing right). You will find it useful to compare the output PDF document with the .rmd document. This latter item is the thing you edit to produce the PDF. We can produce outputs in various formats (manuscript, journal) as well as PDF or MS Word.

Of course, there's more effort in installing and learning and correctly marking up your document in the first place, but it is worth it.

Unanswered questions:

What is the best way to define figure sizes so you get nice and/or consistent sizing across document formats?

References

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., ... Iannone, R. (2020). *Rmarkdown: Dynamic documents for r*. Retrieved from https://github.com/rstudio/rmarkdown

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*, 150.

Holcombe, A. O., Kovacs, M., Aust, F., & Aczel, B. (2020). Documenting contributions to scholarly articles using CRediT and tenzing. *PLoS One*, *15*(12), e0244611.

Stafford, T., Pirrone, A., Croucher, M., & Krystalli, A. (2020). Quantifying the benefits of using decision models with response time and accuracy data. *Behavior Research Methods*, *52*, 2142--2155.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85.