

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

Tom Stafford¹ & Able Coauthor²

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

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.

Example Subheading

Here are some example references in the following sentence. For reviews of this topic see Wickelgren (1977); Heitz (2014). Here is another example reference (e.g. Stafford, Pirrone, Croucher, & Krystalli, 2020)¹. 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.

Method

Rmarkdown also lets us track figure labels, and updates them automatically. Look! Kittens! Illustrated in Figure 1.



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). Read more <https://medium.com/@ceptional/announcing-tenzing-ceca6789d88c> and <https://apastyle.apa.org/blog/author-note>.

Contents of this author note autogenerated using Tenzing, visit <https://martonbalazskovacs.shinyapps.io/tenzing/> to complete for your team

The authors made the following contributions. Tom Stafford: Conceptualization, Project Administration, Visualization, Writing - Original Draft Preparation, Writing - Review & Editing; Able Coauthor: Funding Acquisition, Conceptualization.

Correspondence concerning this article should be addressed to Tom Stafford, Enter postal address here. E-mail: t.stafford@sheffield.ac.uk

¹As well as an example of a footnote.

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()`.

Do adjust start by visting <https://martonbalazskovacs.shinyapps.io/tenzing/> or edit the yaml directly. The file `tenzing.csv` can be uploaded to 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, 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.

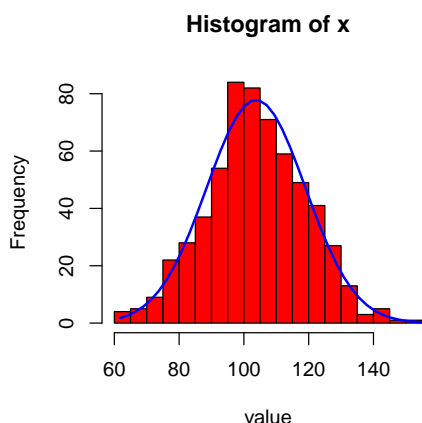


Figure 2. Histogram of all data, grouped

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 ($t=-4.76$ (587.65), $p = 0.00$). 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).

Unanswered questions: Is this the best way to integrate values into text? Why is the df not an integer? What is the best way to define figure sizes so you get nice and/or consistent sizing across document formats?

Also, everyone loves tables. See Table 1. Currently I can't work out how to get it to appear inside a single column of the two-column "jou" output format. So it appears on the final page after the references. If you know how to solve this, get in touch please.

Discussion

Make your document by opening .Rmd file in RStudio and clicking 'knit'

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 40 'class : "man"' gives you manuscript style; "jou" gives you two column style.

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

- `example_manuscript.rmd`
- `apa6.cls` - style file which makes everything look APA format nice
- `references.bib` - information on references in bibtex format
- `figs` folder - where images integrated into the manuscript are kept

Main conclusions

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.

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.