# How to use papaja: An Example Manuscript Including Basic Instructions

Cengiz Zopluoglu[University of Oregon]

# Abstract

This manuscript demonstrates how to use R Markdown and papaja to create an APA conform manuscript. papaja builds on R Markdown, which uses pandoc to turn Markdown into PDF or Word documents. The conversion to Word documents currently supports only a limited set of features.

How to use papaja: An Example Manuscript Including Basic Instructions

# What is papaja?

Reproducible data analysis is an easy to implement and important aspect of the strive towards reproducibility in science. For *R* users, R Markdown has been suggested as one possible framework for reproducible analyses. `papaja` is a R-package in the making including a R Markdown template that can be used with (or without) RStudio to produce documents, which conform to the American Psychological Association (APA) manuscript guidelines (6th Edition). The package uses the LATEXdocument class apa6 and a .docx-reference file, so you can create PDF documents, or Word documents if you have to. Moreover, `papaja` supplies R-functions that facilitate reporting results of your analyses in accordance with APA guidelines.

Markdown is a simple formatting syntax that can be used to author HTML, PDF, and MS Word documents (among others). In the following I will assume you know how to use R Markdown to conduct and comment your analyses. If this is not the case, I recommend you familiarize yourself with R Markdown first. I use RStudio to create my documents, but the general process works with any text editor.

# How to use papaja

Once you have installed `papaja` and all other required software, you can select the APA template when creating a new R Markdown file through the RStudio menus, see Figure **??**. When you click RStudio's *Knit* button, `papaja`, `bookdown`, `rmarkdown`, and `knitr` work together to create an APA conform manuscript that includes both your text and the output of any embedded R code chunks within the manuscript.

If you don't use RStudio, you can create new `papaja` documents via `rmarkdown::draft()` and `rmarkdown::render()`.

## Printing R output

Any output from R is included as you usually would using R Markdown. By default the R code will not be displayed in the final documents. If you wish to show off your code you need to set `echo = TRUE` in the chunk options. For example, to include summary statistics of your data you could use the following code:

```
##       sid           reading             ses               race
##  Min.   :1002   Min.   : 33.14   Min.   :-2.33000   A  :1259
##  1st Qu.:1342   1st Qu.: 46.73   1st Qu.:-0.66500   AI : 338
##  Median :1678   Median : 53.03   Median :-0.13900   B  :2295
##  Mean   :1672   Mean   : 54.44   Mean   :-0.04386   H  :2930
##  3rd Qu.:2006   3rd Qu.: 58.94   3rd Qu.: 0.53500   HPI: 173
##  Max.   :2330   Max.   :164.33   Max.   : 2.59600   W  :7074
```

But, surely, this is not what you want your submission to look like.

**Print tables.** For prettier tables, I suggest you try `apa_table()`, which builds on `knitr`'s `kable()`, and `printnum()`, which can be used to properly round and report numbers. For the table to display correctly set the chunk option `results = "asis"` in the chunk that produces the table.

Of course popular packages like `xtable`[1] or `tables` can also be used to create tables when knitting PDF documents. These packages, however, cannot be used when you want to create Microsoft Word documents because they rely on LaTeXfor typesetting. `apa_table()` creates tables that conform to APA guidelines and are correctly rendered in PDF and Word documents. But don't get too excited; table formatting is somewhat limited for Word documents due to missing functionality in pandoc (e.g., it is not possible to have cells or headers span across multiple columns).

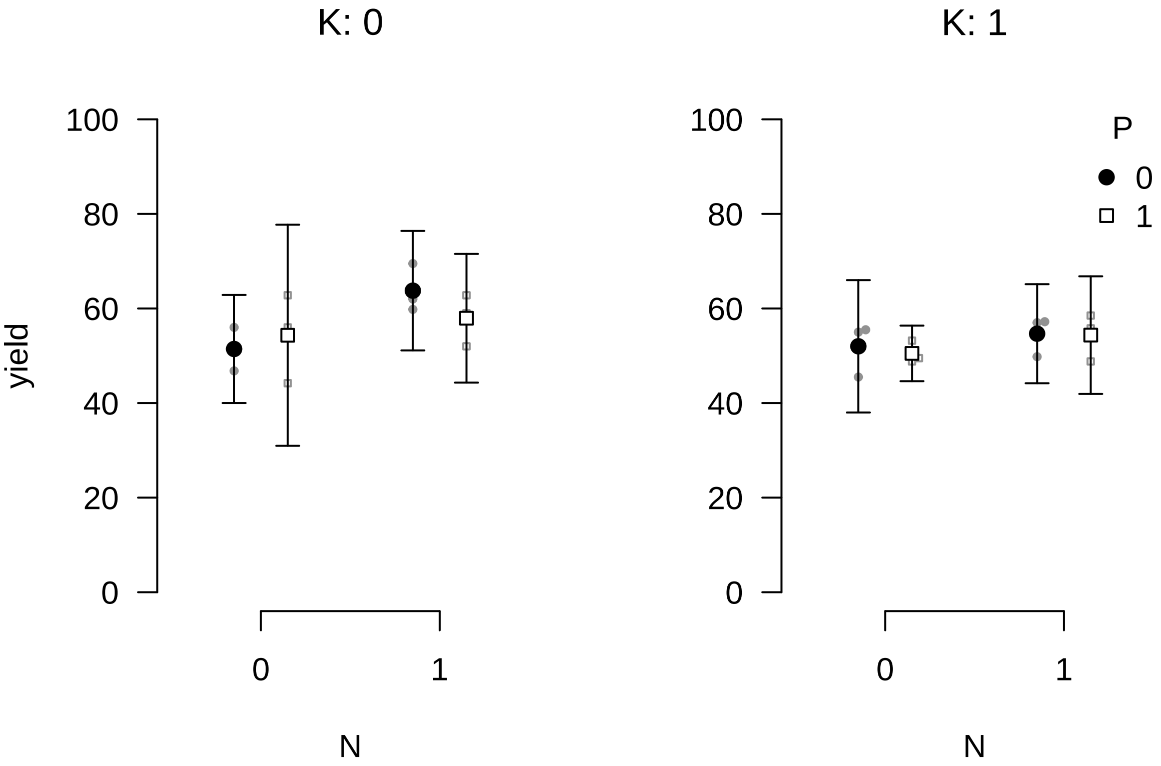
[1] When you use `xtable()`, table captions are set to the left page margin.

As required by the APA guidelines, tables are deferred to the final pages of the manuscript when creating a PDF. Again, this is not the case in Word documents due to limited pandoc functionality. To place tables and figures in your text instead, set the `figsintext` parameter in the YAML header to `yes` or `true`, as I have done in this document.

The bottom line is, Word documents will be less polished than PDF. The resulting documents should suffice to enable collaboration with Wordy colleagues and prepare a journal submission with limited manual labor.

**Embed plots.** As usual in R Markdown, you can embed R-generated plots into your document, see Figure .

\begin{figure}



{

}

\caption{Bee plot of the example data set. Small points represent individual observations, large points represent means, and error bars represent 95% confidence intervals.}
\end{figure}

Again, as required by the APA guidelines, figures are deferred to the final pages of the document unless you set `figsintext` to `yes`.

**Referencing figures and tables.** `papaja` builds on the `bookdown` package, which provides limited cross-referencing capabilities within documents. By default you can insert figure and table numbers into the text using `\@ref(fig:chunk-name)` for figures or `\@ref(tab:chunk-name)` for tables. Note that for this syntax to work chunk names cannot include `_`. If you need to embed an external image that is not generated by R use the `knitr::include_graphics()` function. See the great book on `bookdown` for details. Cross-referencing is currently not available for equations in `bookdown`. However, as anywhere in R Markdown documents you can use LaTeXcommands if the functionality is not provided by `rmarkdown/bookdown` and you don't need to create Word documents.

**Report statistical analyses.** `apa_print()` will help you report the results of your statistical analyses. The function will format the contents of R objects and produce readily reportable text.

Now, you can report the results of your analyses like so:

Race is related to reading scores after accounting for SES, $F(5, 14, 063) = 184.91$, $MSE = 127.63$, $p < .001$, $\hat{\eta}^2_p = .062$.

What's even more fun, you can easily create a complete ANOVA table using by passing `mod_results$table` to `apa_table()`, see Table 2.

## Citations

No manuscript is complete without citation. In order for citations to work, you need to supply a .bib-file to the `bibliography` parameter in the YAML front matter. Once this is done, `[e.g., @james_1890; @bem_2011]` produces a regular citation within parentheses (e.g., Bem, 2011; James, 1890). To cite a source in text simply omit the brackets; for example, write `@james_1890` to cite James (1890). For other options see the overview of the R Markdown citation syntax.

The citation style is automatically set to APA style. If you need to use a different citation style, you can set in the YAML front matter by providing the `csl` parameter. See the R Markdown documentation and Citation Style Language for further details.

If you use RStudio, I have created an easy-to-use add-in that facilitates inserting citations into a document. The relevant references will, of course, be added to the documents reference section automatically. Moreover, the addin can directly access you Zotero database.

I think it is important to credit the software we use. A lot of R packages are developed by academics free of charge. As citations are the currency of science, it's easy to compensate volunteers for their work by citing the R packages we use. I suspect that, among other things, this is rarely done because it is tedious work. That's why papaja makes citing R and its packages easy:

`r_refs()` creates a BibTeX file containing citations for R and all currently loaded packages. `cite_r()` takes these citations and turns them into readily reportable text. `my_citation` now contains the following text that you can use in your document: R (Version 4.0.2; R Core Team, 2020) and the R-packages *afex* (Version 0.28.0; Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2020), *dplyr* (Version 1.0.2; Wickham, François, Henry, & Müller, 2020), *lme4* (Version 1.1.23; Bates, Mächler, Bolker, & Walker, 2015), *Matrix* (Version 1.2.18; Bates & Maechler, 2019), and *papaja* (Version 0.1.0.9997; Aust & Barth, 2020)

## Math

If you need to report formulas, you can use the flexible LaTeXsyntax (it will work in Word documents, too). Inline math must be enclosed in `$` or `\(` and `\)` and the result will look like this: $d' = z(H) - z(FA)$. For larger formulas displayed equations are more appropriate; they are enclosed in `$$` or `\[`and `\]`,

$$d' = \frac{\mu_{old} - \mu_{new}}{\sqrt{0.5(\sigma_{old}^2 + \sigma_{new}^2)}}.$$

## Document options

This text is set as manuscript. If you want a thesis-like document you can change the `class` in the YAML front matter from `man` to `doc`. You can also preview a polished journal typesetting by changing the `class` to `jou`. Refer to the `apa6` document class documentation for further `class` options, such as paper size or draft watermarks.

When creating PDF documents, line numbering can be activated by setting the `lineno` argument in the YAML front matter to `yes`. Moreover, you can create lists of figure or table captions at the end of the document by setting `figurelist` or `tablelist` to `yes`, respectively. These option have no effect on Word documents.

## Last words

That's all I have; enjoy writing your manuscript. If you have any trouble or ideas for improvements, open an issue on GitHub or open a pull request. If you want to contribute, take a look at the open issues if you need inspiration. Other than that, there are many output objects from analysis methods that we would like `apa_print()` to support. Any new S3/S4-method for this function are always appreciated (e.g., `factanal`, `fa`, `lavaan`, `BFBayesFactor`).

## References

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

Bates, D., & Maechler, M. (2019). *Matrix: Sparse and dense matrix classes and methods.* Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407—425. https://doi.org/10.1037/a0021524

James, W. (1890). *The principles of psychology.* Holt: New York.

R Core Team. (2020). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). *Afex: Analysis of factorial experiments.* Retrieved from https://CRAN.R-project.org/package=afex

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *Dplyr: A grammar of data manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

Table 1

*Descriptive statistics of reading scores by race*

| race | Mean | SD | Min | Max |
|---|---|---|---|---|
| A | 60.65 | 16.14 | 33.71 | 115.04 |
| AI | 51.29 | 9.68 | 33.78 | 102.66 |
| B | 52.63 | 10.37 | 33.41 | 113.45 |
| H | 50.44 | 9.68 | 33.14 | 107.71 |
| HPI | 54.58 | 12.52 | 34.30 | 104.81 |
| W | 55.72 | 11.20 | 33.22 | 164.33 |

*Note.* This table was created with apa_table()

Table 2

*ANOVA table for the analyis of the example data set.*

| Effect | $F$ | $df_1$ | $df_2$ | $MSE$ | $p$ | $\hat{\eta}^2_p$ |
|---|---|---|---|---|---|---|
| Race | 184.91 | 5 | 14,063 | 127.63 | $<$ .001 | .062 |

*Note.* This is a table created using apa_print() and apa_table().