

How to use papaja: An Example Manuscript Including Basic Instructions

Frederik Aust¹

¹ University of Cologne

Author Note

papaja has not yet been submitted to CRAN; a development version is available at <https://github.com/crsh/papaja>.

Correspondence concerning this article should be addressed to Frederik Aust, Department Psychology, University of Cologne, Herbert-Lewin-Str. 2, 50931 Köln, Germany. E-mail: frederik.aust@uni-koeln.de

Abstract

10

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

15 *Keywords:* APA style, knitr, R, R markdown, papaja

16 Word count: Too lazy to count

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 \LaTeX document 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 1](#). When you click RStudio's *Knit* button (see [Figure 2](#)), `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.

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

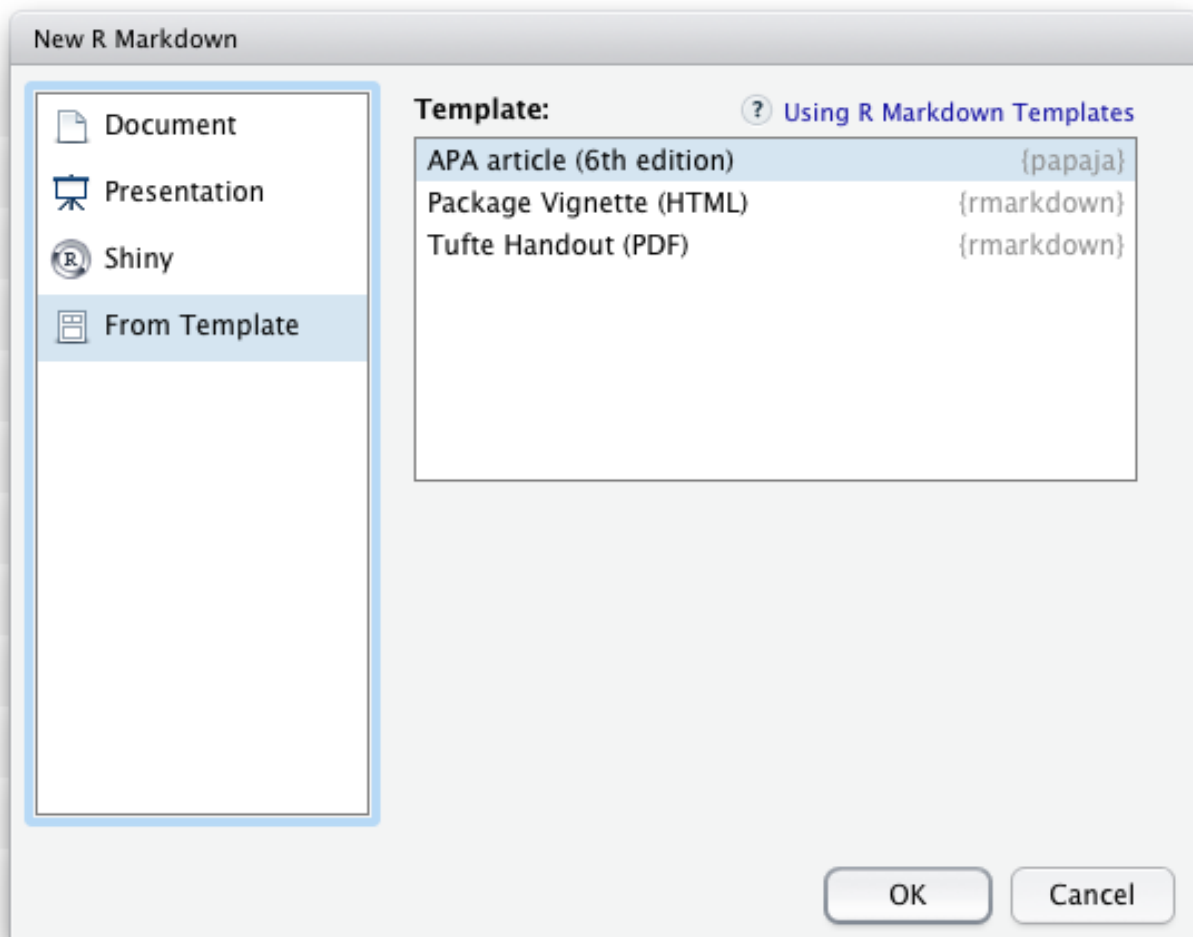


Figure 1. papaja’s APA6 template is available through the RStudio menus.

42 you need to set `echo = TRUE` in the chunk options. For example, to include summary
 43 statistics of your data you could use the following code:

```
summary(mixed_data[, -1])
```

```
44 ##      Subject  Gender Dosage Task  Valence      Recall
45 ##  A      : 6    F:54   A:36   C:54   Neg:36   Min.    : 4.00
46 ##  B      : 6    M:54   B:36   F:54   Neu:36   1st Qu.:13.00
47 ##  C      : 6                C:36                Pos:36   Median :15.00
48 ##  D      : 6                                Mean    :15.63
49 ##  E      : 6                                3rd Qu.:19.00
```

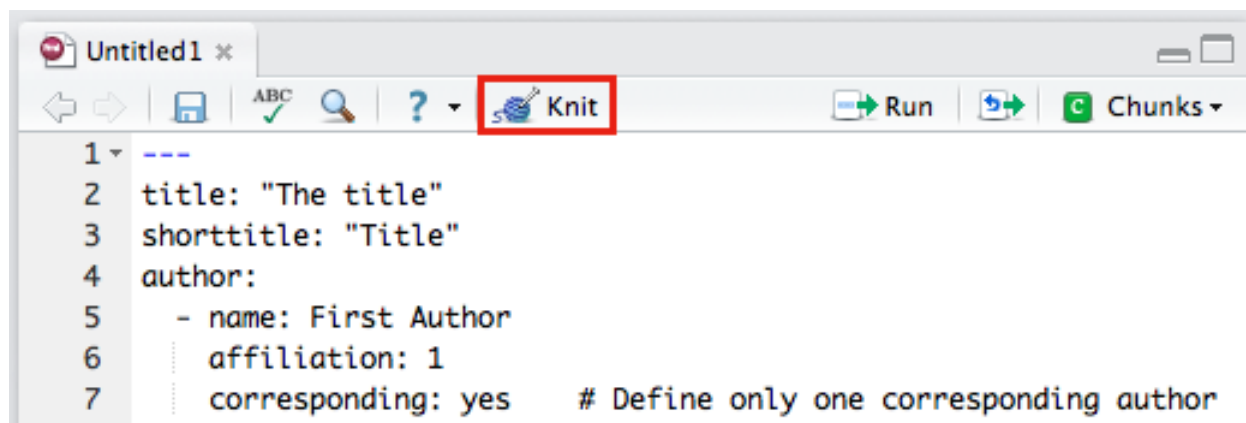


Figure 2. The Knit button in the RStudio.

```
50 ## F : 6 Max. :25.00
```

```
51 ## (Other):72
```

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

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

```
descriptives <- mixed_data %>% group_by(Dosage) %>%
  summarize(
    Mean = mean(Recall)
    , Median = median(Recall)
    , SD = sd(Recall)
    , Min = min(Recall)
    , Max = max(Recall)
  )
descriptives[, -1] <- printnum(descriptives[, -1])
```

Table 1

Descriptive statistics of correct recall by dosage.

Dosage	Mean	Median	SD	Min	Max
A	14.19	14.00	4.45	5.00	25.00
B	13.50	14.00	5.15	4.00	22.00
C	19.19	19.00	3.52	13.00	25.00

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

```
apa_table(
  descriptives
  , caption = "Descriptive statistics of correct recall by dosage."
  , note = "This table was created with apa_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 \LaTeX for 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).

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

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

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 3.

```
apa_beeplot(  
  mixed_data  
  , id = "Subject"  
  , dv = "Recall"  
  , factors = c("Task", "Valence", "Dosage")  
  , dispersion = conf_int  
  , ylim = c(0, 30)  
  , las = 1  
  , args_points = list(cex = 1.5)  
  , args_arrows = list(length = 0.025)  
)
```

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 $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ commands if the functionality is not provided by `rmarkdown/bookdown` and you don't need to create Word documents.

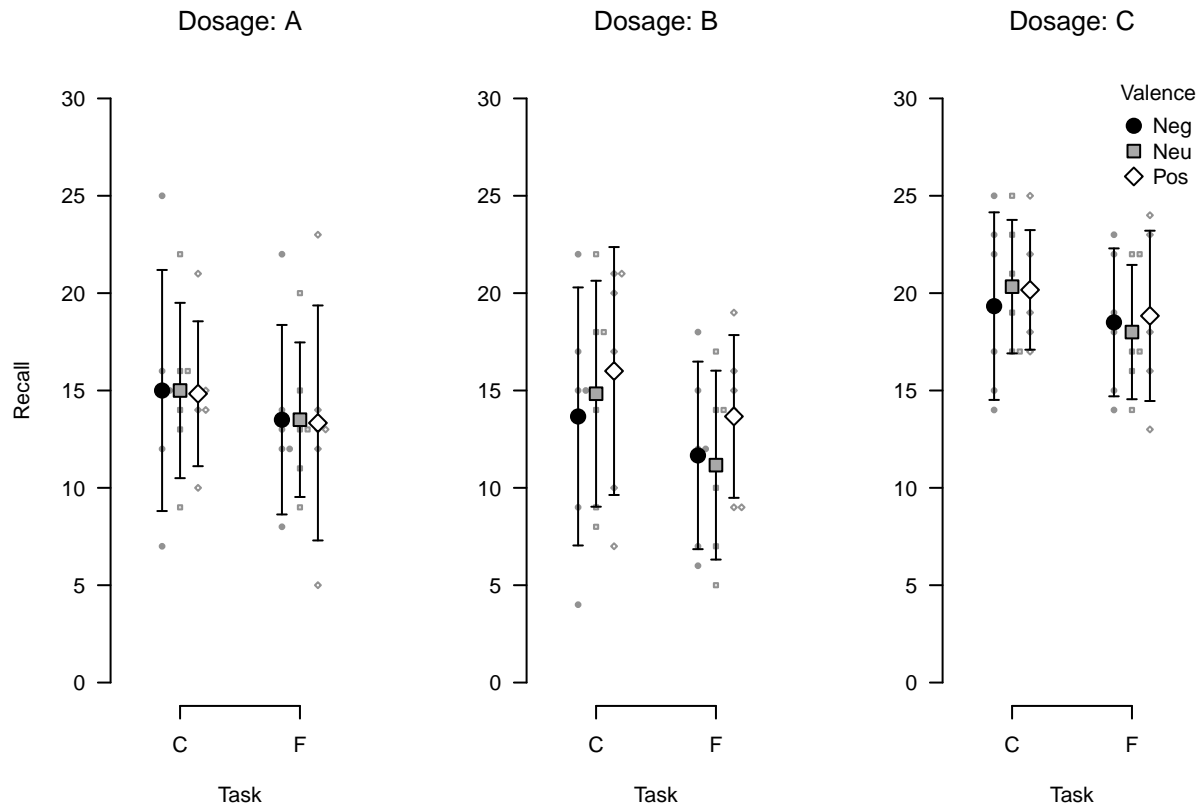


Figure 3. Bee plot of the example data set. Small points represent individual observations, large points represent means, and error bars represent 95% confidence intervals.

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

```
recall_anova <- afex::aov_car(
  Recall ~ (Task * Valence * Dosage) + Error(Subject/(Task * Valence)) + Dosage
  , data = mixed_data
  , type = 3
)
recall_anova_results <- apa_print(recall_anova, es = "pes")
recall_anova_results_p <- apa_print(recall_anova, es = "pes", in_paren = TRUE)
```

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


```
Item valence (`r anova_results_p$full$Valence`) and the task affected recall
performance, `r anova_results$full$Task`; the dosage, however, had no effect
on recall, `r anova_results$full$Dosage`. There was no significant interaction.
```

88 Item valence ($F[1.62, 24.36] = 3.46$, $MSE = 2.62$, $p = .056$, $\hat{\eta}_p^2 = .187$) and the
 89 task affected recall performance, $F(1, 15) = 43.13$, $MSE = 2.23$, $p < .001$,
 90 $\hat{\eta}_p^2 = .742$; the dosage, however, had no effect on recall, $F(2, 15) = 2.97$,
 91 $MSE = 117.17$, $p = .082$, $\hat{\eta}_p^2 = .283$. There was no significant interaction.

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

```
apa_table(
  recall_anova_results$table
  , align = c("l", "r", "c", "r", "r", "r")
  , caption = "ANOVA table for the analysis of the example data set."
  , note = "This is a table created using apa\\_print() and apa\\_table()."
  , escape = FALSE
)
```

94 Citations

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

Table 2

ANOVA table for the analysis of the example data set.

Effect	F	df_1^{GG}	df_2^{GG}	MSE	p	$\hat{\eta}_p^2$
Dosage	2.97	2	15	117.17	.082	.283
Task	43.13	1	15	2.23	< .001	.742
Valence	3.46	1.62	24.36	2.62	.056	.187
Dosage \times Task	1.83	2	15	2.23	.195	.196
Dosage \times Valence	2.38	3.25	24.36	2.62	.090	.241
Task \times Valence	1.50	1.35	20.2	2.67	.242	.091
Dosage \times Task \times Valence	0.39	2.69	20.2	2.67	.743	.049

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

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 `cs1` 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 your 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(file = "r-references.bib")
my_citation <- cite_r(file = "r-references.bib")
```

`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 3.4.2; R Core Team, 2015) and the R-packages *afex* (Version 0.18.0; Singmann, Bolker, Westfall, & Aust, 2016), *bindrcpp* (Version 0.2; Müller, 2017), *boot* (Version 1.3.20; Davison & Hinkley, 1997), *broom* (Version 0.4.3; Robinson, 2016), *dplyr* (Version 0.7.4; Wickham & Francois, 2016), *estimability* (Version 1.2; Lenth, 2015), *knitr* (Version 1.17; Xie, 2015), *lme4* (Version 1.1.14; Bates, Mächler, Bolker, & Walker, 2015), *lsmeans* (Version 2.27.2; Lenth, 2016), *Matrix* (Version 1.2.11; Bates & Maechler, 2016), *MBESS* (Version 4.4.1; Kelley, 2016), *papaja* (Version 0.1.0.9655; Aust & Barth, 2015), *reshape2* (Version 1.4.2; Wickham, 2007), *rmarkdown* (Version 1.8; Allaire et al., 2016), and *testthat* (Version 1.0.2; Wickham, 2011)

Math

If you need to report formulas, you can use the flexible \LaTeX syntax (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., `glm`, `factanal`, `fa`, `lavaan`, `BFBayesFactor`).

References

- Allaire, J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., ... Hyndman, R. (2016). *Rmarkdown: Dynamic documents for r*. Retrieved from <https://CRAN.R-project.org/package=rmarkdown>
- Aust, F., & Barth, M. (2015). *Papaja: Create apa manuscripts with rmarkdown*.
- Bates, D., & Maechler, M. (2016). *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. doi:[10.18637/jss.v067.i01](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. doi:[10.1037/a0021524](https://doi.org/10.1037/a0021524)
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their applications*. Cambridge: Cambridge University Press. Retrieved from

<http://statwww.epfl.ch/davison/BMA/>

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

Kelley, K. (2016). *MBESS: The mbess r package*. Retrieved from

<https://CRAN.R-project.org/package=MBESS>

Lenth, R. V. (2015). *Estimability: Tools for assessing estimability of linear predictions*.

Retrieved from <https://CRAN.R-project.org/package=estimability>

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical*

Software, 69(1), 1–33. doi:10.18637/jss.v069.i01

Müller, K. (2017). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from

<https://CRAN.R-project.org/package=bindrcpp>

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna,

Austria: R Foundation for Statistical Computing. Retrieved from

<http://www.R-project.org/>

Robinson, D. (2016). *Broom: Convert statistical analysis objects into tidy data frames*.

Retrieved from <https://CRAN.R-project.org/package=broom>

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *Afex: Analysis of factorial*

experiments. Retrieved from <https://CRAN.R-project.org/package=afex>

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical*

Software, 21(12), 1–20. Retrieved from <http://www.jstatsoft.org/v21/i12/>

Wickham, H. (2011). Testthat: Get started with testing. *The R Journal*, 3, 5–10. Retrieved

from http://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf

Wickham, H., & Francois, R. (2016). *Dplyr: A grammar of data manipulation*. Retrieved

from <https://CRAN.R-project.org/package=dplyr>

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida:

Chapman; Hall/CRC. Retrieved from <http://yihui.name/knitr/>