

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?

As you may have heard, recently, interest in reproducible research has been growing. Reproducible data analysis is an easy to implement and important aspect of the strive towards reproducibility. 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 L^AT_EXdocument 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 have hopped onto the band wagon and know how to use R Markdown to conduct and comment your analyses. If this is not the case, I recommend you get to grips with [R Markdown](#) first. I use [RStudio](#) (which makes use of [pandoc](#)) to create my documents, but the general process works using any other 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 Markdown file through the RStudio menus, see [Figure 2](#). When you click RStudio's *Knit* button (see ??), **papaja**, R Markdown, 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.

41 Printing R output

42 Any output from R is included as you usually would using R Markdown. By default
 43 the R code will not be displayed in the final documents. If you wish to show off your code
 44 you need to set `echo = TRUE` in the chunk options.

```
summary(mixed_data)
```

```
45 ##           Obs           Subject  Gender Dosage Task  Valence
46 ##  Min.      : 1.00    A          : 6    F:54   A:36   C:54   Neg:36
47 ## 1st Qu.: 27.75    B          : 6    M:54   B:36   F:54   Neu:36
48 ## Median : 54.50    C          : 6                C:36        Pos:36
49 ## Mean    : 54.50    D          : 6
50 ## 3rd Qu.: 81.25    E          : 6
51 ## Max.    :108.00    F          : 6
52 ##                               (Other):72
53 ##           Recall
54 ##  Min.      : 4.00
55 ## 1st Qu.:13.00
56 ## Median :15.00
57 ## Mean    :15.63
58 ## 3rd Qu.:19.00
59 ## Max.    :25.00
60 ##
```

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

62 **Print tables.** For prettier tables, I suggest you try `apa_table()`, which builds on
 63 `knitr's kable()`.

```

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])

apa_table(

  descriptives

  , caption = "Descriptive statistics of correct recall by dosage."

  , note = "This table was created with apa_table()"

  , escape = TRUE

)

```

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

71 As required by the APA guidelines, tables are deferred to the final pages of the
 72 manuscript when creating a PDF. To place tables and figures in your text instead, set the
 73 `figsintext` parameter in the YAML header to `yes` or `true`, as I have done in this

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

document. Again, this is not the case in Word documents due to limited pandoc functionality. 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 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. Currently, `rmarkdown` does not support proper cross-referencing with the document. For this document, I have simply interspersed my R markdown document with \LaTeX to accomplish cross-references. A new version of pandoc (1.16), that is not *yet* shipped with RStudio, supports proper cross-referencing that will work with Word documents as well.

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.

```
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)
```

90 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.
```

91 Item valence ($F[1.62, 24.36] = 3.46$, $MSE = 2.62$, $p = .056$, $\eta_p^2 = .187$) and the
 92 task affected recall performance, $F(1, 15) = 43.13$, $MSE = 2.23$, $p < .001$,
 93 $\eta_p^2 = .742$; the dosage, however, had no effect on recall, $F(2, 15) = 2.97$,
 94 $MSE = 117.17$, $p = .082$, $\eta_p^2 = .283$. There was no significant interaction.

95 What's even more fun, you can easily create a complete ANOVA table using by passing
 96 `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 analyis of the example data set. \\label{ref:anova}"
  , note = "This is a table created using apa\\_print() and apa\\_table()."
)
```

97 Citations

98 No manuscript is complete without citation. In order for citations to work, you need to
 99 supply a .bib-file to the `bibliography` parameter in the YAML front matter. Once this is
 100 done, [e.g., `@james_1890; @bem_2011`] produces a regular citation within parentheses
 101 (Bem, 2011; e.g., James, 1890). To cite a source in text simply omit the brackets; for
 102 example, write `@james_1890` to cite James (1890). For other options see the [overview of the](#)
 103 [R Markdown citation syntax](#). If you use RStudio, I have created an [easy-to-use add-in](#) that
 104 facilitates inserting citations. The relevant references will, of course, be added to the
 105 documents reference section automatically.

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

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

111 `r_refs()` creates a BibTeX file containing citations for R and all currently loaded
 112 packages. `cite_r()` takes these citations and turns them into readily reportable text.
 113 `my_citation` now contains the following text that you can use in your document: R (3.3.2,
 114 R Core Team, 2015) and the R-packages *afex* (0.16.1, Singmann, Bolker, Westfall, and Aust,
 115 2016), *boot* (1.3.18, Davison & Hinkley, 1997), *broom* (0.4.1, Robinson, 2016), *dplyr* (0.5.0,
 116 Wickham & Francois, 2016), *estimability* (1.1.1, Lenth, 2015), *knitr* (1.15.6, Xie, 2015), *lme4*
 117 (1.1.12, Bates, Mächler, Bolker, & Walker, 2015), *lsmeans* (2.23.5, Lenth, 2016), *Matrix*
 118 (1.2.7.1, Bates & Maechler, 2016), *MBESS* (4.0.0, Kelley, 2016), *papaja* (0.1.0.9464, Aust &
 119 Barth, 2015), *reshape2* (1.4.1, Wickham, 2007), *rmarkdown* (1.3, Allaire et al., 2016), and
 120 *testthat* (1.0.2, Wickham, 2011)

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 adding a `cs1` parameter. See the [R Markdown documentation](#) and [Citation Style Language](#) for further details.

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-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](https://doi.org/10.18637/jss.v069.i01)
- 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/>

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()`

Table 2

ANOVA table for the analysis of the example data set.

Effect	F	df_1^{GG}	df_2^{GG}	MSE	p	η_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()`.

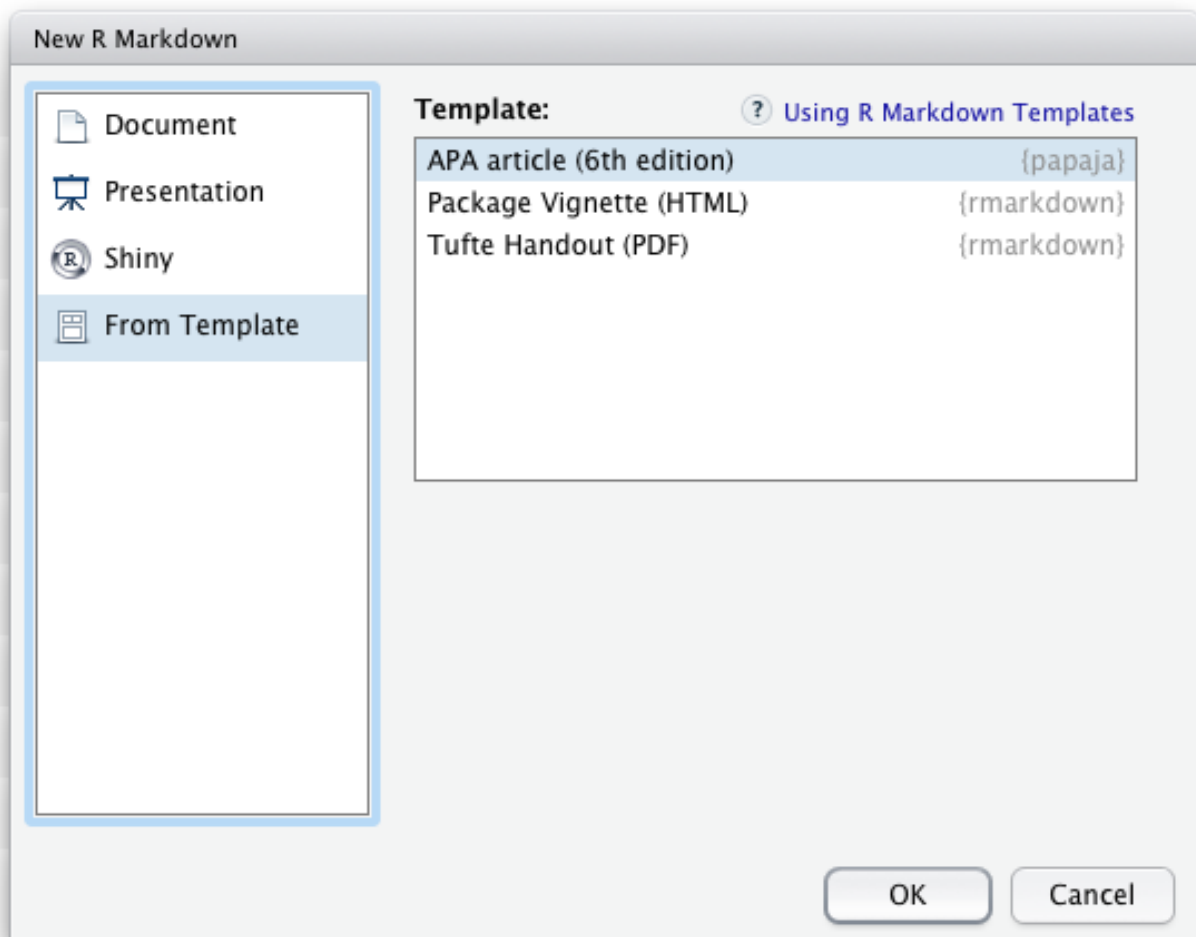


Figure 1. papaja's APA6 template is available through the RStudio menus.

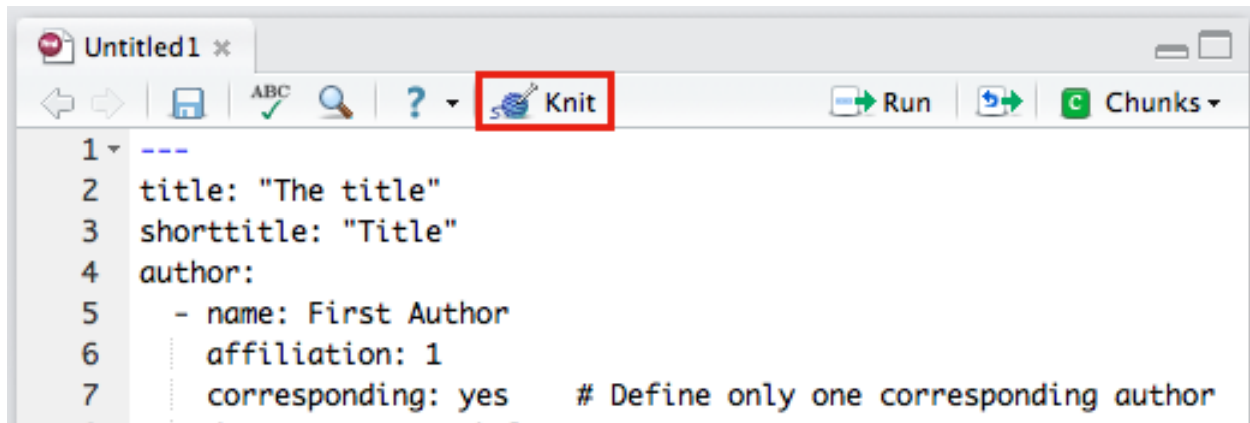


Figure 2. The *Knit* button in the RStudio.

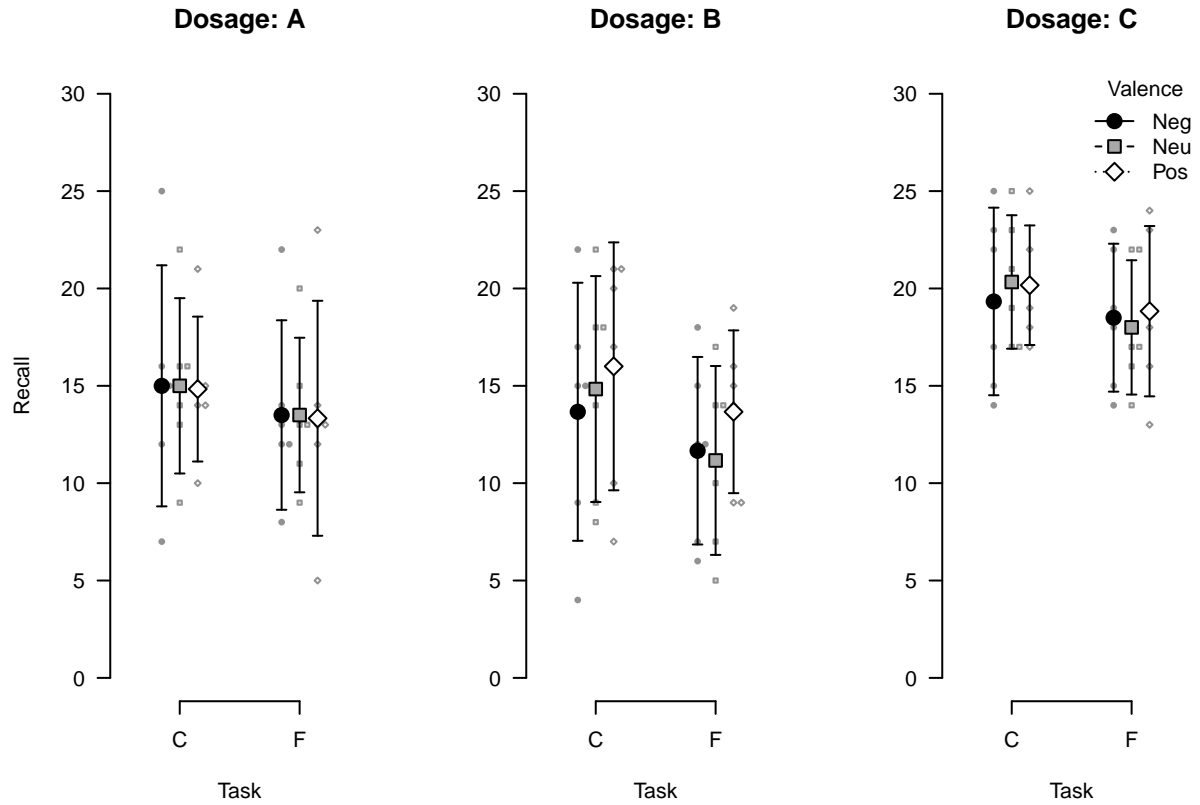


Figure 3. Bee plot of the example data set. Small points represent individual observations, large points represent means, and error bars represent 95% confidence intervals.