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

2   Frederik Aust[1]

3   [1] University of Cologne

8   Abstract

9   This manuscript demonstrates how to use R Markdown and papaja to create an APA

10 conform manuscript. papaja builds on R Markdown, which uses pandoc to turn Markdown

11 into PDF or Word documents. The conversion to Word documents currently supports only a

12 limited set of features.

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

14   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 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 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 menues, see **??**. When you click RStudio's *Knit* button (see Figure 2), `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.

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.

```r
summary(mixed_data)
```

```
##       Obs            Subject   Gender Dosage Task    Valence
##  Min.   :  1.00   A      : 6   F:54   A:36   C:54   Neg:36
##  1st Qu.: 27.75   B      : 6   M:54   B:36   F:54   Neu:36
##  Median : 54.50   C      : 6          C:36          Pos:36
##  Mean   : 54.50   D      : 6
##  3rd Qu.: 81.25   E      : 6
##  Max.   :108.00   F      : 6
##                   (Other):72
##      Recall
##  Min.   : 4.00
##  1st Qu.:13.00
##  Median :15.00
##  Mean   :15.63
##  3rd Qu.:19.00
##  Max.   :25.00
##
```

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()`.

```r
descriptives <- mixed_data %>% group_by(Dosage) %>%

  summarize(

    Mean = printnum( mean(Recall) )

    , Median = printnum( median(Recall) )

    , SD = printnum( sd(Recall) )

    , Min = printnum( min(Recall) )

    , Max = printnum( max(Recall) )

  )


apa_table(

  descriptives

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

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

  , escape = TRUE

)
```

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. In papaja, 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 defered to the final pages of the manuscript when creating a PDF. 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. Again, this is not the case in Word documents due to limited pandoc

---

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

₇₃ functionality. The bottom line is, Word documents will be less polished than PDF. The

₇₄ resulting documents should suffice to enable collaboration with Wordy colleagues and

₇₅ prepare a journal submission with limited manual labor.

₇₆ **Embed plots.** As usual in R Markdown, you can embed R-generated plots into your

₇₇ 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
)
```

₇₈ Again, as required by the APA guidelines, figures are defered to the final pages of the

₇₉ document unless you set `figsintext` to `yes`.

₈₀ **Referencing figures and tables.** Currently, `rmarkdown` does not support proper

₈₁ cross-referencing with the document. For this document, I have simply interspersed my R

₈₂ markdown document with LaTeXto accomplish cross-references. A new version of pandoc

₈₃ (1.16), that is not *yet* shipped with RStudio, supports proper cross-referencing that will work

₈₄ with Word documents as well.

₈₅ **Report statistical analyses.** `apa_print()` will help you report the results of your

₈₆ statistical analyses. The function will format the contents of R objects and produce readily

₈₇ 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)
```

88   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.
```

89   Item valence ($F[1.62, 24.36] = 3.46$, $MSE = 2.62$, $p = .056$, $\eta_p^2 = .187$) and the

90   task affected recall performance, $F(1, 15) = 43.13$, $MSE = 2.23$, $p < .001$,

91        $\eta_p^2 = .742$; the dosage, however, had no effect on recall, $F(2, 15) = 2.97$,

92        $MSE = 117.17$, $p = .082$, $\eta_p^2 = .283$. There was no significant interaction.

93   What's even more fun, you can easily create a complete ANOVA table using by passing

94   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()."

)
```

⁹⁵ **Citations**

⁹⁶   No manuscript is complete without citation. In order for citations to work, you need to

⁹⁷ supply a .bib-file to the `bibliography` parameter in the YAML front matter. Once this is

⁹⁸ done, `[e.g., @james_1890; @bem_2011]` produces a regular citation within parenthes

⁹⁹ (Bem, 2011; e.g., James, 1890). To cite a source in text simply omit the brackets; for

¹⁰⁰ example, write `@james_1890` to cite James (1890). For other options see the overview of the

¹⁰¹ R Markdown citation syntax. If you use RStudio, I have created an easy-to-use add-in that

¹⁰² facilitates inserting citations. The relevant references will, of course, be added to the

¹⁰³ documents reference section automatically.

¹⁰⁴   I think it is important to credit the software we use. A lot of R packages are developed

¹⁰⁵ by academics free of charge. As citations are the currency of science, it's easy to compensate

¹⁰⁶ volunteers for their work by citing the R packages we use. I suspect that, among other

¹⁰⁷ things, this is rarely done because it is tedious work. That's why papaja makes citing R and

¹⁰⁸ its packages easy:

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

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

¹⁰⁹   `r_refs()` creates a BibTeX file containing citations for R and all currently loaded

¹¹⁰ packages. `cite_r()` takes these citations and turns them into readily reportable text.

¹¹¹ `my_citation` now contains the following text that you can use in your document: R (3.3.1,

¹¹² R Core Team, 2015) and the R-packages *afex* (0.16.1, Singmann, Bolker, Westfall, & Aust,

¹¹³ 2016), *dplyr* (0.5.0, Wickham & Francois, 2016), *estimability* (1.1.1, Lenth, 2015), *lme4*

¹¹⁴ (1.1.12, Bates, Mächler, Bolker, & Walker, 2015), *lsmeans* (2.23, Lenth, 2016), *Matrix* (1.2.6,

¹¹⁵ Bates & Maechler, 2016), *papaja* (0.1.0.9076, Aust & Barth, 2015), and *reshape2* (1.4.1,

¹¹⁶ Wickham, 2007)

¹¹⁷   The citation style is automatically set to APA style. If you need to use a different

¹¹⁸ citation style, you can set in the YAML front matter by adding a `csl` parameter. See the R

119 [Markdown documentation](#) and [Citation Style Language](#) for further details.

Math

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

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

Document options

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

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

Last words

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

References

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

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

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

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

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/

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., & Francois, R. (2016). *dplyr: A Grammar of Data Manipulation.* Retrieved
    from https://CRAN.R-project.org/package=dplyr

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 analyis 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
Dosage × Task	1.83	2	15	2.23	.195	.196
Valence	3.46	1.62	24.36	2.62	.056	.187
Dosage × Valence	2.38	3.25	24.36	2.62	.090	.241
Task × Valence	1.50	1.35	20.2	2.67	.242	.091
Dosage × Task × Valence	0.39	2.69	20.2	2.67	.743	.049

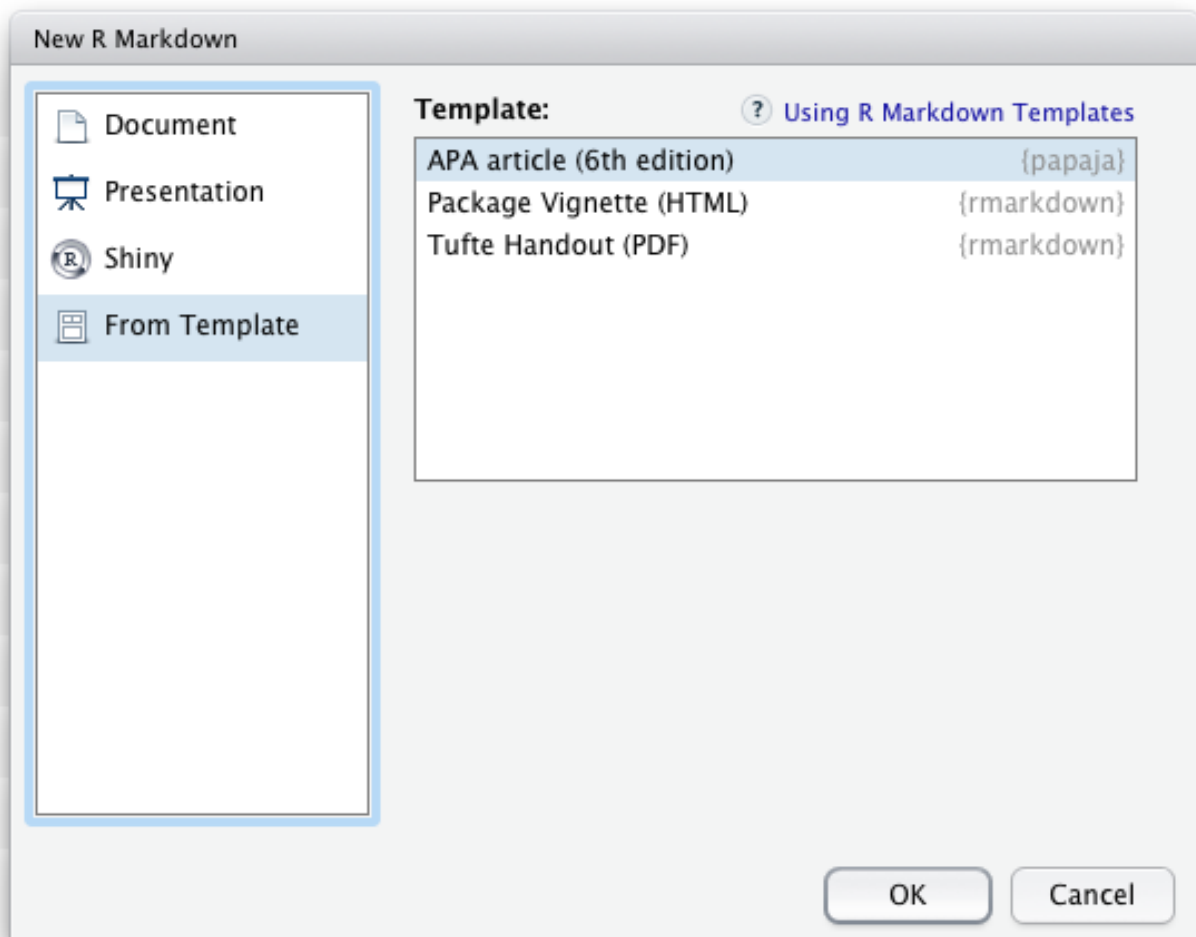*Note.* This is a table created using apa_print() and apa_table().

Figure 1. papaja's APA6 template is available through the RStudio menues.
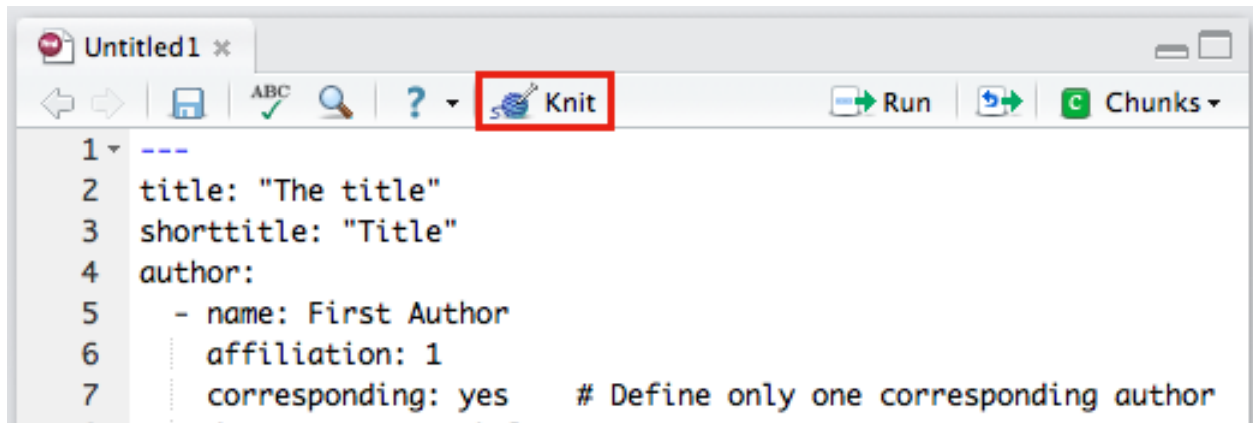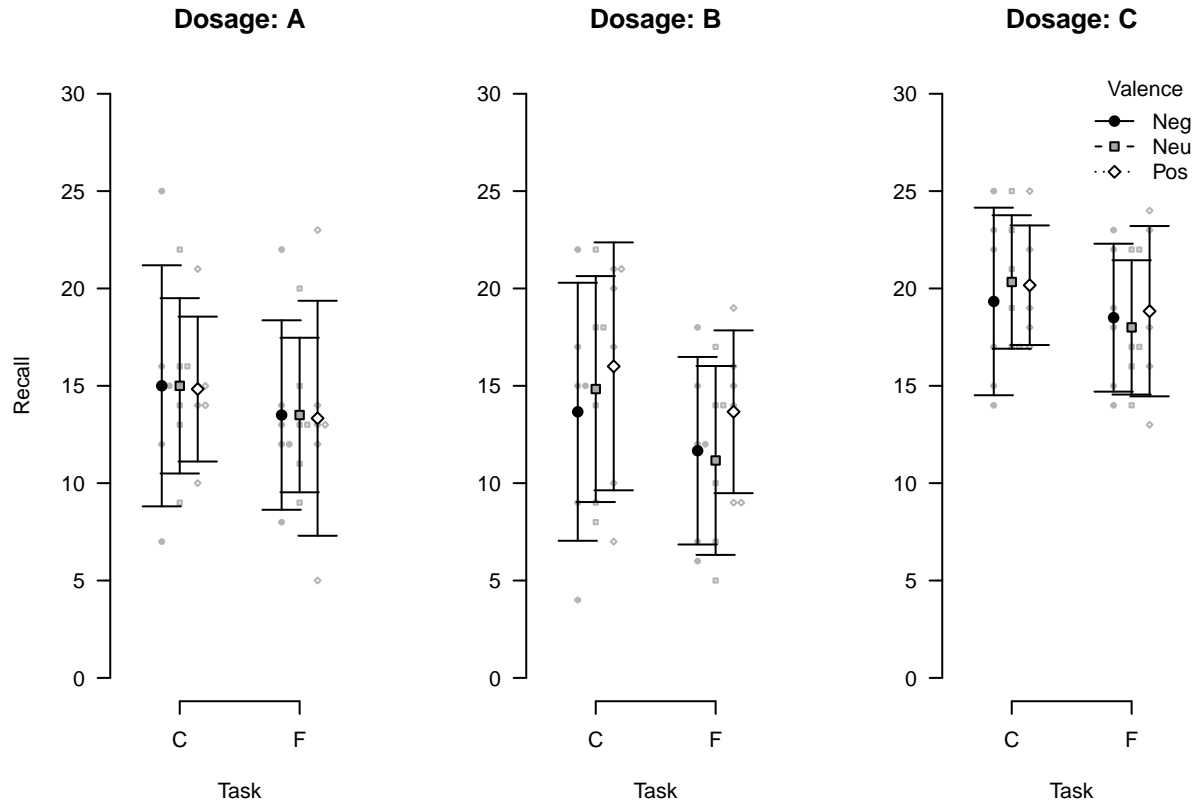
Figure 2. The *Knit* button in the RStudio.

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.