Introduction to papaja

R Markdown for APA-style manuscripts

Johannes Breuer & Frederik Aust

KU Leuven, 27.-28.04.2022

Scope of the package

Preparing APA Journal Articles

1. Designed for APA-style manuscripts
2. Templates for PDF and DOCX documents
3. Functions to facilitate reporting of results, e.g.
    - `apa_print()`, `apa_num()`
    - `apa_table()`
    - `apa_factorial_plot()`, `theme_apa()`

Getting started

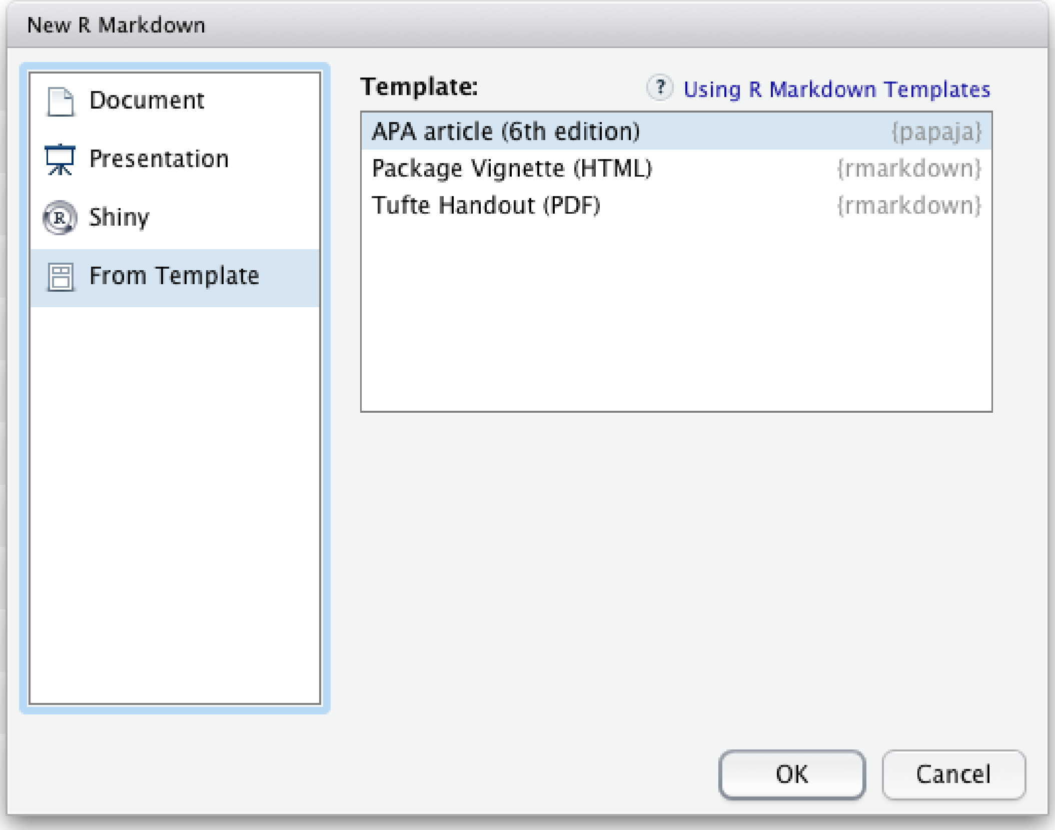
Install papaja from GitHub

```
# Install the stable development verions from GitHub
remotes::install_github("crsh/papaja", build_vignettes = T

# Install the latest development snapshot from GitHub
remotes::install_github("crsh/papaja@devel", build_vignett
```

Ensure the required software is also installed

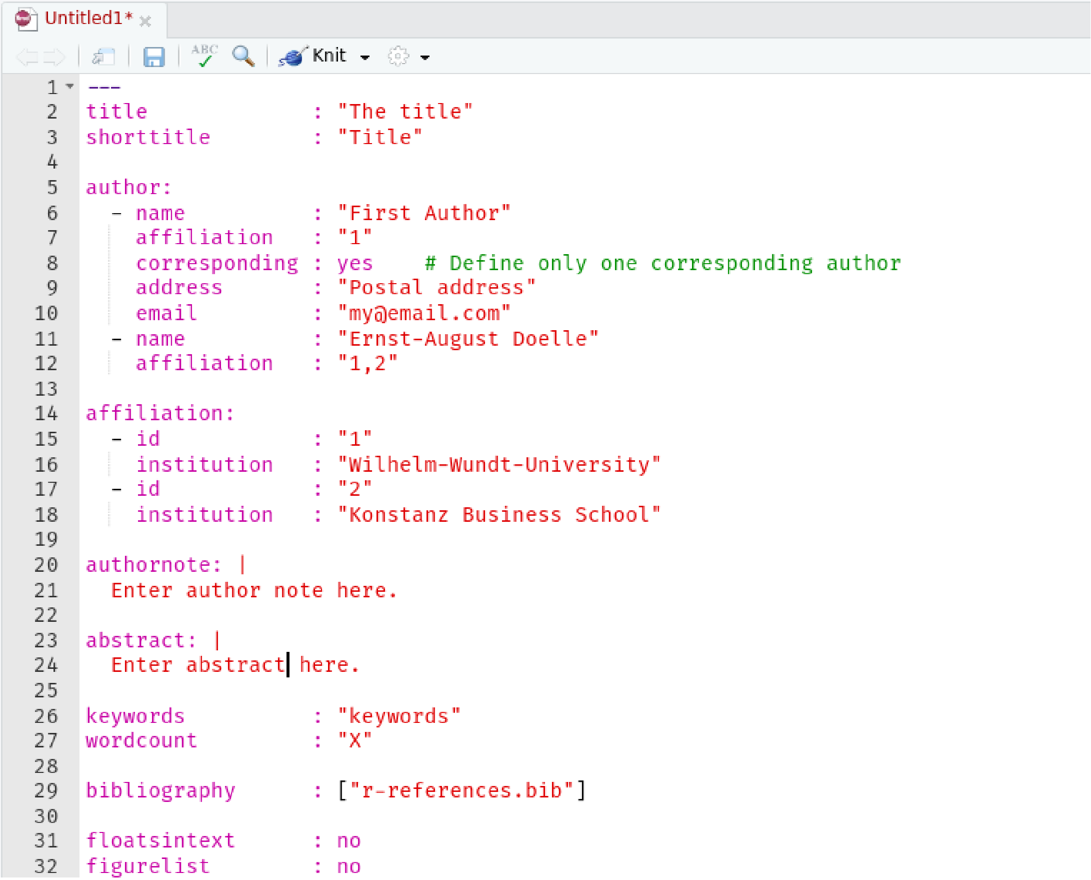
Document templates



Template:

APA article (6th edition)	{papaja}
Package Vignette (HTML)	{rmarkdown}
Tufte Handout (PDF)	{rmarkdown}

Document templates



```
---
title             : "The title"
shorttitle        : "Title"

author:
  - name          : "First Author"
    affiliation   : "1"
    corresponding : yes    # Define only one corresponding author
    address       : "Postal address"
    email         : "my@email.com"
  - name          : "Ernst-August Doelle"
    affiliation   : "1,2"

affiliation:
  - id            : "1"
    institution   : "Wilhelm-Wundt-University"
  - id            : "2"
    institution   : "Konstanz Business School"

authornote: |
  Enter author note here.

abstract: |
  Enter abstract here.

keywords          : "keywords"
wordcount         : "X"

bibliography      : ["r-references.bib"]

floatsintext      : no
figurelist        : no
```

YAML field	Metadata
`author`	list of author information (e.g., `name` and `affiliation`; start each new author with `-`)
`affiliation`	list of institutional information (`id` and `institution`)
`authornote`	automatically contains corresponding author line
`keywords`	article keywords
`wordcount`	article word count
`note`	text to add above author note on the title page (e.g. "Preprint submitted for publication")

Rendering options

YAML field	Effect
`bibliography`	List of bibliography files
`draft`*	Add "DRAFT" watermark across all pages
`figurelist`* `tablelist`* `footnotelist`*	Create lists of figure captions, table captions, or footnotes

* Only available for PDF documents

Rendering options

YAML field	Effect
`floatsintext`*	Place figures and tables in the text rather than at the end
`linenumbers`*	Add line numbers in margins
`mask`	Omit identifying information from title page
`classoption`*	control the style of the document (e.g., `man` or `doc`, see `apa6` LaTeX class options)

* Only available for PDF documents

Citations

`citeproc` is a `pandoc` extension that formats references

- works well for both PDF and DOCX documents
- requires a separate reference file (e.g., CSL-JSON, Bib(La)TeX, EndNote)

Add the following to the YAML front matter:

```
bibliography: references.bib
```

In papaja, the default citation style is APA, 6th edition.

Citations

The reference handle is used to select citations

Citation type	Syntax	Rendered citation
Citation within parentheses	`[@james_1890]`	(James, 1890)
Multiple citations	`[@james_1890; @bem_2011]`	(Bem, 2011; James, 1890)
In-text citations	`@james_1890`	James (1890)
Year only	`[-@bem_2011]`	(2011)

Citations

You can add pre- and post-fixes to individual citations

- `[e.g., @bem_2011]` yields "(e.g., Bem, 2011)"
- `[see @bem_2011 for a surprising result]` yields "(see Bem, 2011, for a surpising result)"

Citations

Insert citations via

1. (Copy-and-paste)
2. Visual editor
3. RStudio addin `citr`

Both connect directly to Zotero if the Zotero extension Better Bib(La)TeX is installed and Zotero is running

A quick demonstration!

Citing R packages

Reward volunteers who develop R packages for free! ;)

```r
citation("papaja")
```

```
##
##   Aust, F. & Barth, M. (2020). papaja: Prepare reproducible A
##   articles with R Markdown. R package version 0.1.0.9999. Ret
##   from https://github.com/crsh/papaja
##
## Ein BibTeX-Eintrag für LaTeX-Benutzer ist
##
##   @Manual{,
##     title = {{papaja}: {Prepare} reproducible {APA} journal a
##     author = {Frederik Aust and Marius Barth},
##     year = {2020},
##     note = {R package version 0.1.0.9999},
##     url = {https://github.com/crsh/papaja},
```

Citing R packages

- `r_refs()` creates a BibTeX file with references for all loaded packages (place at the end of document)
- `cite_r()` automates citing R and R packages

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

```
We used `r r_citations` for all analyses.
```

> We used R (Version 4.1.3; R Core Team, 2022) and the R-package papaja (Version 0.1.0.9999; Aust & Barth, 2022) for all analyses.

Citing R packages

Cite only selected packages or place package citations in a footnote

```
r_citations <- cite_r(
  file = "r-references.bib"
  , pkgs = c("afex", "emmeans", "papaja"), withhold = FALS
  , footnote = TRUE
)
r_citations
```

```
## Warning in cite_r(file = "../../exercises/3_papaja_example_ma
## r-references.bib", : File ../../exercises/3_papaja_example_ma
## references.bib not found. Cannot cite R-packages. If knitting
## solve the problem, please check file path.

## $r
```

Citing R packages

Cite only selected packages or place package citations in a footnote

```
r_citations <- cite_r(
  file = "r-references.bib"
  , pkgs = c("afex", "emmeans", "papaja"), withhold = FALS
  , footnote = TRUE
)
```

```
We used `r r_citations$r` for all analyses.

`r r_citations$pkgs`
```

Report statistical analyses

Numerical values reported inline will be rounded

```
Participants mean age was `r age_mean` years
(*SD* = `r age_sd`).
```

> Participants mean age was 32.35 years
> (*SD* = 6.23).

Report statistical analyses

Typeset numerical values for greater control

```r
apa_num(c(143234.34557, Inf))
```

```
## [1] "143,234.35" "$\\infty$"
```

```r
apa_num(42L, numerals = FALSE, capitalize = TRUE)
```

```
## [1] "Forty-two"
```

```r
apa_num(1.7e10, format = "e")
```

```
## [1] "$1.70 \\times 10^{10}$"
```

Report statistical analyses

Special-purpose wrappers for convenience

```
apa_p(c(1, 0.0008, 0))
```

```
## [1] "> .999" ".001"   "< .001"
```

```
apa_df(c(1, 15.93))
```

```
## [1] "1"     "15.93"
```

```
apa_confint(c(0.01, 0.8), conf.int = 0.95)
```

```
## [1] "95\\% CI [0.01, 0.80]"
```

Report statistical analyses

```r
# Data from Field, Miles, & Field (2012)
load("cosmetic_surgery.Rdata")

(cor_res <- with(cosmetic_surgery, cor.test(Post_QoL, BDI)
```

```
##
##      Pearson's product-moment correlation
##
## data:  Post_QoL and BDI
## t = 7.7581, df = 274, p-value = 1.71e-13
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3224754 0.5165716
## sample estimates:
##       cor
## 0.4243863
```

Report statistical analyses

`apa_print()` facilitates reporting of results. It returns a `list` with the following elements:

- `estimate`: Effect size estimate
- `statistic`: Hypothesis test statistic
- `full_result`: Combined estimates and statistics
- `table`

Report statistical analyses

```r
cor_apa_res <- apa_print(cor_res)
cor_apa_res[c("estimate", "statistic", "full_result")]
```

```
## $estimate
## [1] "$r = .42$, 95\\% CI $[.32, .52]$"
##
## $statistic
## [1] "$t(274) = 7.76$, $p < .001$"
##
## $full_result
## [1] "$r = .42$, 95\\% CI $[.32, .52]$, $t(274) = 7.76$, $p <
```

Report statistical analyses

```
cor_apa_res$table
```

```
## A data.frame with 5 labelled columns:
##
##   estimate   conf.int statistic  df p.value
## 1      .42 [.32, .52]      7.76 274  < .001
##
## estimate : $r$
## conf.int : 95\\% CI
## statistic: $t$
## df       : $\\mathit{df}$
## p.value  : $p$
```

Report statistical analyses

`apa_print()` adjusts behavior according to input

```
class(cor_res) # Result from cor.test()
```

```
## [1] "htest"
```

```
?apa_print.htest
```

`methods(apa_print)` provides a list of supported classes

Report statistical analyses

A-B	D-L	L-S	S-Z
afex_aov	default	lsmobj	summary.glht
anova	emmGrid	manova	summary.glm
anova.lme	glht	merMod	summary.lm
Anova.mlm	glm	mixed	summary.manova
aov	htest	papaja_wsci	summary.ref.grid
aovlist	list	summary.Anova.mlm	summary_emm
BFBayesFactor	lm	summary.aov	
BFBayesFactorTop	lme	summary.aovlist	

See `vignette("extending_apa_print", package = "papaja")`.

Report statistical analyses

```r
lm_res <- lm(Post_QoL ~ Base_QoL + BDI, data = cosmetic_su

lm_res_apa <- apa_print(lm_res, observed = TRUE)
str(lm_res_apa, max.level = 2)
```

```
## List of 4
##  $ estimate   :List of 4
##   ..$ Intercept: chr "$b = 18.50$, 95\\% CI $[13.10, 23.91]$"
##   ..$ Base_QoL : chr "$b = 0.59$, 95\\% CI $[0.50, 0.67]$"
##   ..$ BDI      : chr "$b = 0.17$, 95\\% CI $[0.11, 0.22]$"
##   ..$ modelfit :List of 4
##  $ statistic  :List of 4
##   ..$ Intercept: chr "$t(273) = 6.74$, $p < .001$"
##   ..$ Base_QoL : chr "$t(273) = 13.23$, $p < .001$"
##   ..$ BDI      : chr "$t(273) = 6.08$, $p < .001$"
##   ..$ modelfit :List of 1
##  $ full_result:List of 4
```

Report statistical analyses

```r
lm_res_apa$estimate$Intercept
```

```
## [1] "$b = 18.50$, 95\\% CI $[13.10, 23.91]$"
```

> $b = 18.50$, 95% CI $[13.10, 23.91]$

```r
lm_res_apa$full_result$modelfit$r2
```

```
## [1] "$R^2 = .50$, 90\\% CI $[0.42, 0.57]$, $F(2, 273) = 136.7
```

> $R^2 = .50$, 90% CI $[0.42, 0.57]$, $F(2, 273) = 136.78$, $p < .001$

Rendering Tables

Tables returned by `apa_print()` have variable labels

```
lm_res_apa$table
```

```
## A data.frame with 6 labelled columns:
## 
##        term estimate       conf.int statistic  df p.value
## 1 Intercept    18.50 [13.10, 23.91]      6.74 273  < .001
## 2  Base QoL     0.59   [0.50, 0.67]     13.23 273  < .001
## 3       BDI     0.17   [0.11, 0.22]      6.08 273  < .001
## 
## term     : Predictor
## estimate : $b$
## conf.int : 95\\% CI
## statistic: $t$
## df       : $\\mathit{df}$
## p.value  : $p$
```

Rendering Tables

Tables returned by `apa_print()` have variable labels

```
variable_labels(lm_res_apa$table)
```

```
## $term
## [1] "Predictor"
##
## $estimate
## [1] "$b$"
##
## $conf.int
## [1] "95\\% CI"
##
## $statistic
## [1] "$t$"
##
## $df
```

Rendering Tables

`apa_table()` renders tables with variable labels

```r
apa_table(
  lm_res_apa$table
  , caption = "Cosmetic surgery regression table."
)
```

Table 1. *Cosmetic surgery regression table.*

Predictor	b	95%CI	t	df	p
Intercept	18.50	[13.10, 23.91]	6.74	273	< .001
Base QoL	0.59	[0.50, 0.67]	13.23	273	< .001
BDI	0.17	[0.11, 0.22]	6.08	273	< .001

Rendering Tables

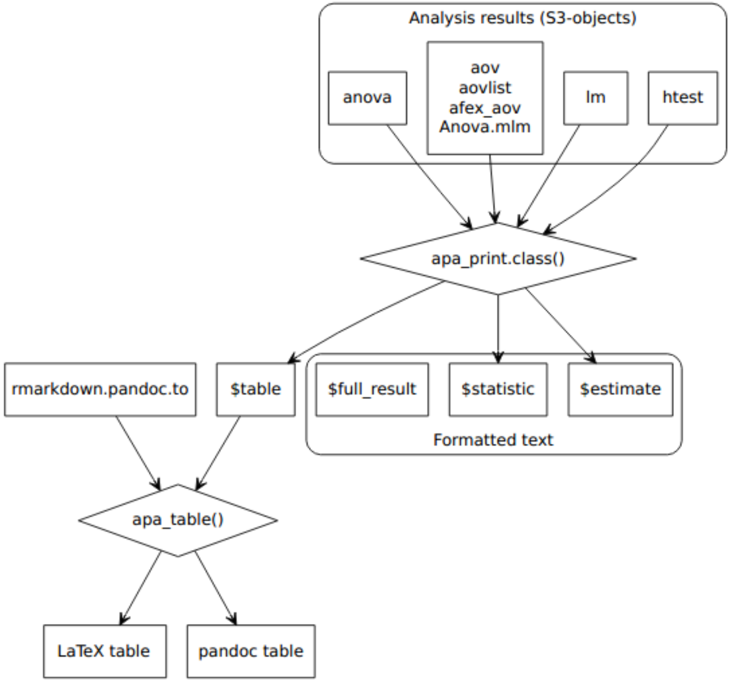
`apa_table()`

- was designed with table examples from the APA manual in mind
- is much more powerful in PDF documents

A quick demonstration!

GitHub folder

Report statistical analyses

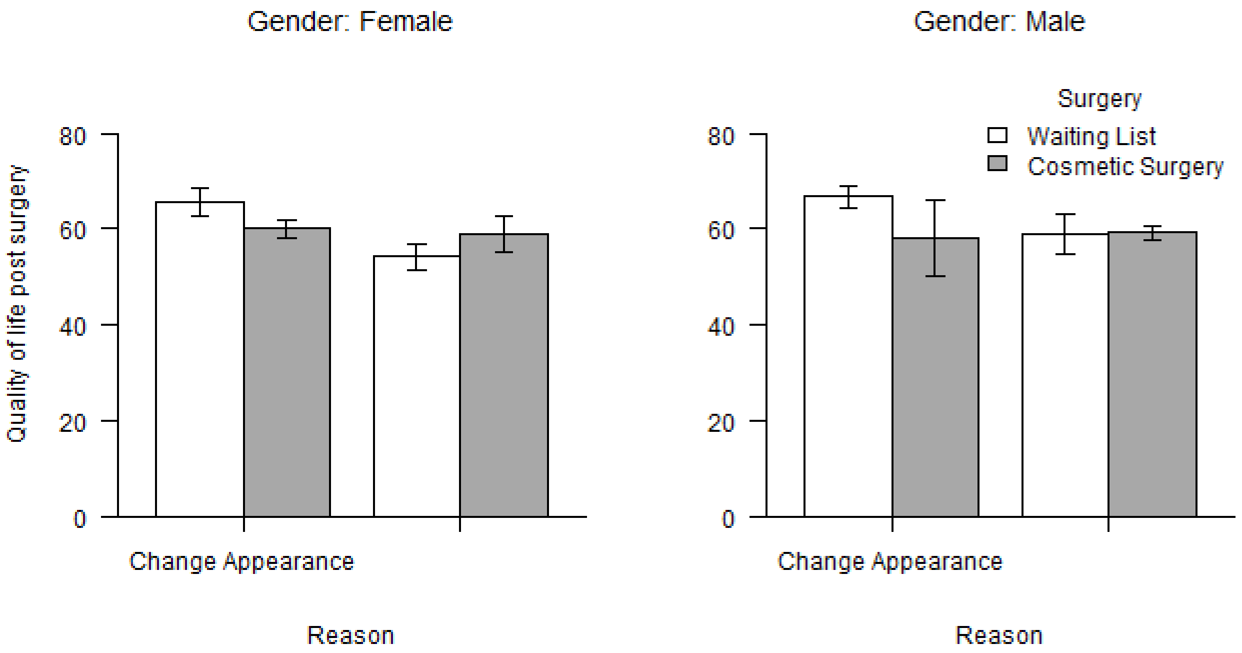


Creating figures

`apa_barplot()`, `apa_beeplot()`, `apa_lineplot()` can be used to visualize factorial designs

```
apa_barplot(
  id = "ID"
  , dv = "Post_QoL"
  , factors = c("Reason", "Surgery", "Gender")
  , data = cosmetic_surgery
  , ylab = "Quality of life post surgery"
  , las = 1
)
```

Creating figures

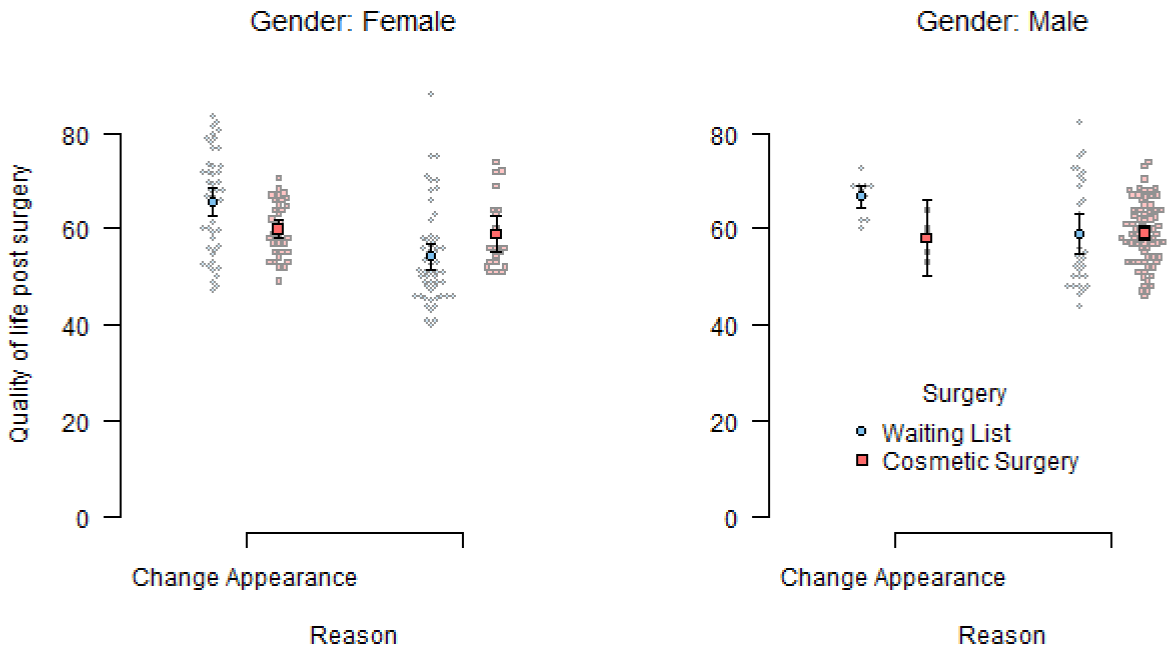


Creating figures

`apa_barplot()`, `apa_beeplot()`, `apa_lineplot()` can be used to visualize factorial designs

```r
# This time with bees
apa_beeplot(
  id = "ID"
  , dv = "Post_QoL"
  , factors = c("Reason", "Surgery", "Gender")
  , data = cosmetic_surgery
  , ylab = "Quality of life post surgery"
  , las = 1
  , args_legend = list(x = 0.25, y = 30)
  , args_points = list(bg = c("skyblue2", "indianred1"))
  , args_error_bars = list(length = 0.03)
)
```

Creating figures



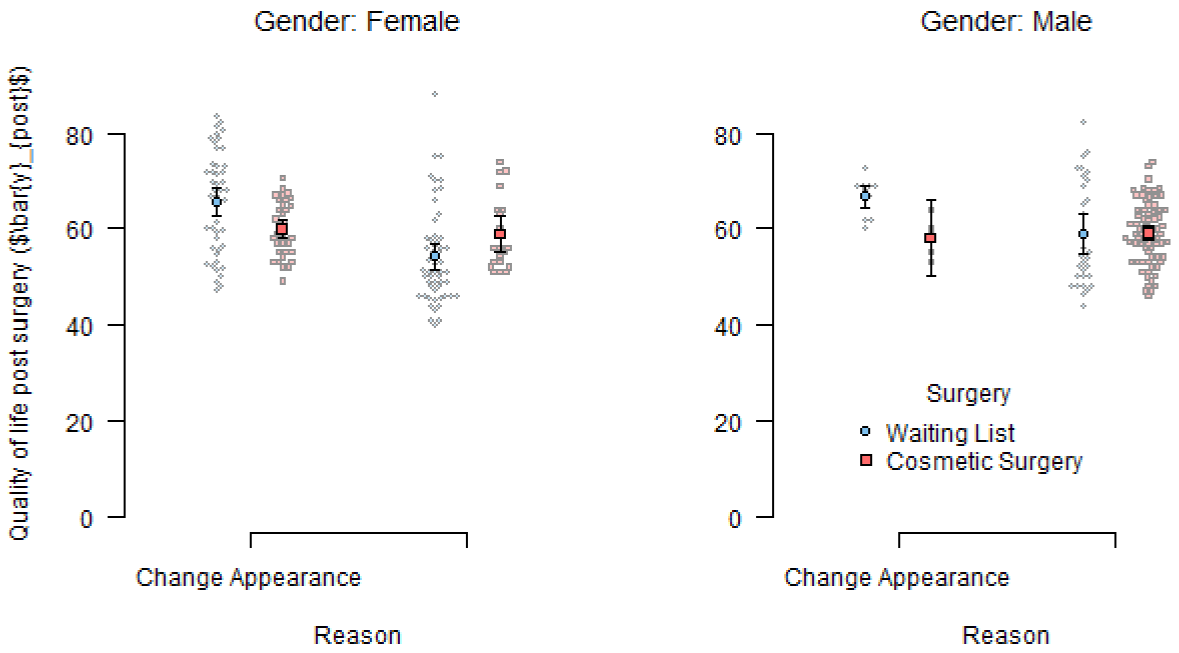
Creating figures

All plot functions render variable labels, with some LaTeX math support (see `?latex2exp::TeX`)

```r
variable_labels(cosmetic_surgery) <- c(
  Post_QoL = "Quality of life post surgery ($\\bar{y}_{pos
)

apa_beeplot(
  id = "ID"
  , dv = "Post_QoL"
  , factors = c("Reason", "Surgery", "Gender")
  , data = cosmetic_surgery
  , las = 1
  , args_legend = list(x = 0.25, y = 30)
  , args_points = list(bg = c("skyblue2", "indianred1"))
  , args_error_bars = list(length = 0.03)
)
```

Creating figures



Creating figures

For `ggplot2` users, `papaja` provides `theme_apa()`

```r
ggplot(
    cosmetic_surgery
    , aes(x = Base_QoL, y = Post_QoL, color = Reason)
  ) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(
    x = "Baseline quality of life"
    , y = "Quality of life post surgery"
  ) +
  scale_color_brewer(palette = "Set1") +
  theme_apa(box = TRUE) +
  theme(legend.position = c(0.2, 0.8))
```

Creating figures

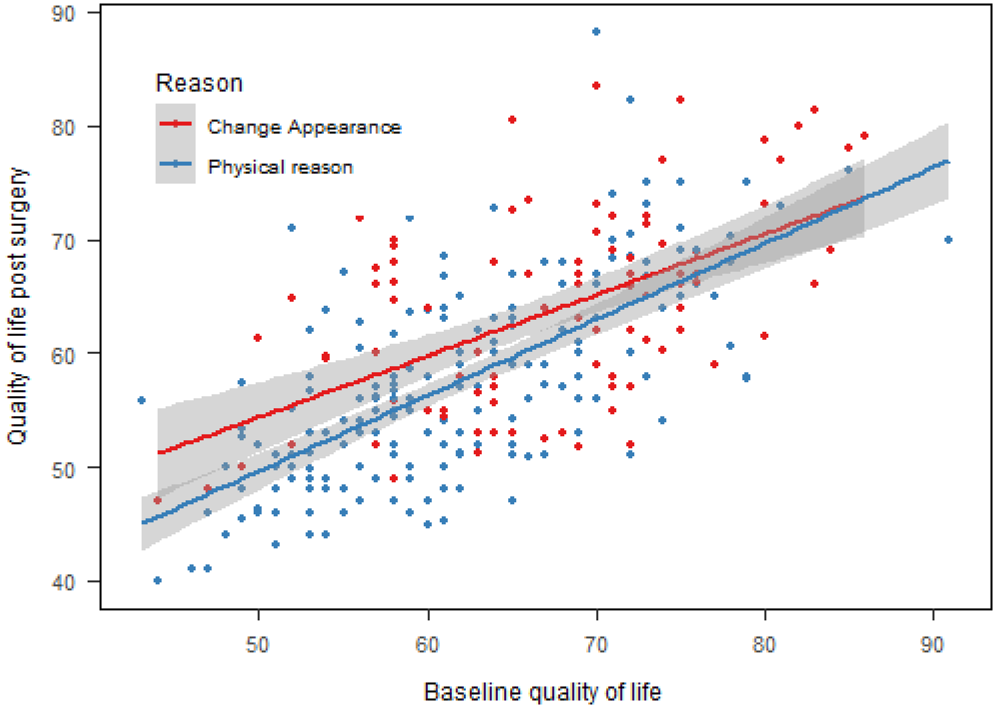


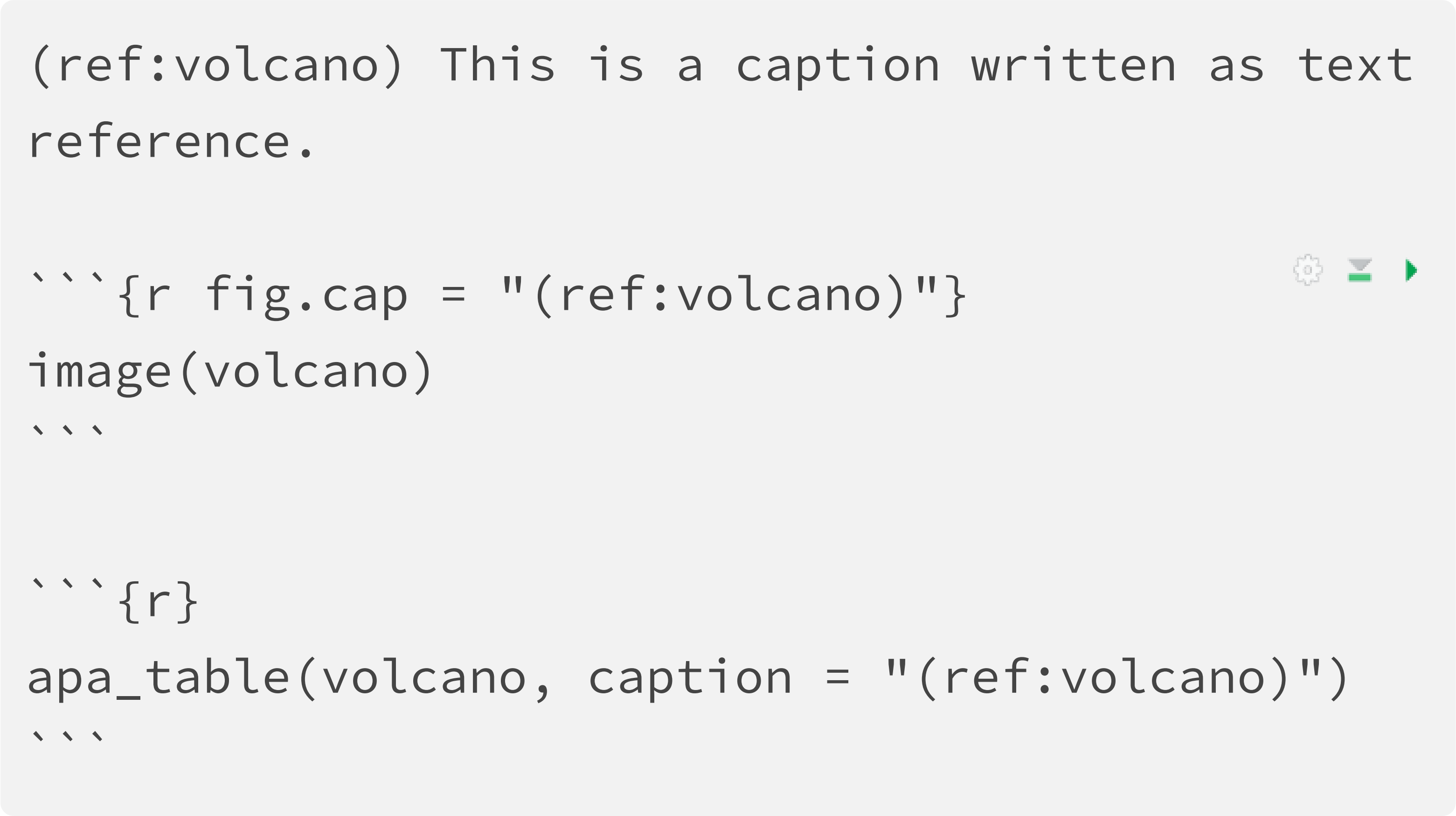
Figure and table captions

Add a figure caption with the chunk option `fig.cap`

- Caption is reused for every figure in a chunk
  - Only one figure per chunk
  - Combine plots into multi-panel figures (e.g., `layout()`, `cowplot::plot_grid()`, or the **patchwork** package)

Figure and table captions

It's recommended to use "text-references"



~~~
(ref:volcano) This is a caption written as text reference.

```{r fig.cap = "(ref:volcano)"}
image(volcano)
```


```{r}
apa_table(volcano, caption = "(ref:volcano)")
```
~~~

Cross-referencing

To cross-reference figures and tables use `\@ref(fig:chunk-label)` or `\@ref(tab:chunk-label)`

- Chunk labels must not contain `_`
- Precede by non-breaking spaces, e.g. `Figure\ \@ref(fig:chunk-label)`

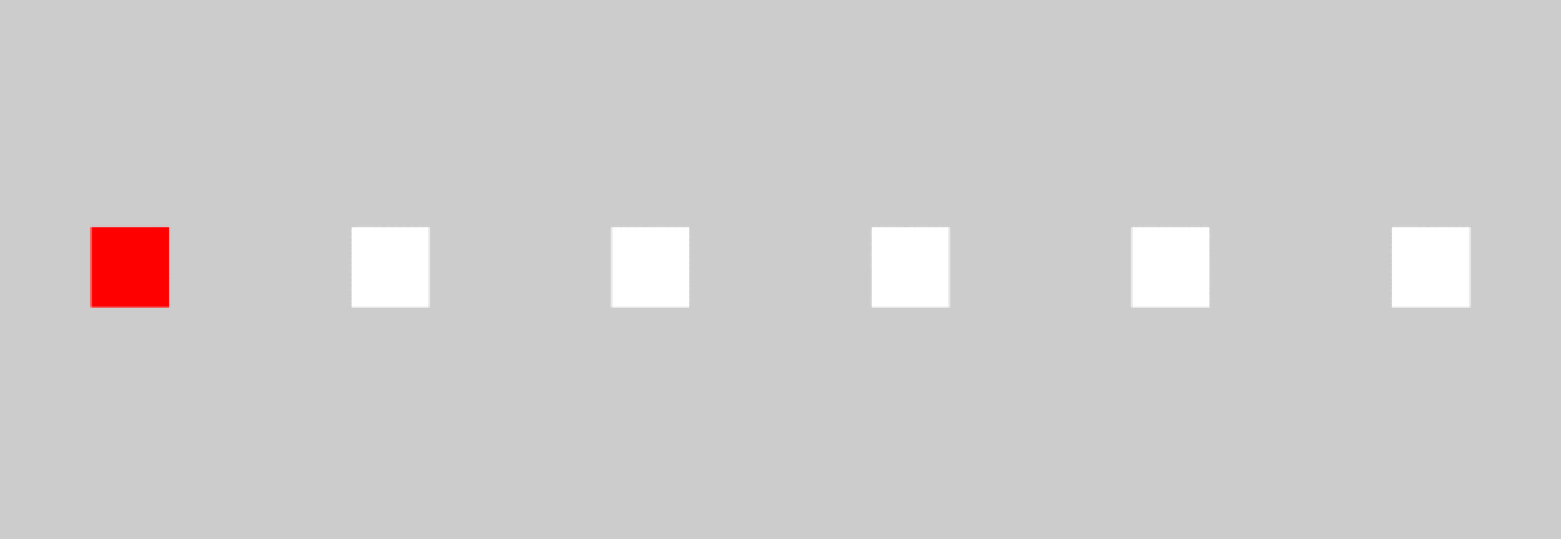
If you are stuck

1. Try our work-in-progress manual
2. Ask on StackOverflow using the `papaja`-tag
3. Open a GitHub issue

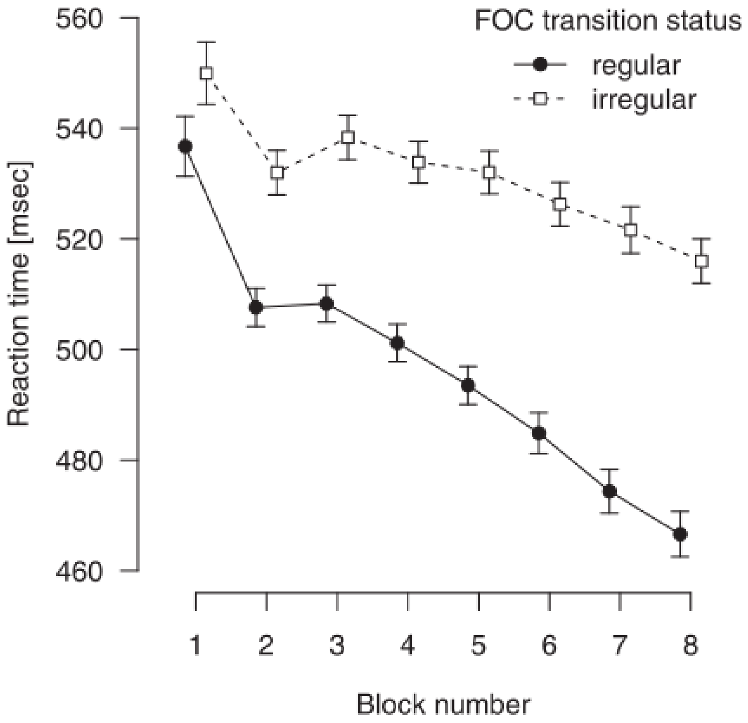
Example manuscript

GitHub folder

Serial response time task



Serial response time task



Process Dissociation

Distinguish implicit and explicit learning

- Implicit: automatic, not controllable
- Explicit: controllable, may be used intentionally

Inclusion

"Generate a sequence that is as *similar* as possible"

Exclusion

"Generate a sequence that is as *dissimilar* as possible"

The present study

Do variants of the PD task differ with respect to "baseline" performance?

- Generation task (**free** vs. **cued** generation tasks)
- Types of "random" material (**permuted** vs. **random** material)
- Performance without prior task exposure (**no-learning** group)

(Stahl, Barth, & Haider, 2015)

The present study

The following files are provided:

- The paper `manuscript.pdf` and `manuscript.docx`
- Bibliography file `references.bib`
- Data in the folder `data`
- Analysis script `analyses.R`

https://tinyurl.com/rrpp-papaja

Exercise time

Exercise

Solutions