RMarkdown based APA Manuscript with ‘papaja’

Jenny Rieck[1] & Derek Beaton[1]

[1] Rotman Research Institute

Author Note

We can include author notes for, e.g., attribution of data or expanding on roles/locations.

Abstract

This is a short example of a completely reproducible manuscript made entirely in RStudio with RMarkdown and various R scripts, functions, and packages.

Keywords: reproducible, manuscript, multivariate, APA, papaja, knitr, R, Rmarkdown

RMarkdown based APA Manuscript with 'papaja'

Introduction

The aim of this RMarkdown example is to show how to write a reproducible manuscript, which includes numerous bells-and-whistles—with contributions from `R`, `Python`, RMarkdown, and `LaTeX`. This "manuscript" will include only the best (nicest looking) parts from `1_a_Simple_RMarkdown_PDF.Rmd`

In order to help make this manuscript look nicer, we are changing some of the YAML ("yet another markup language") header options, so that we remove line numbers, and allow for tables & figures to appear in text, as opposed to the end.

Methods

We first make our data with R. This will be followed by a paragraph break because this text precedes a chunk.

Then call off to python for the `.describe()` method. This, too, will be followed by a paragraph break because this text precedes a chunk.

And then pass the `desc` object back to `R` and use the `kable()` and `kableExtra` packages to make a nice table of summary statistics for the measures of interest. In this particular part, we also show the code chunk. Furthermore, some particular packages—such as `papaja`—and certain advanced features from `LaTeX` require the use of `results='asis'` in chunk header.

```
apa_table(py$desc, caption = "A descriptive statistics table.",
  note = "This formatted through LaTeX via papaja::apa_table() from a python object fr
```

Table 1

A descriptive statistics table.

	AGE	MOCA	CDRSB	WholeBrain	Hippocampus	MidTemp
count	665.00	665.00	665.00	665.00	665.00	665.00
mean	71.92	23.89	1.20	1,057,025.55	7,149.61	20,301.93
std	6.87	3.28	1.34	103,672.74	1,086.04	2,675.57
min	55.00	16.00	0.00	817,421.23	3,731.00	12,213.00
25%	67.20	22.00	0.00	984,409.91	6,510.00	18,535.00
50%	71.90	24.00	1.00	1,051,621.33	7,223.00	20,186.00
75%	76.60	26.00	2.00	1,120,569.50	7,834.00	22,088.00
max	89.60	30.00	5.50	1,486,035.64	10,602.00	32,189.00

Note. This formatted through LaTeX via papaja::apa_table() from a python object from the .describe() method, loaded from data in R and written through RMarkdown.

Procedure

Next we are going to use more features from `LaTeX`, including various additional `LaTeX` packages that we define in the YAML header. We can use a number of `LaTeX` features like inline calls, numbered equations, and, for eaxmple, algorithms. We will use each of those features to describe the covSTATIS method (see also our covSTATIS project repository). Our description of covSTATIS is extremely truncated here and is only meant to illustrate features of writing a manuscript in RMarkdown.

CovSTATIS is a multi-table principal components analysis, specifically designed to integrate and analyze multiple correlation or covariance matrices. Each correlation matrix—$\mathbf{R}_{[k]}$—is double-centered by way of a centering matrix as $\mathbf{\Xi} = \mathbf{I} - \mathbf{1}(I^{-1})\mathbf{1}^T$ as

$$\mathbf{S}_{[k]} = \frac{1}{2}\mathbf{\Xi}\mathbf{R}_{[k]}\mathbf{\Xi}. \tag{1}$$

After we perform the double-centering in Eq. @ref(eq:double_center), we then compute α weights of each matrix where first we vectorize each $\mathbf{S}_{[k]}$ and storing those each column vector in a new matrix as $\mathbf{Z} = [\text{vec}\{\mathbf{S}_{[1]}, \dots, \mathbf{S}_{[\text{k}]}, \dots, \mathbf{S}_{[\text{K}]}\}]$ and then decompose $\mathbf{Z}$ with the singular value decomposition (SVD):

$$\mathbf{Z} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^T. \tag{2}$$

The *alpha* weights are $\boldsymbol{\alpha} = \mathbf{v}_1 \times (\mathbf{v}_1^T\mathbf{1})^{-1}$. We then compute the compromise cross-product matrix as $\mathbf{S}_{[+]} = \sum_{i=1}^{K} \alpha_k \mathbf{S}_{[k]}$, and finally decompose $\mathbf{S}_{[+]}$ with the eigenvalue decomposition (EVD) as

$$\mathbf{S}_{[+]} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T. \tag{3}$$

We can also outline these steps algorithmically as

Data analysis

The `papaja` package includes the ability to generate a bibliography directly from all the loaded packages (via `cite_r()`). This document includes R (Version 3.5.1; R Core Team, 2018) and the R-packages *covstatis* (Version 0.1.0.0; it, n.d.), *dplyr* (Version 0.7.6; Wickham et al., 2018), *ExPosition* (Version 2.8.23; Beaton et al., 2014a), *factoextra* (Version 1.0.5; Kassambara & Mundt, 2017), *forcats* (Version 0.3.0; Wickham, 2018a), *ggplot2* (Version 3.0.0; Wickham, 2016), *gridExtra* (Version 2.3; Auguie, 2017), *GSVD* (Version 0.2.0; Beaton, n.d.), *here* (Version 0.1; Müller, 2017), *kableExtra* (Version 0.9.0; Zhu, 2018), *knitr* (Version 1.22.8; Xie, 2015), *ours* (Version 0.0.0.9000; Sunderland & Beaton, n.d.), *papaja* (Version

Result: Decomposition of the compromise matrix—$\mathbf{S}_{[+]}$—in covSTATIS

Input : $[\mathbf{R}_{[1]} ... \mathbf{R}_{[k]} ... \mathbf{R}_{[K]}]$

Output: $\mathbf{Q}$, $\mathbf{\Lambda}$

Define $\mathbf{S}_{[+]} = \mathbf{0}$

for $k = 1, \dots, K$ **do**

$\quad \mathbf{S}_{[k]} \leftarrow \frac{1}{2}\mathbf{\Xi}\mathbf{R}_{[k]}\mathbf{\Xi}$

end

$\mathbf{Z} \leftarrow [\text{vec}\{\mathbf{S}_{[1]}, \dots, \mathbf{S}_{[\text{k}]}, \dots, \mathbf{S}_{[\text{K}]}\}]$

$\mathbf{U}\mathbf{\Delta}\mathbf{V}^T \leftarrow \text{SVD}(\mathbf{Z})$

$\boldsymbol{\alpha} \leftarrow \mathbf{v}_1 \times (\mathbf{v}_1^T\mathbf{1})^{-1}$

for $k = 1, \dots, K$ **do**

$\quad \mathbf{S}_{[+]} \leftarrow \mathbf{S}_{[+]} + (\mathbf{S}_{[k]} \times \alpha_{[k]})$

end

$\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \leftarrow \text{EVD}(\mathbf{S}_{[+]})$

Algorithm 1: CovSTATIS algorithm

0.1.0.9842; Aust & Barth, 2018), *prettyGraphs* (Version 2.1.6; Beaton et al., 2014b), *purrr* (Version 0.2.5; Henry & Wickham, 2018), *readr* (Version 1.1.1; Wickham et al., 2017), *reticulate* (Version 1.9; Allaire, Ushey, & Tang, 2018), *RevoUtils* (Version 11.0.1; Corporation, 2018b, 2018a), *RevoUtilsMath* (Version 11.0.0; Corporation, 2018a), *stringr* (Version 1.3.1; Wickham, 2018b), *tibble* (Version 1.4.2; Müller & Wickham, 2018), *tidyr* (Version 0.8.1; Wickham & Henry, 2018), and *tidyverse* (Version 1.2.1; Wickham, 2017) for all our analyses.

Results

Like in the other example RMarkdown file, we call off to an `R` script

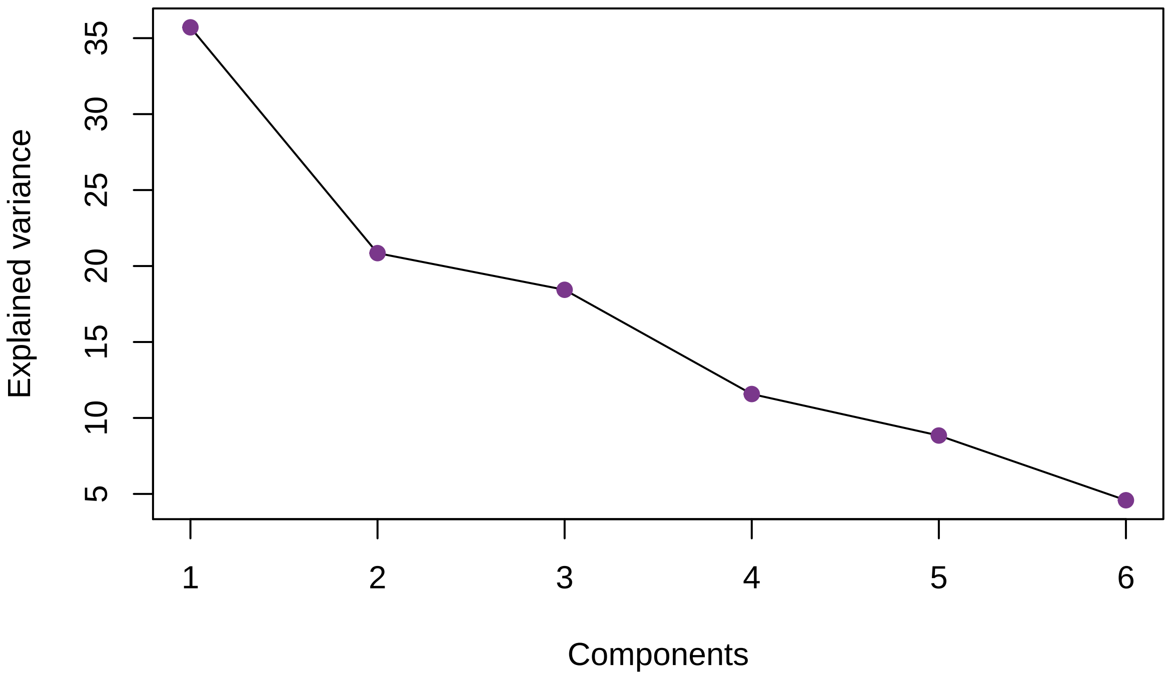

Figure 1. Scree plot of the covSTATIS compromise results.

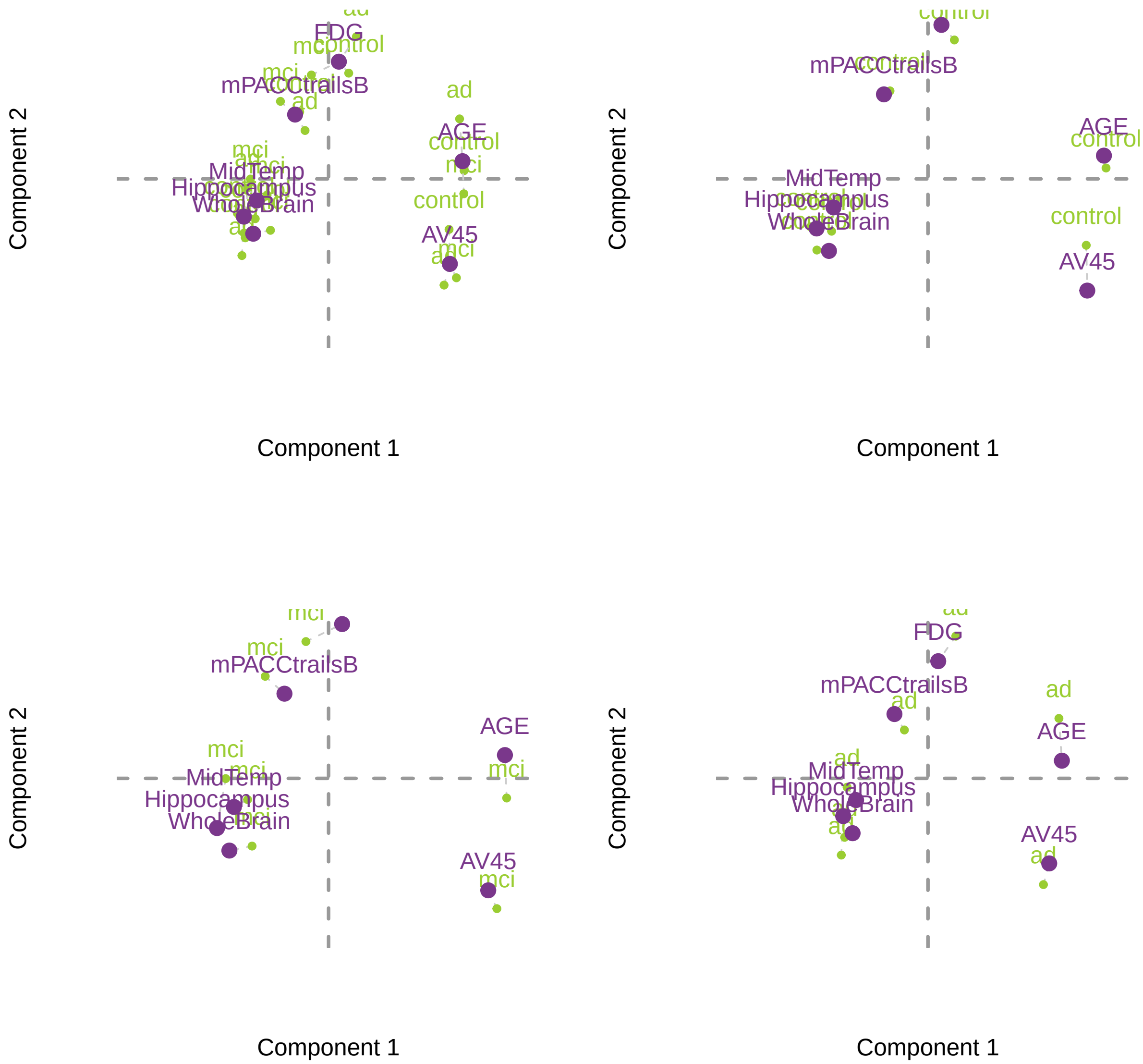

Figure 2. Component maps with the compromise (purple) and each group's results (green) projected onto the compromise.

Discussion

This example manuscript provides an illustrative view of how to write a reproducible paper in RMarkdown with the help of various external tools (e.g., `LaTeX`, `Python`) and

`R`/RStudio tools. References can be included in two ways: through `LaTeX` bibliography files or through the use of additional `R` packages and Rstudio plugins, such as `citr`.

References

Allaire, J., Ushey, K., & Tang, Y. (2018). *Reticulate: Interface to 'python'*. Retrieved from https://CRAN.R-project.org/package=reticulate

Auguie, B. (2017). *GridExtra: Miscellaneous functions for "grid" graphics*. Retrieved from https://CRAN.R-project.org/package=gridExtra

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Beaton, D. (n.d.). *GSVD: The generalized singular value decomposition*.

Beaton, D., Fatt, C. R. C., & Abdi, H. (2014a). An ExPosition of multivariate analysis with the singular value decomposition in R. *Computational Statistics & Data Analysis*, *72*(0), 176–189. https://doi.org/10.1016/j.csda.2013.11.006

Beaton, D., Fatt, C. R. C., & Abdi, H. (2014b). An ExPosition of multivariate analysis with the singular value decomposition in R. *Computational Statistics & Data Analysis*, *72*(0), 176–189. https://doi.org/10.1016/j.csda.2013.11.006

Corporation, M. (2018a). *RevoUtilsMath: Microsoft r services math utilities package*.

Corporation, M. (2018b). *RevoUtils: Microsoft r utility package*.

Henry, L., & Wickham, H. (2018). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

it, W. wrote. (n.d.). *Covstatis: What the package does (title case)*.

Kassambara, A., & Mundt, F. (2017). *Factoextra: Extract and visualize the results of multivariate data analyses*. Retrieved from https://CRAN.R-project.org/package=factoextra

Müller, K. (2017). *Here: A simpler way to find your files*. Retrieved from https://CRAN.R-project.org/package=here

Müller, K., & Wickham, H. (2018). *Tibble: Simple data frames*. Retrieved from https://CRAN.R-project.org/package=tibble

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Sunderland, K., & Beaton, D. (n.d.). *Ours: OUtliers and robust structures*.

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wickham, H. (2018a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2018b). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2018). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & Francois, R. (2017). *Readr: Read rectangular text data*. Retrieved from https://CRAN.R-project.org/package=readr

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.name/knitr/

Zhu, H. (2018). *KableExtra: Construct complex table with 'kable' and pipe syntax.* Retrieved from https://CRAN.R-project.org/package=kableExtra