



Reproducible Workflow

R Markdown

Michelle Donzallaz

Slides and Material

You can find the slides and materials here: <https://github.com/jstbcs/ReproducibleWorkflowWorkshop>.

Materials heavily build on previous workshops by Julia Haaf

Overview

1. Minimizing mistakes and reproducibility.
2. What is RMarkdown and how can I use it (e.g., to write a thesis)?
3. What is `git` and how can I use it?

1. Minimizing mistakes and reproducibility.

Replicability Crisis

- Failures to replicate (e.g. Ebersole et al., 2016; Open Science Collaboration, 2015; Wagenmakers et al., 2016).
- Fraud (Bhattacharjee, 2013).
- Improbable findings have been published in top-tier journals (e.g. Bem, 2011).



Proposed Solutions

- Change the incentive structure (e.g., Nosek et al., 2015; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).
- Be transparent and open (e.g. Rouder, 2016; Wicherts, Bakker, & Molenaar, 2011).
- Change the statistical approach (e.g. Benjamin et al., 2018; Erdfelder, 2010; Rouder, Morey, Verhagen, Province, & Wagenmakers, 2016)



Proposed Solutions

- Change the incentive structure (e.g., Nosek et al., 2015; Wagenmakers et al., 2012).
- Be transparent and open (e.g. Rouder, 2016; Wicherts et al., 2011).
- Change the statistical approach (e.g. Benjamin et al., 2018; Erdfelder, 2010; Rouder et al., 2016)

We assume people do stuff on purpose.



Mistakes in Psychological Science

Sources of mistakes:

- Errors when programming an experiment or study (e.g. randomization).
- Equipment failure (e.g. responses are collected unreliably).
- Lost data.
- Errors when coding the analysis (e.g. with data cleaning).
- Errors when reporting the analysis (e.g. typos).



Consequences in data science

- Prevalence: Roughly half the publications in 30 years of literature contained at least one malformed statement of a statistical test (Nuijten, Hartgerink, Assen, Epskamp, & Wicherts, 2016). Do you think these would be caught in an internal presentation in a company?
- Bias: Simple mistakes tend to go in scientists' preferred direction (Gould, 1996). You want to present clear results to your client.
- Persistence: Once in the literature mistakes are almost impossible to detect (Rouder, Haaf, & Snyder, 2019). Or, once a change is implemented on a website/at a company it is much more difficult to change it back.

High-Reliability Organizations

Principles for Avoiding Mistakes

1. Sensitivity to operations: Focus on processes instead of outcomes.
2. Preoccupation with failure: Look for ways to proactively anticipate and avoid mistakes, and take small mistakes seriously.
3. Resilience in the face of failure and reluctance to simplify: In a resilient team, when things go wrong — and they will — people talk about them, document them, and learn from them.
4. Deference to expertise: Each team member has certain expertise.



From Principles to Practices

1. Adopting a team culture focused on learning from mistakes.
2. Implementing radical computer automation.
3. Standardizing organizational strategies across team members (everyone documents solutions the same way).
4. **Ensuring that statistical analyses are coded.**
5. **Developing expanded reports in which documentation of analyses is woven into the documentation files.**



Reproducibility

Fully Reproducible

- Reproducible analysis.
- Reproducible graphs and tables.
- Reproducible numbers in text.

Who Can Reproduce When?

- Ideal: [Code containerization](#).
- Minimal: Provide a list of packages and software needed (Open source!).
- Utopian: “I will be able to fully reproduce my analysis by 2035.”

Tools for Reproducibility

git

- Versioning tool for collaboratively working on a product.
- Avoid retaining multiple versions of the same work product.
- ‘analysis_final_final_B.R’.
- Tutorial: Vuorre & Curley (2018).

R Markdown

- Document format embedding code chunks into text documents.
- Avoid copy-and-paste of results.
- Book: <https://bookdown.org/yihui/rmarkdown/>.

Short Break (5 minutes)

If you still need to install some things from [the instructions](#) now is a good time.

link: <https://github.com/jstbcs/ReproducibleWorkflowWorkshop#how-to-prepare>

2. What is R Markdown and how can I use it?

General Remarks About Working with R

(or, really, just doing data analysis)

- Standardization: Think about your folder structure.
- Clean coding: Develop your own style guide. Note that you might have to adopt your future team's style.
- Use comments throughout the analysis (R Markdown can help).
- Use relative paths (“../material/example-paper.pdf” instead of “/home/julia/Documents/Git/ReproducibleWorkflowWorkshop/material/example-paper.pdf”)!
- We could have an entire lecture just on this topic. Now is the time for you to adopt good habits!

What is R Markdown?

- Document format embedding code chunks into Markdown documents.
- What's Markdown? “A lightweight markup language.”
- Easy-to-read and easy-to-write plain text format. Easier than latex (but supports latex).
- Supports many output formats (html, pdf, ms word, ...)

Basic Rules of Markdown

Header

Level 1 Header

Level 2 Header

Level 3 Header

Level 1 Header

Level 2 Header

Level 3 Header

Basic Rules of Markdown

Bold and Italic

To write text in bold or in italic, just add `*`'s around the word(s)

You can print one word `**bold**` and another `*italic*`.

Or `*multiple words*` can `**be altered**`.

You can print one word **bold** and another word *italic*.

Or *multiple words* can **be altered**.

Basic Rules of Markdown

Lists

- This
- Is
- A
- List

- This
- Is
- A
- List

Basic Rules of Markdown

Overview

Check out the cheat sheets in the folder [material](#).

Task I

Go to stackedit.io ➤ Write Markdown!

Use the cheat sheet!

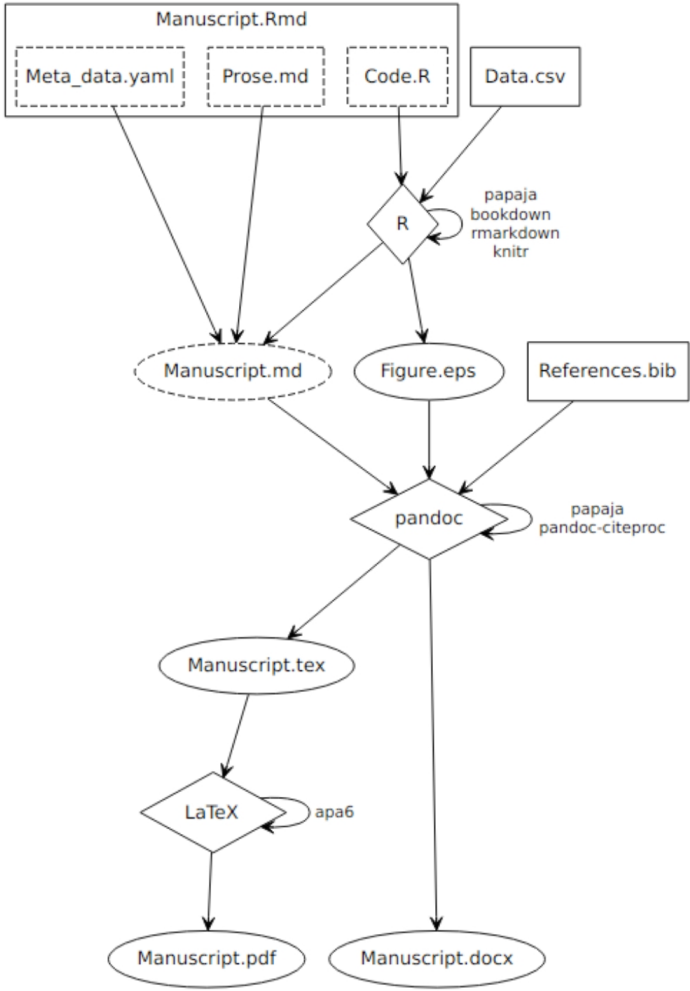


R Markdown

What is Possible?

- Scientific papers
- Slides
- HTML sites
- CVs
- Books
- ...

From Rmd to PDF



Task II

- Open RStudio.
- Go to 'File' ► 'New File' ► 'R Markdown'
- Add title and author, choose PDF, and click 'ok'.
- Save the file.
- Click the 'knit' button.



YAML header

YAML = “Yet Another Markup Language”

```
---  
title: "My first Markdown"  
author: "My name"  
output: pdf_document  
---
```

YAML header

YAML = “YAML Ain’t Markup Language”

```
title      : "Creative or Not? Hierarchical Diffusion Modeling of the Creative Evaluation"
shorttitle : "Creative or Not?"
```

author:

```
- name      : "Michelle C. Donzallaz"
  affiliation : "1"
  address    : "Nieuwe Achtergracht 129-B, 1018 WS Amsterdam"
  corresponding : yes    # Define only one corresponding author
  email      : "m.c.donzallaz@uva.nl"
- name      : "Julia M. Haaf"
  affiliation : "1"
- name      : "Claire E. Stevenson"
  affiliation : "1"
```

YAML header

continued...

affiliation:

```
- id          : "1"
  institution : "University of Amsterdam"
```

authornote: |

This report was written in R-Markdown with code for data analysis integrated into the text.

note: "\\copyright American Psychological Association, 2022. This paper (...)"

abstract: |

```
(...)
<!-- https://tinyurl.com/ybremelq -->
```

bibliography : ["references.bib", "r-references.bib"]

YAML header

continued...

```
keywords      : "creativity, evaluation, diffusion model, Bayesian hierarchical modeling"
floatsintext  : yes
figurelist    : no
tablelist     : no
footnotelist  : no
figsintext    : yes
linenumbers   : no
mask          : no
draft         : no
nocite: ["@Singmann:2018", "@Barchard:2012", "@DonzallazEtAl:2022"]
header-includes:
  - \raggedbottom
documentclass : "apa6"
classoption   : "man, noextraspace"
output        : papaja::apa6_pdf
```

R code chunks

```
```{r name, options = ...}  
1 + 1
```
```

```
## [1] 2
```

- Code chunks include code (languages other than R also possible, e.g., Python).
- Code is executed when knitting a R Markdown file.
- In R Studio you can use the shortcut `Ctrl + alt + i` to insert a code block.
- No underscore in chunk name, no name twice.

Chunk options

- `eval = FALSE`
- `echo = FALSE`
- `warnings = FALSE`
- `cache = TRUE`
- `child = "rmdfilename.Rmd"`
- See [rmarkdown-cheatsheet2](#) for most important options.

Global chunk options

Adjust all chunk options in a Rmd file with the following command:












```
knitr::opts_chunk$set(echo = TRUE)
```

Pro-tip: If the R chunks are too long use `source()` to load R-files, and the chunk option `child = myanalysis.Rmd` to include other R Markdown files.

Check out papaja!

- “Prepare APA Journal Articles with R Markdown”
- [R package](#) developed by Frederik Aust with great [documentation](#)!
- create computationally reproducible, submission-ready manuscripts (or theses!)

Example: folder structure on GitHub

| | | |
|---|-------------------------|--|
| main  con-diffusion / Manuscript / | | |
| michelle Merge branch 'main' of https://github.com/mdllz/con-diffusion into main | | |
| .. | | |
|  | Analysis | updated files |
|  | Chapters | changed prefix of table & figure numbers |
|  | Data | added data description |
|  | Figures | fixed typo in figure description |
|  | Supplementary-materials | changed prefix of table & figure numbers |
|  | Tables | updated files |
|  | Manuscript.Rmd | changed note |
|  | Manuscript.pdf | changed note |
|  | r-references.bib | updated files |
|  | references.bib | added OSF repo to ref list |

Example: Manuscript/Manuscript.Rmd

```
80   ```{r 'Intro', child="Chapters/Introduction.Rmd"}
81   ```
82
83
84
85   ```{r 'Model and Procedure S1', child="Chapters/Procedure_and_Model.Rmd"}
86   ```
87
88   ## Results Study 1
89   ```{r 'Results S1', child="Chapters/Results.Rmd"}
90   ```
91
92
93   ```{r 'Model and Procedure S2', child="Chapters/Procedure_and_ModelS2.Rmd"}
94   ```
95
96   ## Results Study 2
97
98   ```{r 'Results S2', child="Chapters/ResultsS2.Rmd"}
99   ```
100
101   \clearpage
102
103   # Discussion
104
105   ```{r 'Conclusion', child="Chapters/Discussion.Rmd"}
106   ```
```

Example: Manuscript/Chapters/Introduction.Rmd

```
1 ---
2 title: "Creative or not, Introduction"
3 output:
4   pdf_document: papaja::apa6_pdf
5   html_document: default
6   word_document: papaja::apa6_word
7 ---
8 Creative ideas are essential for tackling today's problems, from personal obsta
9 While the standard definition of creativity states that originality and utilit
10 In order to examine the evaluation phase of the AUT, we focus on the decision-m
11 To gain insight into the cognitive basis of the evaluation process, we model C
12 The DDM has commonly been applied to studies of language, perception, or memo
13 Taking into account both responses and RTs, the DDM allows us to explore the r
14
15 \vspace{5mm}
16 ```{r 'figureDDM', child = "../Figures/DDM_figure.Rmd"}
17
18 ```
19
20 As a plausibility check for our cognitive modeling approach, we also examine how
```

Example: Manuscript/Chapters/Results.Rmd

```
1 ---
2 title: "Creative or not, Results Study 1"
3 output:
4   pdf_document: papaja::apa6_pdf
5   html_document: default
6   word_document: papaja::apa6_word
7 ---
8
9 ```{r 'Analysis S1', child = "../Analysis/Analysis.Rmd"}
10 ```
11
12 The data cleaning is described in Appendix B. The cleaned dataset comprised `r length(unique(dat1$pers))` participants
13
14 ### Model Fit
15
```

Task III

- Go back to your Rmd file.
- Add a chunk calculating the standard deviation (`sd()`) of the speed of cars.
- Print the results in a sentence below the chunk.
- Knit the PDF.
- Not happy with result? Round it!



Inline R-Code

This is an inline expression 5.2876444 in the output.

```
sdcars <- sd(cars$speed)
```

The standard deviation of the speed of cars is 5.29.

L^AT_EX

- You can add **L^AT_EX** code to your text.
- Math mode for formulas.

L^AT_EX

- You can add **L^AT_EX** code to your text.
- Math mode for formulas.

You can simply add a formula: `$\frac{SD}{\sqrt{N}}$`.

Result:

You can simply add a formula: $\frac{SD}{\sqrt{N}}$.

Math mode and R code

You can use R within math mode.

Math mode and R code

You can use R within math mode.

```
sd1 <- 15
```

```
n1 <- 17
```

Math mode and R code

You can use R within math mode.

```
sd1 <- 15
```

```
n1 <- 17
```

```
$se = \frac{\`r sd1\`}{\sqrt{\`r n1\`}} \approx \`r round(sd1/(sqrt(n1)), 2)\`$
```

$$se = \frac{15}{\sqrt{17}} \approx 3.64$$

References

- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407–425. Retrieved from <http://dx.doi.org/10.1037/a0021524>
- Benjamin, D. J., Berger, J., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., ... Johnson, V. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2, 6.
- Bhattacharjee, Y. (2013). The mind of a con man. New York Times, April 26, 2013. Retrieved from <http://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html?pagewanted=all>
- Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., ... Nosek, B. A. (2016). Many labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, 67, 68–82. Retrieved from <http://ezid.cdlib.org/id/doi:10.17605/OSF.IO/QGJM5>
- Erdfelder, E. (2010). A note on statistical analysis. *Experimental Psychology*, 57(1-4). Retrieved from [10.1027/1618-3169/a000001](http://dx.doi.org/10.1027/1618-3169/a000001)
- Gould, S. J. (1996). *The mismeasure of man*. New York: WW Norton & Company.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425.
- Nuijten, M. B., Hartgerink, C. H., Assen, M. A. van, Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, 48(4), 1205–1226.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6521), 943. Retrieved from dx.doi.org/10.1126/science.aac4716
- Rouder, J. N. (2016). The what, why, and how of born-open data. *Behavioral Research Methods*, 48, 1062–1069. Retrieved from [10.3758/s13428-015-0630-z](http://dx.doi.org/10.3758/s13428-015-0630-z)
- Rouder, J. N., Haaf, J. M., & Snyder, H. K. (2019). Minimizing mistakes in psychological science. *Advances in Methods and Practices in Psychological Science*.
- Rouder, J. N., Morey, R. D., Verhagen, J., Province, J. M., & Wagenmakers, E.-J. (2016). Is there a free lunch in inference? *Topics in Cognitive Science*, 8, 520–547.