Reproducible Workflow

Git and Rmarkdown: Two Useful Tools for a Fully Reproducible Workflow

Julia Haaf

Format

- This is a workshop.
- But it is short.
- We will do some hands-on stuff but there is not a lot of time for practice.
- Towards the end of the session today we can discuss whether it makes sense to extend the workshop at another time.

Slides and Material

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

Who Am I?

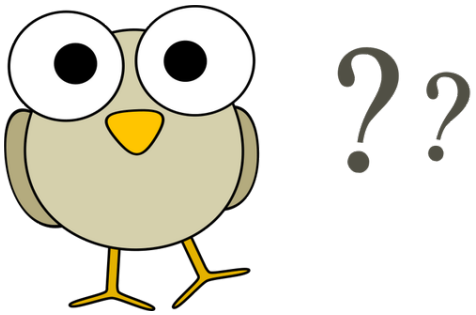
Julia Haaf

Postdoc in the Bayesian lab (Psychological Methods)

email: j.m.haaf@uva.nl

Room: G0.34

Who Are You?



Overview

1. How can we minimize mistakes in psychological science?
2. What does it mean to have a fully reproducible pipeline?
3. What is `RMarkdown` and how can I use it?
4. Writing an APA-style paper that is fully reproducible.
5. What is `git` and how can I use it?

1. How can we minimize mistakes in psychological science?

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 et al., 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

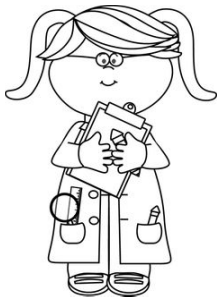
Sources of mistakes:

- Errors when programming the experiment (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).

Lab Practices Under the Microscope

Think about your own experience:

- Is there time pressure to collect data?
- Are there checks for coding experiments/surveys?
- Are there checks for running analyses?



Consequences

- 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).
- Bias: Simple mistakes tend to go in researchers' preferred direction (Gould, 1996).
- Persistence: Once in the literature mistakes are almost impossible to detect (Rouder, Haaf, & Snyder, 2019).

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 lab, when things go wrong — and they will — people talk about them, document them, and learn from them.
4. Deference to expertise: Each lab member has certain expertise.

From Principles to Practices

1. Adopting a lab culture focused on learning from mistakes.
2. Implementing radical computer automation.
3. Standardizing organizational strategies across lab members.
4. **Ensuring that statistical analyses are coded.**
5. **Developing expanded manuscripts in which documentation of analyses is woven into the manuscript files.**

2. What does it mean to have a fully reproducible pipeline?

Science vs. Pseudo-Science

> “An article about computational results is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result.” Claerbout & Karrenbach, 1992

Different Reproducibilities

- Empirical Reproducibility
- Computational Reproducibility
- Statistical Reproducibility
- Plus: machine vs. human readability

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.
- 'paper_final_final_B.docx'.
- Tutorial: Vuorre & Curley (n.d.).

`R Markdown`

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

Short Break (5 minutes)

If you need help installing some things from the instructions now is a good time.

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

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

Slides are at rmarkdown.html.

5. What is `git` and how can I use it?

Slides are at git.html.

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

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

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.

Vuorre, M., & Curley, J. P. (n.d.). *Curating research assets: A tutorial on the git version*. Retrieved from https://psyarxiv.com/6tzh8

Wagenmakers, E.-J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., R. B. Adams, J., … Zwaan, R. A. (2016). Registered replication report: Strack, martin, & stepper (1988). *Perspectives on Psychological Science*, *11*(6), 917–928. Retrieved from https://doi.org/10.1177/1745691616674458

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*, 627–633. Retrieved from https://doi.org/10.1177/1745691612463078

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, *6*(11), e26828. Retrieved from http://www.plosone.org/annotation/listThread.action?root=19627