

A tutorial on setting up a reproducible workflow in R and R Studio

Richard Ramsey¹

¹ ETH Zürich

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line. Enter author note here.

The authors made the following contributions. Richard Ramsey: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Richard Ramsey, Postal address. E-mail: richard.ramsey@hest.ethz.ch

11

Abstract

12 tbc

13

Keywords: keywords

14

Word count: X

A tutorial on setting up a reproducible workflow in R and R Studio

Introduction

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (Simmons, Nelson, & Simonsohn, 2012).

Participants

Material

Procedure

Data analysis

We used R (Version 4.4.1; R Core Team, 2023) and the R-packages *citr* (Version 0.3.2; Aust, 2019), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.1; Wickham, 2016), *lubridate* (Version 1.9.4; Grolemund & Wickham, 2011), *papaja* (Version 0.1.3; Aust & Barth, 2023), *patchwork* (Version 1.3.0; Pedersen, 2024), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *RColorBrewer* (Version 1.1.3; Neuwirth, 2022), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019), *tinylab* (Version 0.2.4; Barth, 2023) and *tinytex* (Version 0.54; Xie, 2019) for all our analyses.

Results

Descriptive statistics

Raw data plots

Accuracy by condition violin plot

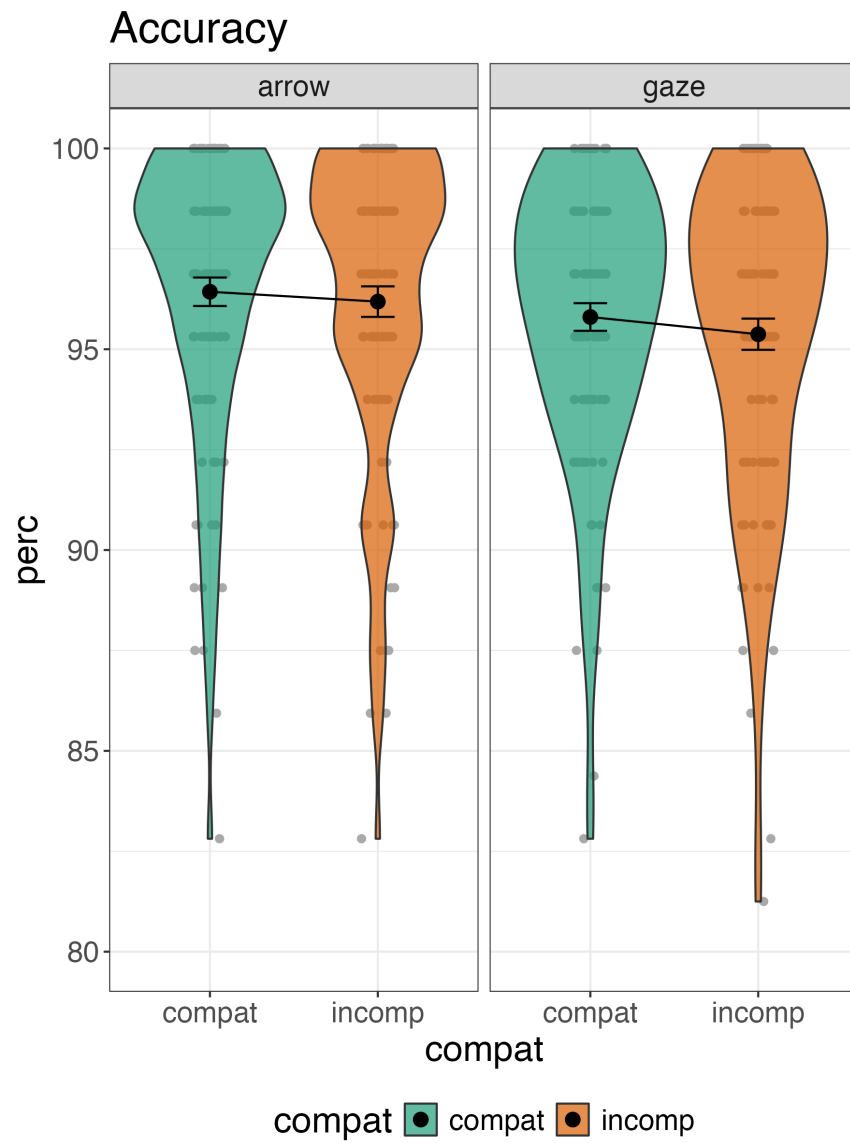


Figure 1. Accuracy is quite high for both types of stimuli

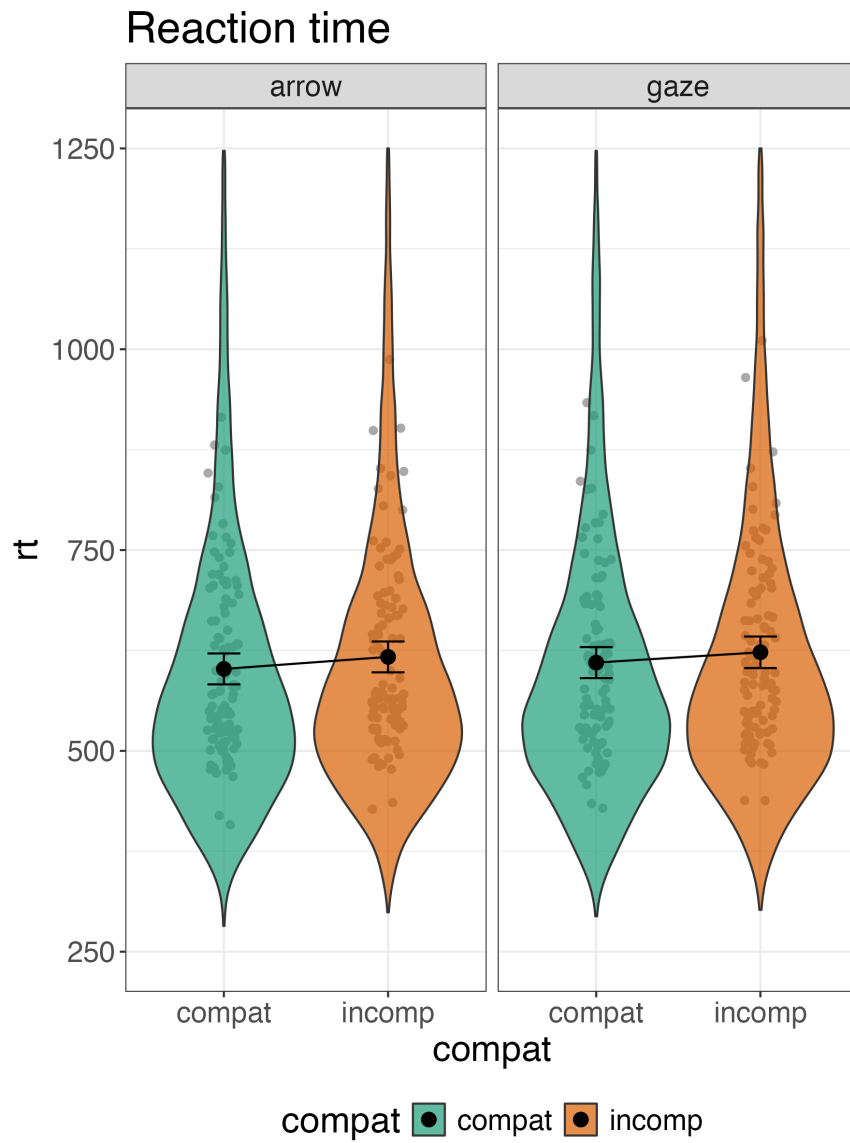


Figure 2. RT as a function of stimuli and conditions.

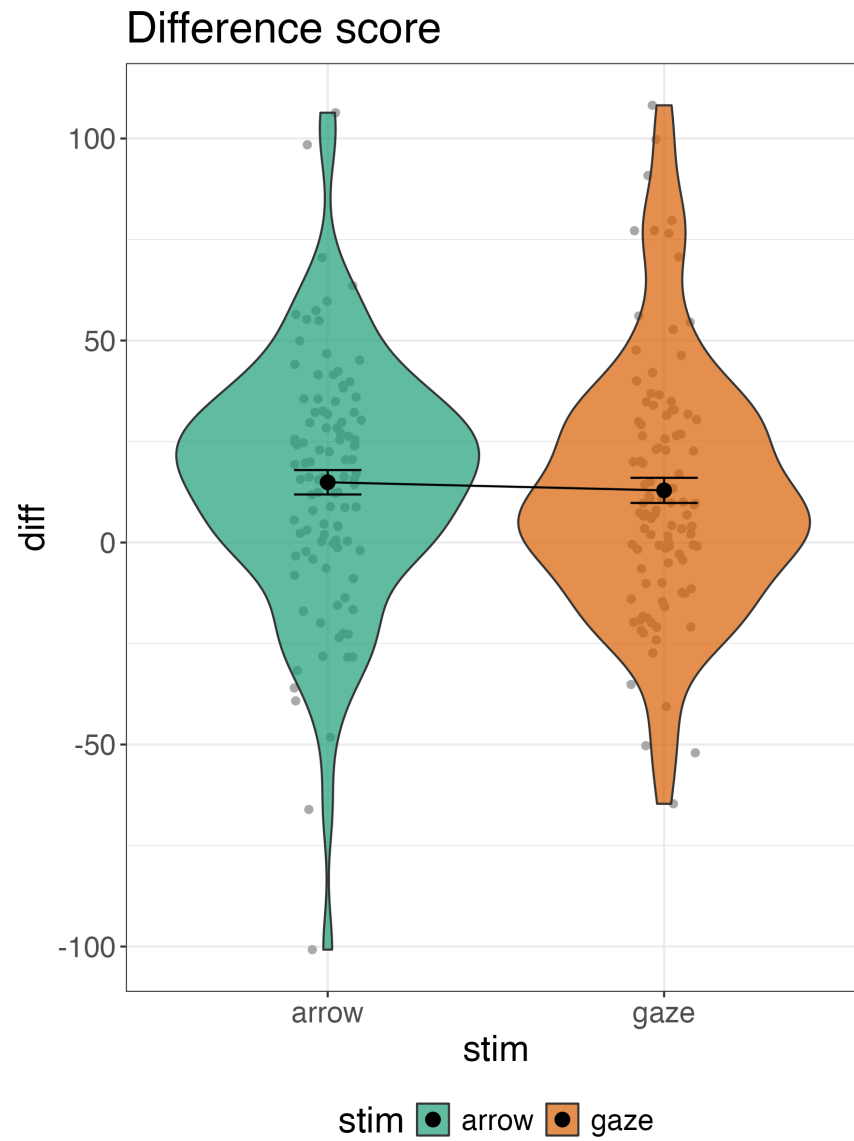


Figure 3. RT difference scores (incomp - compat) by stimulus type.

Difference score density by quantile

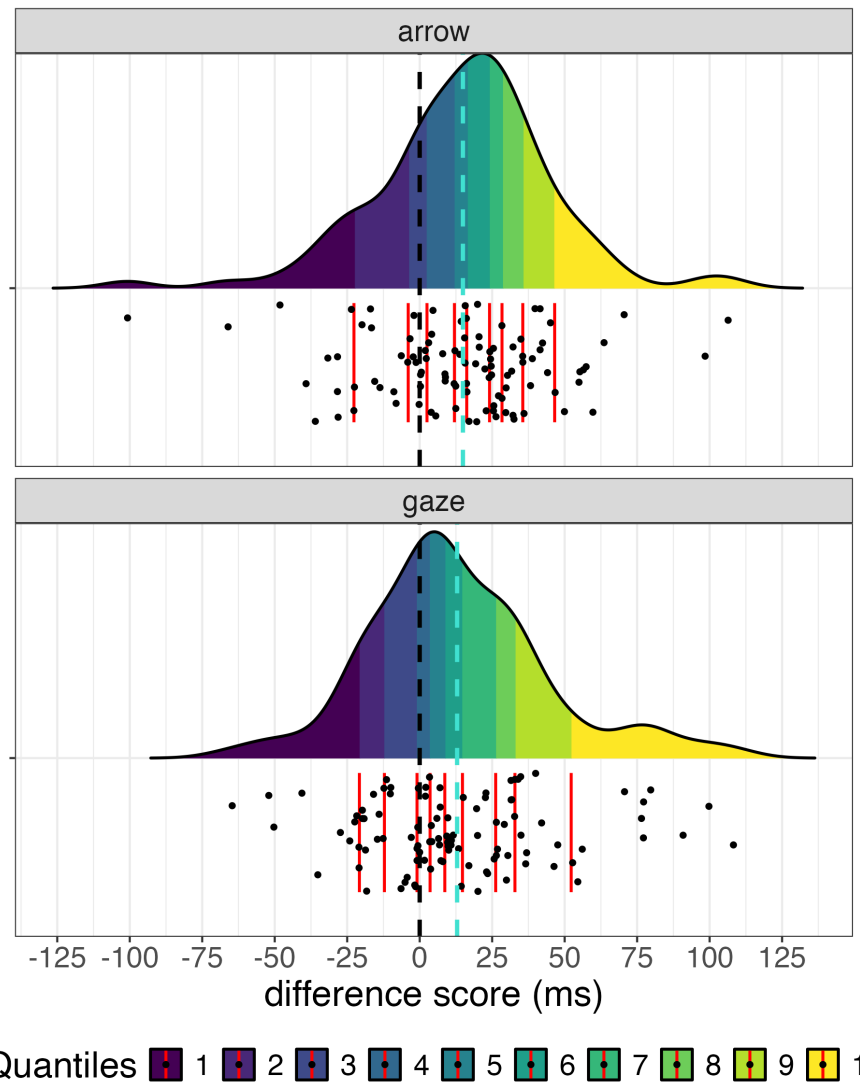


Figure 4. RT difference scores (incomp - compat) as a density plot.

41 Inferential statistics

42 Now we plot and tabulate parameters from the posterior distribution.

43 Fixed effects from model b2

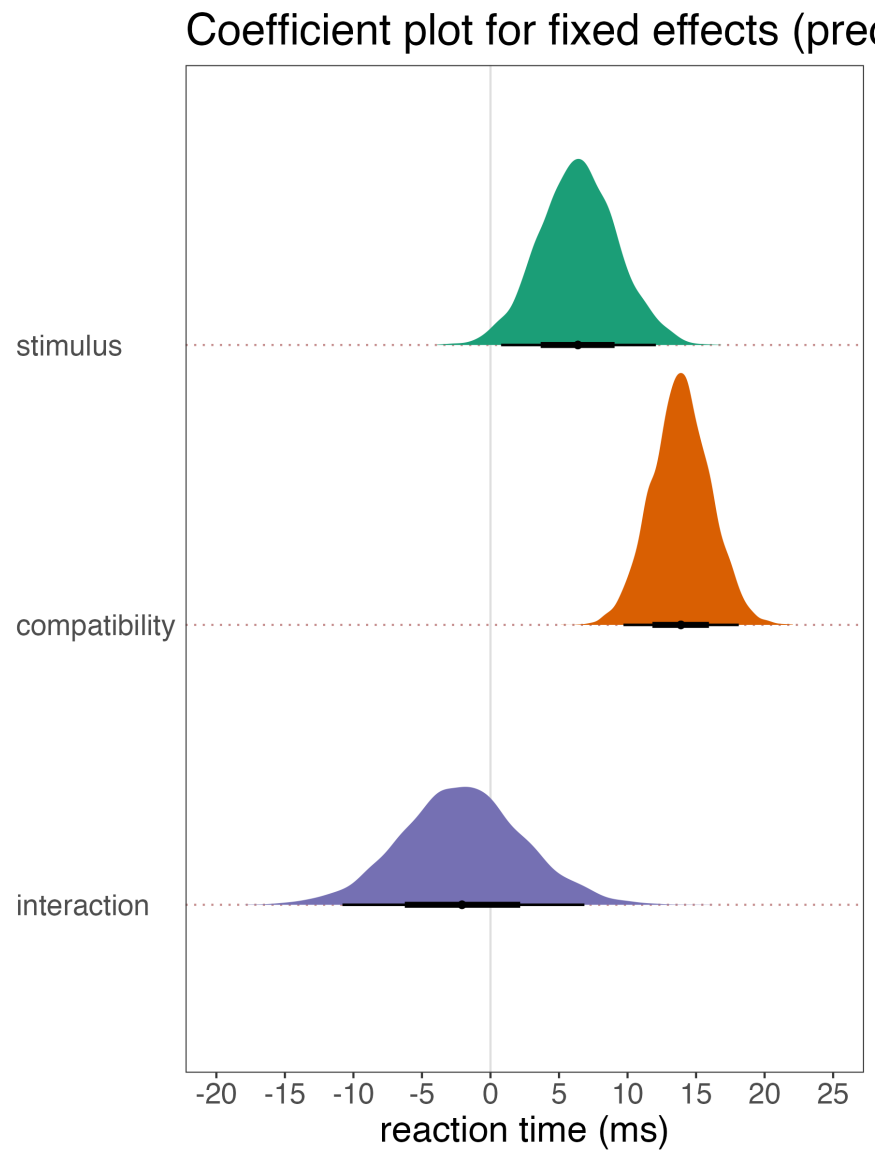


Figure 5. Fixed effects from model b2.

Table 1
Fixed effects from model b2.

term	value	.lower	.upper
intercept	613.94	593.62	634.79
stimulus	6.38	0.77	12.07
compatibility	13.89	9.70	18.11
interaction	-2.08	-10.80	6.85

Note. Median point estimates and 95% quantile intervals shown.

References

- Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from <https://github.com/crsh/citr>
- Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Barth, M. (2023). *tinylabels: Lightweight variable labels*. Retrieved from <https://cran.r-project.org/package=tinylabels>
- Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3), 1–25. Retrieved from <https://www.jstatsoft.org/v40/i03/>
- Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from <https://CRAN.R-project.org/package=tibble>
- Neuwirth, E. (2022). *RColorBrewer: ColorBrewer palettes*. Retrieved from <https://CRAN.R-project.org/package=RColorBrewer>
- Pedersen, T. L. (2024). *Patchwork: The composer of plots*. Retrieved from <https://CRAN.R-project.org/package=patchwork>
- R Core Team. (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). *A 21 Word Solution* (SSRN Scholarly Paper No. 2160588). Rochester, NY: Social Science Research Network. <https://doi.org/10.2139/ssrn.2160588>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from <https://CRAN.R-project.org/package=forcats>
- Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from <https://CRAN.R-project.org/package=stringr>

- 72 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . . Yutani,
73 H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
74 <https://doi.org/10.21105/joss.01686>
- 75 Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar*
76 *of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- 77 Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from
78 <https://CRAN.R-project.org/package=purrr>
- 79 Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. Retrieved
80 from <https://CRAN.R-project.org/package=readr>
- 81 Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. Retrieved from
82 <https://CRAN.R-project.org/package=tidyr>
- 83 Xie, Y. (2019). TinyTeX: A lightweight, cross-platform, and easy-to-maintain LaTeX
84 distribution based on TeX live. *TUGboat*, 40(1), 30–32. Retrieved from
85 <https://tug.org/TUGboat/Contents/contents40-1.html>