Reproducible research with R and RMarkdown: lessons learned

Anónimo

Abstract Recently I published my first first-author paper [OMITTED FOR REVIEW], which was was my first attempt to do reproducible research. This resume describes which was the software and packages used and some of the challenges faced.

Palabras clave: Reproducible Research - RMarkdown - Quarto

Introduction

In May 2022, I published my first first-author paper [OMITTED FOR REVIEW]. This work was my first attempt to do reproducible research. This resume describes which was the software and packages used and some of the challenges faced. In addition, the complete code used in the paper is available in a public repository on GitHub: [OMITTED FOR REVIEW].

Software and packages used

RMarkdown (Xie, Allaire, and Grolemund 2018; Xie, Dervieux, and Riederer 2020) is a format for writing reproducible reports with R: authors can add text, code, and results (such as calculations, tables, and figures) in this format. Then, authors can export the final report to PDF, Word, and HTML (web pages). RMarkdown is the format used in the study. The software used to develop the study was R (R Core Team 2021), RStudio IDE (RStudio Team 2020), Zotero (a reference manager), Git (for version control), and GitHub (for online hosting of the code).

The packages used to structure the paper were: Rmarkdown (Allaire et al. 2022), bookdown (Xie 2016, 2022), flextable (to create tables) (Gohel 2022), and ggplot2 with patchwork (to create and arrange graphs) (Wickham et al. 2022; Pedersen 2020). In addition, the paper lists the packages used to perform the data and statistical analysis.

Challenges faced

The journal the authors chose to submit the study only accepted submissions using word files (`.docx`). The first challenge was that only two authors were R programmers. Only the first author knew how to use RMarkdown, Git, and GitHub. The workflow of doing the analysis and writing the paper consisted in writing in RMarkdown and exporting a Word file to send to all the authors. After each iteration of contributions and reviews, the authors returned the Word files with their contributions, and the first author added the contributions to the RMarkdown file. The task of adding the contributions from Word to RMarkdown was time-consuming.

Another challenge was generating tables in a suitable format when exported to Word. After testing several R packages and functions to generate tables, the chosen package was flextable. The package has extensive documentation published in a book format, which is an excellent asset for learning how to use it: https://ardata-fr.github.io/flextable-book/.

Organizing the code: R packages

The first author of the paper chose to organize the code and files using the structure of an R package. A package structure creates a consistent way to organize the files. It is also interesting because all the dependencies (the R packages used) are stored in the DESCRIPTION file and can be easily installed on other computers using the function `devtools::install_deps()`.

The organization of the code and files of the paper follows this pattern: The R functions are created and stored in the `R/` directory; the raw dataset and the code to clean the data are stored in the `data-raw/` directory; the clean data stored in the `data/` directory; all the dependencies stored in the `DESCRIPTION` file; and all the other content stored in the `inst/` directory: the directory for the analysis performed, the word template, the references files (`.bib`) both exported from Zotero and also created with R and the RMarkdown file of the article. The R package usethis (Wickham, Bryan, and Barrett 2022) is an important tool to create the structure of a package. The book R packages is a great source of content about how to create a package.

Conclusion

The course of development and writing of this paper offered much learning. The first author learned much about writing a reproducible document and collaborating with RMarkdown. Since the other authors did not use RMarkdown, it is essential to have feedback about the reproducibility of the study. In addition, it can help to do research more reproducible in the future. In the future, it would be good to have a free feature on RStudio IDE or other software that allowed contributions to an RMarkdown or Quarto file from researchers who both are programmers and no programmers.

References

10 Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2022. *Rmarkdown: Dynamic Documents for r*.

Gohel, David. 2022. *Flextable: Functions for Tabular Reporting*. https://CRAN.R-project.org/package=flextable.

Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

RStudio Team. 2020. *RStudio: Integrated Development Environment for r*. Boston, MA: RStudio, PBC. http://www.rstudio.com/.

Wickham, Hadley, Jennifer Bryan, and Malcolm Barrett. 2022. *Usethis: Automate Package and Project Setup*. https://CRAN.R-project.org/package=usethis.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2022. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://CRAN.R-project.org/package=ggplot2.

Xie, Yihui. 2016. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/bookdown.

———. 2022. *Bookdown: Authoring Books and Technical Documents with r Markdown*.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown-cookbook.