

Reproducible Research in R

2025-01-21

Contents

Reading: 1

Reading:

Required Reading (everyone):

- R for Data Science, 28.
- Handout: Jonathan Gilligan, “Introduction to Reproducible Research”.
- Handout: Jonathan Gilligan, “Introduction to Git and GitHub Classroom”.

Optional Extra Reading:

- Handout: Lenny Teytelman, “No More Excuses for Non-Reproducible Methods,” *Nature* **560**, 411 (2018), [doi:10.1038/d41586-018-06008-w](https://doi.org/10.1038/d41586-018-06008-w).
- Web Page: Quarto Website.

Reading Notes:

The key reading for today is the handout, “Introduction to Reproducible Research,” which introduces the concepts behind reproducible research and presents historical examples and context for the motivations and goals behind the growing adoption of reproducible research practices.

You should read this handout first.

The other major readings are:

- “Introduction to Git and GitHub Classroom,” which explains the `git` revision control system and the way we use it with GitHub and GitHub Classroom, for homework assignments in this course.
- Chapter 28 of *_R for Data Science*,” which explains the Quarto system for integrating text, data, and R code to create fully reproducible research reports, manuscripts, web pages, blogs, etc.

Tools like Quarto are important because they let you recreate an entire document—even an entire book—with the push of a button, and automatically update it if data or analysis code changes.

The optional readings are:

- A short paper that appeared in the journal *Nature* describing how badly many cancer research projects went wrong because they failed to follow reproducible research methods. The paper argues that “There are no more excuses for non-reproducible research methods.”
- If you want to learn more about using Quarto, the website for the Quarto project has extensive documentation.