Reproducible Workflow with `rmarkdown`

Tyson S. Barrett

Outline

What is "reproducibility"?

How can `rmarkdown` help?

Integrating `rmarkdown` into your workflow

The Ins-and-Outs of `rmarkdown`

References

Goodman, SN, Fanelli, D, Ioannidis, JPA. (2016). What does reproducibility really mean? Science Translational Medicine, 8(341), 1-6.

Reproducibility

Reproducibility

Reproducibility	Definition
Methods	Same data and same methods
Results	Independent data (same population) and same methods
Inferential	Same or independent data and same methods

This class will focus on **Methods Reproducibility**.

"Methods reproducibility refers to the provision of enough detail about study procedures and data so the same procedures could, in theory or in actuality, be exactly repeated," (Goodman et al., 2016).

Methods Reproducibility

"Methods reproducibility refers to the provision of enough detail about study procedures and data so the same procedures could, in theory or in actuality, be exactly repeated," (Goodman et al., 2016).

Requires

- A replicable workflow.
- Avoidance of input/output errors.
- Documentation, documentation, documentation.
- Transparent reporting.
- A single raw data set from which all analyses are done.
- Code (in this class, `R` code!) that works with the raw data.
- Others?

Methods Reproducibility and Rmarkdown

- A replicable workflow.
- Avoidance of input/output errors.
- Documentation, documentation, documentation.
- Transparent reporting.
- A single raw data set from which all analyses are done.
- Code (in this class, `R` code!) that works with the raw data.
- Others?

Each can be aided by `rmarkdown` and its associated packages.

We will demonstrate most of these throughout this course, but especially while we discuss `rmarkdown`.

Benefits of rmarkdown

1. Input (code) and output (tables/figures) are intimately tied---NO MANUAL (ERROR-PRONE) TABLES!

2. Document type is flexible and easy to adjust.

3. Updates to the analyses and output are straightfoward---no need to keep track of 50+ little things to change.

4. Provides quick reports that can be shared across many platforms (Word, Latex, PDF, HTML).

5. Makes transparent reporting easier---the actual steps of analysis are right next to the write up about the actual steps.

6. If we use version control (e.g. *Git*), the file system is much simpler.

7. It's cool.

Integrating rmarkdown into your workflow

If:

- You are reporting findings
- The analyses are iterative (report and re-analyze, report and re-analyze...)
- Working with individuals that understand some Latex or Markdown

Use Rmarkdown for the writing and reporting!

It'll probably stretch you a bit as a researcher -- but that is a good thing. 🤗

Get one started!

...

The Ins-and-Outs of rmarkdown

RStudio has a great website that we will use to understand rmarkdown better.