R-eproducible-science

Rick Gilmore

2017-08-11 15:26:27

Contents

R-eproducible psychological science

Themes

1. What is reproducible psychological science?
2. How can R make my science more transparent, open, and reproducible?

Is there a crisis?

Baker 2016

Not just in psychology

If so, why?

Baker 2016

Here are the data from the Nature survey.

What am I trying to reproduce?

- My own workflow
  - Data collection

  - Cleaning
  - Visualization
  - Analysis
- "Hit by a truck" scenario

Reproducible workflows

- Scripted, automated = minimize human-dependent steps.
- Well-documented
- Transparent to me & colleagues == transparent to others

Using R for reproducible workflows

- Mix R code, output, comments, tables using R Markdown
- R Markdown files = text files
- One input file, multiple outputs to
  - PDF, Word (.docx)
  - HTML for web pages, slides

Example 1

- James' R commands from Day 1
- Raw R script (.R)
- Converted to R Markdown
- Output as | HTML notebook | HTML Slides | PDF | DOCX |

How to

- Add header info in YAML Ain't Markup Language (YAML) format
- Wrap R code "chunks" with triple backticks and {r}
- Separate segments with `---` and/or `##` or `###`
- Render via `knit` button or 'rmarkdown::render(file="my-file.Rmd")

How to

- Create new R Markdown file: `New/New File/R Markdown...`
- Specify default, alternative output formats:
  - `pdf_document`
  - `word_document`
  - `ioslides_document`: HTML slides
  - `github_document`: renders nicely on GitHub

- Create your document
  - Use an "outline" with Header_1, Header_2, Header_3, etc.
  - Header_1 text starts with `# This is a top level header`
  - Header_2 text starts with `## This is a 2nd level header`
  - Header_3 text starts with `### This is a 3rd level header`
- Surround R code with triple back-ticks

- Sections that start with Header_1, Header_2, and `---` will start new slides in `ioslides_presentation` mode.
- **Bold text**: `**This is bold**`; *Italicized text*: `*Italics*`
- Start lists with hyphens `- Item 1` or numbers `1. Item 1`.

Let's try it

- bootcamp-survey.Rmd
- bootcamp-survey.md

Key points

- Use R Markdown files for documents, reports, presentations.
  - One or more output formats from the same file.
  - Analysis/lab notebook.
- Use R scripts to automate different pieces of the pipeline.
- Make README files to explain how to put pieces together.

Toward a reproducible psychological science...

- Transparent, reproducible, open workflows pre-publication
- Openly shared materials + data + code
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, N. P. du, Simonsohn, U., et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021. Retrieved January 10, 2017, from http://www.nature.com/articles/s41562-016-0021.
- Gilmore, R. O., & Adolph, K. E. (2017). Video can make behavioural science more reproducible. *Nature Human Behavior*, *1*. Retrieved from http://dx.doi.org/10.1038/s41562-017-0128.

Advanced topics

- Use R Studio projects
- Version control with git and GitHub
- Web sites, blogs, (even books) with R Markdown
- Scriptable analysis workflows
  - Reports for each participant
  - Example: PEEP-II project

My GitHub workflow

1. Create a repo on GitHub
2. Copy repo URL
3. `File/New Project.../`
4. Version Control, Git
5. Paste repo URL
6. Select local name for repo and directory where it lives.
7. Open project within R Studio `File/Open Project...`
8. Commit early & often

Learn from my mistakes

- Script **everything** you possibly can
  - If you have to repeat something, make a function or write a parameterized script
- Document **all the time**
  - Comments in code
  - Update README files
- Don't be afraid to ask
- Don't be afraid to work in the open
- Learn from others
- Just do it!