R-eproducible-science

Rick Gilmore

2017-08-15 10:54:41

Contents

R-eproducible psychological science

Themes

1. Why worry about reproducible science?
2. What is reproducible psychological science?
3. 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.

(Munafò et al. 2017) manifesto

This recent manifesto from Nature Human Behavior describes the risks to reproducible science at every step of the process. I urge you to read it.

What am I trying to reproduce?

- My own workflow
  - Data collection
  - Cleaning
  - Visualization
  - Analysis
- "Hit by a truck" scenario

But today I want us to think more parochially about our own workflows. How can using R make our own data collection, cleaning, visualization, and analysis workflows more reproducible. Put it this way: If you were hit by a truck tomorrow, could your adviser and collaborators pick up where you left off?

Reproducible workflows

- Scripted, automated = minimize human-dependent steps.
- Well-documented
- Be kind to your future (forgetful) self
- Transparent to me & colleagues == transparent to others

Reproducible workflows are scripted. They minimize human contact with your data files. They are well-documented. And it turns out that workflows that are transparent to you and your colleagues are transparent to others. This makes them easy to share.

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

We've already shown you in this bootcamp how writing R scripts and functions can let you import, clean, munge, reorganize, plot, and analyze data. We've already seen how commenting code fragments makes it easier to read and understand. An extension to R called R Markdown lets us mix R code, analyses, text, tables, and other formatting to make all sorts of products. R Markdown files are just text files. But with this one text file, it's easy to produce multiple output types: PDF or Word formatted documents; HTML for blogs, web sites, or even slide presentation.

Example 1

- James' R commands from Day 1
- Raw R script (.R)
- Converted to R Markdown

- Output as | HTML notebook | HTML Slides | PDF | DOCX |

Just to show you how easy this is, let's look at the R syntax James used yesterday. I'm going to show you how adding just a tiny bit of text to that file transforms it. Here is the original R script. Here is the transformed file with a .Rmd extension.

```
---
title: "R-Workshop-James"
author: "James LeBreton with Rick Gilmore"
date: '2017-08-15 10:54:41'
output:
  pdf_document: default
  html_notebook: default
  html_document: default
---
```

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")

```
---
title: "R-Workshop-James"
author: "James LeBreton with Rick Gilmore"
date: '2017-08-15 10:54:41'
output:
  pdf_document: default
  html_notebook: default
  html_document: default
---
```

- pdf_document
- github_document
- html_notebook
    - Like jupyter notebooks

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`.

```
# Big idea

## Smaller idea in service of bigger

- Supporting point
- Another suppporting point
- a **bold** point
- an *italicized* point
- a [link](http://psu-psychology.github.io/r-bootcamp) to this bootcamp
- an image: ![rawr](https://www.insidehighered.com/sites/default/server_files/media/PennState2.PNG)
```
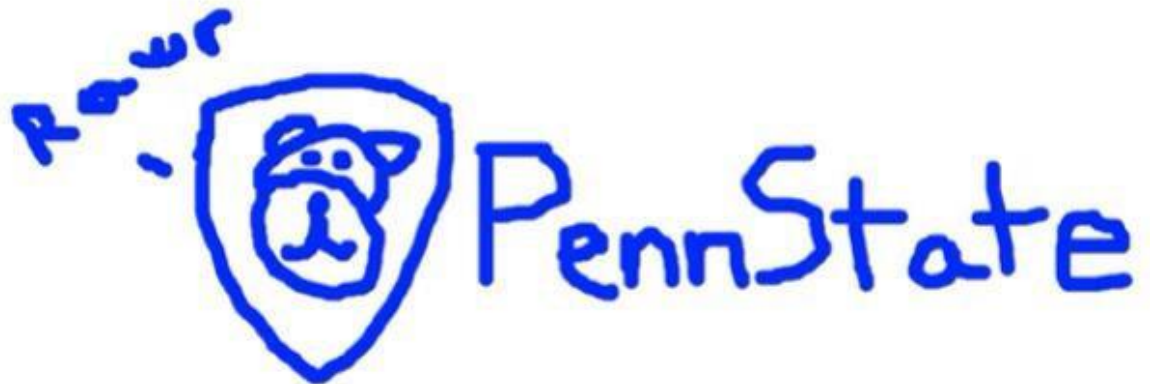
Big idea

Smaller idea in service of bigger

- Supporting point
- Another suppporting point
- a **bold** point
- an *italicized* point
- a link to this bootcamp

- an image:

Let's try it

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

One file, many output options

- **PDF document**: rmarkdown::render('talks/bootcamp-survey.Rmd', output_format = "pdf_document")
- **Word document**: rmarkdown::render('talks/bootcamp-survey.Rmd', output_format = "word_document")
- GitHub document: rmarkdown::render('talks/bootcamp-survey.Rmd', output_format = "github_document")
- **HTML slides**: rmarkdown::render('talks/bootcamp-survey.Rmd', output_format = "ioslides_presentation
- **Many outputs**: rmarkdown::render('talks/bootcamp-survey.Rmd', output_format = c("pdf_document", "word_document", "github_document", "ioslides_presentation")

Scripting the pipeline

```
# Get_bootcamp_googlesheet.R
#
# Script to authenticate to Google, extract R bootcamp survey data

library(googlesheets)
library(tidyverse)


survey_url <- "https://docs.google.com/spreadsheets/d/1Ay56u6g4jyEEdlmV2NHxTLBlcjI2gHavta-Ik0kGrpg/edit

bootcamp_by_url <- survey_url %>%
  extract_key_from_url() %>%
  gs_key()

bootcamp_sheets <- gs_ws_ls(bootcamp_by_url)
```

```
boot_data <- bootcamp_by_url %>%
  gs_read(bootcamp_sheets[1])

names(boot_data) <- c("Timestamp",
                      "R_exp",
                      "GoT",
                      "Age_yrs",
                      "Sleep_hrs",
                      "Fav_date",
                      "Tidy_data")

write_csv(boot_data, path = "data/survey.csv")
```

```
# Update_survey.R
#
# Updates Googlesheet survey data and generates new R Markdown report
#

source("R/Get_bootcamp_googlesheet.R")
rmarkdown::render("talks/bootcamp-survey.Rmd",
                  output_format = c("github_document",
                                    "pdf_document",
```

```
                            "word_document",
                            "ioslides_presentation"))
```

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 or functions 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 et al. 2017)
- (Gilmore and Adolph 2017)

Advanced topics

- Write papers in R Markdown using `papaja`
- 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
  - This bootcamp's `Make_site.R`

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!

References

Gilmore, Rick O, and Karen E Adolph. 2017. "Video Can Make Behavioural Science More Reproducible." *Nature Human Behavior* 1 (12~jun). doi:10.1038/s41562-017-0128.

Munafò, Marcus R, Brian A Nosek, Dorothy V M Bishop, Katherine S Button, Christopher D Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J Ware, and John P A Ioannidis. 2017. "A Manifesto for Reproducible Science." *Nature Human Behaviour* 1 (10~jan): 0021. doi:10.1038/s41562-016-0021.