R-eproducible Psychological Science

Rick Gilmore

2017-08-17 12:48:31

## Themes

1. Is there a reproducibility crisis?
2. What is reproducible psychological science?
3. How can R make my science more transparent, open, and reproducible?

## Is there a reproducibility crisis?

* Yes, a significant crisis
* Yes, a slight crisis
* No crisis
* Don't know

[Baker 2016](http://doi.org/10.1038/533452a)

## Not just in psychology

[Baker 2016](http://doi.org/10.1038/533452a)

Here are the data from the Nature survey.

## [(Munafò et al. 2017)](http://doi.org/10.1038/s41562-016-0021) 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
	+ Reporting
	+ Manuscript generation?
* "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. Ask yourself this: Can you pick up where you left off on a project you were working on yesterday? Last week? Last month? Six months ago? 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

* **Option 1**: All commands in an R script: e.g., project\_analysis.R
* **Option 2a**: Mix R code, output, comments in an [R Markdown](http://rmarkdown.rstudio.com/) document
* **Option 2b**: Use R scripts with some [special formatting](http://rmarkdown.rstudio.com/articles_report_from_r_script.html), [(more info)](https://github.com/jennybc/happy-git-with-r/blob/master/31_workflow-first-use-r-script-and-github.Rmd).

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

# Import data

# Clean data

# Visualize data

# Analyze data

# Report findings

# Import data
my\_data <- read.csv("path/2/data\_file.csv")

# Clean data
my\_data$gender <- tolower(my\_data$gender) # make lower case
...

## Make script that calls sequence of R commands or functions

# Import data
source("R/Import\_data.R") # source() runs scripts, loads functions

# Clean data
source("R/Clean\_data.R")

# Visualize data
source("R/Visualize\_data.R")
...

## Strengths & Weaknesses

* R commands in files that can be re-run
* Separate pieces of workflow kept separate
* "Master.R" script that can be run to regenerate full sequence of results
	+ Error in raw data file?
	+ No problem; fix and re-run "Master.R"
* How to save results or share with collaborators?

## Example 2 - R Markdown

* James' R commands from Day 1: [Raw R script (.R)](R-Workshop-James.R)
* Converted to [R Markdown](R-Workshop-James.Rmd)
* Output as | [HTML notebook](R-Workshop-James.nb.html) | [HTML Slides](R-Workshop-James.html) | [PDF](R-Workshop-James.pdf) | [docx](R-Workshop-James.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.

## Structure of an [R Markdown .Rmd](R-Workshop-James.Rmd) file

* header info in [YAML Ain't Markup Language (YAML) format](http://www.yaml.org/start.html)
* Markdown for formating text (headers, **boldface**/*italics*, code, bulleted or numbered lists, [web links](http://www.psu.edu), etc.
* R code "chunks"

## One R to rule them all and in the console bind them...

* One file, many possible outputs
	+ [pdf\_document](http://rmarkdown.rstudio.com/pdf_document_format.html), [word\_document](http://rmarkdown.rstudio.com/word_document_format.html), or [github\_document](http://rmarkdown.rstudio.com/github_document_format.html)
	+ [ioslides\_presentation](http://rmarkdown.rstudio.com/ioslides_presentation_format.html) for HTML slide show
	+ Cool interactive web-app like Dan's tutorial
	+ Web sites like the one for this [bootcamp](https://github.com/psu-psychology/r-bootcamp), [blogs](https://bookdown.org/yihui/blogdown/), even [books](https://bookdown.org/yihui/bookdown/)

## Your turn

1. Open "File/New File/R Notebook"
2. Change title: "R Notebook" to something else, like title: "Rick's R Notebook"
3. Save the file (default name is Untitled) with an .Rmd extension.
4. Look at the \*.Rmd code.
5. Look at the \*.nb.html file in a browser.

## Things to try if you like

# Big idea

## Smaller idea in service of bigger

- Supporting point
- Another suppporting point

1. an enumerated \*\*bold\*\* point
1. an enumerated \*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)
- an equation: $e = mc^2$

# 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: 
* an equation: $e=mc^{2}$

## Let's try it with some data

* <bootcamp-survey.Rmd>
* <bootcamp-survey.md>

## One file, many output options

* 'Default' for the file: rmarkdown::render("talks/bootcamp-survey.Rmd")
* **PDF document**: rmarkdown::render('talks/bootcamp-survey.Rmd', output\_format = "pdf\_document")
* **Word document**: rmarkdown::render('talks/bootcamp-survey.Rmd', output\_format = "word\_document")
* **HTML slides**: rmarkdown::render('talks/bootcamp-survey.Rmd', output\_format = "ioslides\_presentation")
* **Multiple outputs**: rmarkdown::render('talks/bootcamp-survey.Rmd', output\_format = c("pdf\_document", "word\_document", "github\_document", "ioslides\_presentation")

## Key points

* Use R scripts to capture & reproduce workflows and/or
* 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)](http://doi.org/10.1038/s41562-016-0021): reproducible practices across the workflow
	+ Where to share and when? Lots of options. Let's talk.
* [(Gilmore and Adolph 2017)](http://dx.doi.org/10.1038/s41562-017-0128): video and reproducibility

## Advanced topics

* Write papers in R Markdown using [papaja](https://github.com/crsh/papaja)
	+ Make [this](../papaja_demo/gilmore-lebreton-hallquist.pdf) from [this](../papaja_demo/gilmore-lebreton-hallquist.Rmd)
* Use R Studio [projects](https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects)
* Version control with git and [GitHub](http://github.com)
* Scriptable analysis workflows
	+ Reports for each participant, e.g. [PEEP-II project](https://github.com/gilmore-lab/peep-II-ratings-analysis)
	+ This bootcamp's [Make\_site.R](../R/Make_site.R)
* [Web sites](http://rmarkdown.rstudio.com/rmarkdown_websites.html), [blogs](https://bookdown.org/yihui/blogdown/), (even [books](https://bookdown.org/)) with R Markdown

## R Studio Projects

* Keep files, settings, organized
* Easy to switch between projects
* Reduces mental effort (what directory am I in?)
* Integrates with version control (e.g., GitHub)

## Version control

* Keep track of your past
* Back to the Future
* **git**: a system for software version control
* [**GitHub**](http://github.com): a website for managing projects that use git

## 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

 Your browser does not support the video tag.

 Your browser does not support the video tag.

## 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?usp=sharing"

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"))

## [Web sites](http://rmarkdown.rstudio.com/rmarkdown_websites.html)

* [\_site.yml](../_site.yml): site configuration parameters
* [index.Rmd](../index.Rmd): home page for site
* other \*.Rmd files: other pages
* other directories for files
* rmarkdown::render\_site()
* [GitHub pages](https://pages.github.com/) or other web site hosting service

## 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](https://doi.org/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](https://doi.org/10.1038/s41562-016-0021).