UPPSALA
UNIVERSITET

Uppsala Child & Baby Lab

Reproducible Research Reports i R

R for Psychology Research

Marcus Lindskog, Docent
2019-11-03

Overview

1. Not today!!

Problem

Questions

- How long would it take you to reproduce the results in your latest paper from raw data?
- If a reviewer asked you to add/drop a variable/condition in your analyses, how long would it take to get the new results section ready (including figures)?
- Could an independent researcher reproduce your results if she/he had your raw data and your paper?

Reproducible Research

- If you feel a bit uncomfortable answering the previous questions, reproducible research reports might be for you.
- **Reproducible Research**: A workflow that integrates data and analyses in a way that makes it possible for an independent researcher to recreate the finding in a relatively straightforward way.
- **Reproducible Research Reports (RRRs)** : A report (e.g., your manuscript) that also integrates writing the report into the reproducible research workflow.

The Case for RRRs

1. Adopting a workflow with RRRs increases transparency of your work. Others can easily see what you did.
2. Collaborators can easily double-check the results.
3. Producing RRRs makes it easy to revise and update analyses and visualizations based on feedback from, e.g., reviewers.
4. RRRs maintain a close coupling between data, analyses, and report. You will like that in 10 years time.
5. Using RRRs makes it easy to incorporate the same output from your analyses in a variety of formats (papers, presentations, posters, etc.)

Why is preregistration and/or sharing data not enough?

- Preregistration and/or sharing data is GREAT!!
- But, there is a long way from raw data to final paper.
- With RRRs you can show exactly how you choose to walk down that path and make all of your decisions explicit.

Possible Solutions

The Cumbersome way

- Write down in detail exactly all the steps you take from raw data to the finished Results section.
- Be prepared to share these notes with a reviewer/independent researcher.
- This is a good first step but takes a lot of time. Both to do at first and to use when reproducing.

The Middle way

- Use software like **R**, **SPSS**, or **MATLAB** to create scripts that take you from raw data to results.
- This even better. Someone else can take your script and your data and reproduce your results, and all of your analysis decisions are available from the code.
- If you use **R** or **MATLAB** the scripts can also output all your figures, close to how they would appear in the paper.
- **Downside:** You still have to assemble your report, which might lead to "copy-paste-mistakes" and other issues.

The Full Monty

1. Structure your projects in a consistent way on your computer.
2. Analyse your data by writing scripts in **R**.
3. Write your paper in *RStudio* using *R Markdown* and incorporate your analyses directly into your paper.
4. When your paper is finished, make your entire project publicly available (within the limits of your ethics approval) on, e.g., osf.io.
5. (Optional) Use version control to keep track of all changes you make, both to your analyses and to your paper.

But...

What could be possible reasons for not trying out RRRs?

- You feel that Word is God's gift to man and can't consider ever writing in any other software.
- You enjoy spending hours trying to get **SPSS** to output within-subjects error bars on your split plot.
- You think that the best way to spend your Sunday is to figure out why you excluded participant 31 but not 38 form your analyses.

R Markdown

What is R Markdown?

- A unified authoring framework that combines:
  - code
  - results
  - text.
- Combines various tools like, R, markdown, and pandoc to go from code to reports in wide range of formats (e.g., Word, PDF, etc.)
- R Markdown is available in the RStudio IDE.

Three types of content

A *YAML* header

```
---
title: "Untitled"
author: "Marcus Lindskog"
date: '2019-11-03'
output: pdf_document
---
```

```
---
title: "Reproducible Research Reports i R"
subtitle: R for Psychology Research
author: "Marcus Lindskog, Docent"
date: "2019-11-03"
output:
  xaringan::moon_reader:
    lib_dir: libs
    css: ["rutgers", "marcus_uu.css"]
    nature:
      highlightStyle: github
      highlightLines: true
      countIncrementalSlides: false
      beforeInit: "../templates/macros.js"
---
```

Chunks of R code

- A code chunk starts with three backticks like ```` ```{r} ```` where `r` indicates the language name and ends with three backticks. You can write chunk options in the curly braces (e.g., set the figure height to 5 inches: ```` ```{r, fig.height=5} ````).

````
```{r}
fit = lm(dist ~ speed, data = cars)
b   = coef(fit)
plot(cars)
abline(fit)
```
```
````
````

Inline R code

- An inline R code expression starts with `` `r `` and ends with a backtick `` ` ``.

```
x <- 5
```

```
For a circle with the radius `r x`,
its area is `r pi * x^2`.
```

- For a circle with the radius 5, its area is 78.5398163.

```
The formula for the area is $A = \pi \cdot r^2$.
For a circle with the radius `r x`,
its area is `r pi * x^2`.
```

- The formula for the area is $A = \pi \cdot r^2$. For a circle with the radius 5, its area is 78.5398163.

Chunk options

Option	Run code	Show code	Output	Plots	Messages	Warnings
`eval = FALSE`	-		-	-	-	-
`include = FALSE`		-	-	-	-	-
`echo = FALSE`		-				
`results = "hide"`			-			
`fig.show = "hide"`				-		
`message = FALSE`					-	
`warning = FALSE`						-

New notebook and Compile

- Open a new *.Rmd* file from RStudio and choose your document type.
- To complie the notebook pres *Knit* or Cmd/Ctrl-Shift-K

Formating

- Text

```
> *italic* or _italic_
> **bold** or __bold__
> `code`
```

- Headers

```
> # 1st level
> ## 2nd level
> ### 3rd level
```

Tables

```
Header A| Header B
--------|--------
Cont. A1| Cont. B1
Cont. A2| Cont. B2
```

Header A	**Header B**
Cont. A1	Cont. B1
Cont. A2	Cont. B2

- You can also produce nice looking tables using `kable` from the `knitr` package. Or functions from the `kableExtra` package.

Bibliographies and Citations

- You might want to do some citations in your reports.
- Then you need your bibliography in a `.bib` file. Most citation managers will produce that for you.

```
bibliography: rmarkdown.bib
csl: apa.csl
```

```
Separate multiple citations with a `;`:
Blah blah [@smith04; @doe99].

You can add arbitrary comments inside the square brackets:
Blah blah [see @doe99, pp. 33-35; also @smith04, ch. 1].

Remove the square brackets to create an in-text citation:
@smith04 says blah, or @smith04 [p. 33] says blah.

Add a `-` before the citation to suppress the author's name:
Smith says blah [-@smith04].
```

An example

```
x <- rnorm(100, .6, .2)
t_test_res <- t.test(x, mu = .5)
```

- A single sample t-test revealed that the mean *x* (*M* = 0.589) was significantly different (*t*(99) = 4.36, *p* = 0) from .5.

```
A single sample t-test revealed that the mean *x*
(*M* =  `r round(mean(x), 3)`) was
significantly different (*t*( `r t_test_res$parameter` ) =
`r round(t_test_res$statistic,2)`, *p* =
`r round(t_test_res$p.value,3)`) from .5.
```

Packages to help you...

- `papaja`: Reproducible APA manuscripts with R Markdown (https://crsh.github.io/papaja_man/)
- `kableExtra`: To get much more flexibility in formatting tables.
- `xaringan`: To make nice and easy presentations (https://github.com/yihui/xaringan/wiki). All presentations for this course was produced with xaringan.
- `citr`: So you get bibliographies into your markdown and can easily cite.

That's all folks!