

Reproducible Research

Andrew Caines | ALTA-DTAL | apc38

23 February 2015

Overview

- reproducible research

Overview

- reproducible research
- case studies

Overview

- reproducible research
- case studies
- *break*

Overview

- reproducible research
- case studies
- *break*
- R Markdown

What is Reproducible Research?

- available data

What is Reproducible Research?

- available data
- available, functioning code

What is Reproducible Research?

- available data
- available, functioning code
- Reproducibility movement: e.g. Political Science Replication

Why reproducible research?

- incremental Science

Why reproducible research?

- incremental Science
- collaborate with anyone anywhere

Why reproducible research?

- incremental Science
- collaborate with anyone anywhere
- avoid scandals: don't end up in Retraction Watch

Why reproducible research?

- incremental Science
- collaborate with anyone anywhere
- avoid scandals: don't end up in Retraction Watch
- be good and keep up / go with the flow: e.g. –

Why reproducible research?

- incremental Science
- collaborate with anyone anywhere
- avoid scandals: don't end up in Retraction Watch
- be good and keep up / go with the flow: e.g. –
- Psychology

Why reproducible research?

- incremental Science
- collaborate with anyone anywhere
- avoid scandals: don't end up in Retraction Watch
- be good and keep up / go with the flow: e.g. –
- Psychology
- SocPsych

Why reproducible research?

- incremental Science
- collaborate with anyone anywhere
- avoid scandals: don't end up in Retraction Watch
- be good and keep up / go with the flow: e.g. –
- Psychology
- SocPsych
- PolSci

Why reproducible research?

- incremental Science
- collaborate with anyone anywhere
- avoid scandals: don't end up in Retraction Watch
- be good and keep up / go with the flow: e.g. –
- Psychology
- SocPsych
- PolSci
- CompSci

How to achieve reproducibility?

- paper + supplementary information / external website

How to achieve reproducibility?

- paper + supplementary information / external website
- e.g.1: independent repository: Eckhoff, Nasset & Janda

How to achieve reproducibility?

- paper + supplementary information / external website
- e.g.1: independent repository: Eckhoff, Nasset & Janda
- e.g.2: dedicated websites: Dodds et al, + GitHub repo

How to achieve reproducibility?

- paper + supplementary information / external website
- e.g.1: independent repository: Eckhoff, Nasset & Janda
- e.g.2: dedicated websites: Dodds et al, + GitHub repo
- e.g.3: personal website + independent repository: Martijn Wieling

Reproducibility (and research) tools

- GitHub, Dropbox, institutional repositories, your webpages

Reproducibility (and research) tools

- GitHub, Dropbox, institutional repositories, your webpages
- Open Science Framework

Reproducibility (and research) tools

- GitHub, Dropbox, institutional repositories, your webpages
- Open Science Framework
- Plotly

Reproducibility (and research) tools

- GitHub, Dropbox, institutional repositories, your webpages
- Open Science Framework
- Plotly
- Shiny

A word of warning

- Nature article by Mina Bissell, 2013

Intermission

- questions?

Intermission

- questions?
- next: R Markdown

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R
- markdown = plain-text authoring syntax, increasingly popular (e.g. GitHub)

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R
- markdown = plain-text authoring syntax, increasingly popular (e.g. GitHub)
- R Markdown = implementation of markdown, can process output from R

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R
- markdown = plain-text authoring syntax, increasingly popular (e.g. GitHub)
- R Markdown = implementation of markdown, can process output from R
- uses `knitr` package (successor to Sweave) and 'Pandoc' text file conversion program

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R
- markdown = plain-text authoring syntax, increasingly popular (e.g. GitHub)
- R Markdown = implementation of markdown, can process output from R
- uses `knitr` package (successor to Sweave) and 'Pandoc' text file conversion program
- can produce HTML, PDF and Word documents

What is R Markdown?

- **simple syntax:** shallow learning curve

What is R Markdown?

- **simple syntax**: shallow learning curve
- **human-readable**: emphasis on communication and clarity, rather than technique and complexity

What is R Markdown?

- **simple syntax**: shallow learning curve
- **human-readable**: emphasis on communication and clarity, rather than technique and complexity
- **transparent formatting**: wysiwyg

What is R Markdown?

- **simple syntax**: shallow learning curve
- **human-readable**: emphasis on communication and clarity, rather than technique and complexity
- **transparent formatting**: wysiwyg
- **embedded computation**: share with collaborators, reviewers, colleagues, students, critics

What is R Markdown?

*... integrating source code, statistical output, and text in R Markdown is **a model of reproducibility**.*

Such transparency facilitates comprehension, defensibility, and further research or testing.

R Markdown helps to bring the vision for reproducibility in statistical analysis articulated by Gentleman & Temple Lang [2004] to reality.

- Udwin & Baumer 2015, 'R Markdown', arXiv.org

What is R Markdown?

This vision — in which the barriers to verify another's statistical computations from start to finish are low — is the intellectual descendant of Claerbout [1994], Buckheit & Donoho [1995], and began with Knuth [1984], the creator of TEX.

Moreover, R Markdown is dynamic: each time the document is rendered, the commands therein are run anew. If data are altered or different data are called in advance of rendering, then the output is dependent and calculated accordingly.

- Udwin & Baumer 2015, 'R Markdown', arXiv.org

How to use R Markdown

installation

- `install.packages('rmarkdown')`

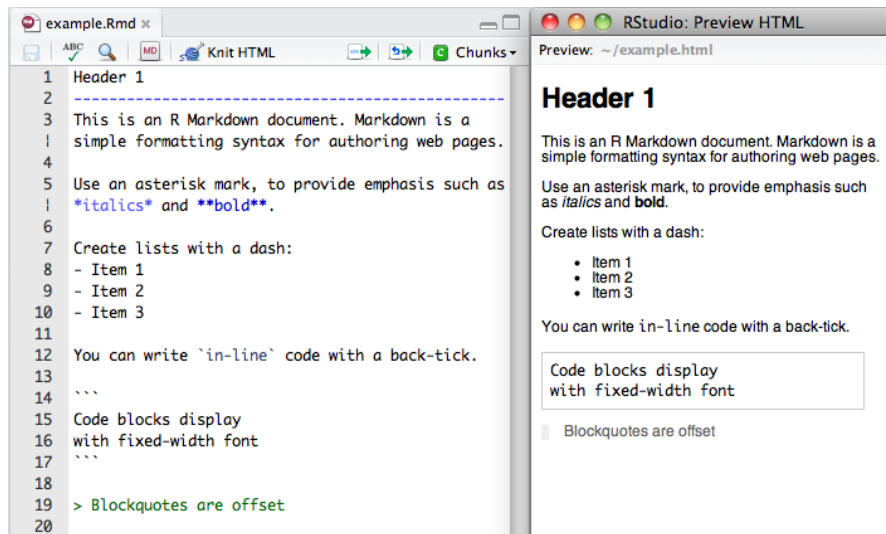
How to use R Markdown

installation

- `install.packages('rmarkdown')`
- `library(rmarkdown)`

How to use R Markdown

simple syntax, human-readable, transparent formatting



The screenshot displays the RStudio interface with two panes. The left pane shows the R Markdown source file 'example.Rmd' with line numbers 1 through 20. The right pane shows the HTML preview of the document.

Source File (example.Rmd):

```
1 Header 1
2 -----
3 This is an R Markdown document. Markdown is a
4 | simple formatting syntax for authoring web pages.
5
6 Use an asterisk mark, to provide emphasis such as
7 | italics and bold.
8
9 Create lists with a dash:
10 - Item 1
11 - Item 2
12 - Item 3
13
14 You can write `in-line` code with a back-tick.
15
16 Code blocks display
17 | with fixed-width font
18 |
19 | > Blockquotes are offset
20
```

HTML Preview:

Header 1

This is an R Markdown document. Markdown is a simple formatting syntax for authoring web pages.

Use an asterisk mark, to provide emphasis such as *italics* and **bold**.

Create lists with a dash:

- Item 1
- Item 2
- Item 3

You can write in-line code with a back-tick.

```
Code blocks display
with fixed-width font
```

Blockquotes are offset

How to use R Markdown

embedded computation

chunks.Rmd x

ABC 🔍 NO Knit HTML ➡️ ⚙️ Chunks ▾

```
1 R Code Chunks
2 =====
3
4 With R Markdown, you can insert R code
5 chunks including plots:
6
7 ```{r qplot, fig.width=4, fig.height=3,
8   message=FALSE}
9 # quick summary and plot
10 library(ggplot2)
11 summary(cars)
12 qplot(speed, dist, data=cars) +
13   geom_smooth()
```

RStudio: Preview HTML

Preview: ~/chunks.html | Save As | Publish

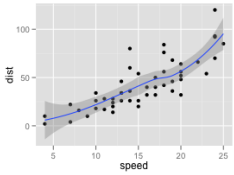
R Code Chunks

With R Markdown, you can insert R code chunks including plots:

```
# quick summary and plot
library(ggplot2)
summary(cars)
```

| ## | speed | dist |
|----|--------------|-------------|
| ## | Min. : 4.0 | Min. : 2 |
| ## | 1st Qu.:12.0 | 1st Qu.: 26 |
| ## | Median :15.0 | Median : 36 |
| ## | Mean :15.4 | Mean : 43 |
| ## | 3rd Qu.:19.0 | 3rd Qu.: 56 |
| ## | Max. :25.0 | Max. :120 |

```
qplot(speed, dist, data = cars) + geom_smooth()
```



R Markdown and Reproducibility

- example 1: Beamer presentation

R Markdown and Reproducibility

- example 1: Beamer presentation
- example 2: HTML output demo

R Markdown and Reproducibility

- example 1: Beamer presentation
- example 2: HTML output demo
- example 3: PDF output demo

R Markdown and Reproducibility

- example 1: Beamer presentation
- example 2: HTML output demo
- example 3: PDF output demo
- *n.b.* Tufte handouts

R Markdown: some useful links

- RStudio: R Markdown
- Udwin & Baumer arXiv.org paper
- Shiny & R Markdown
- my GitHub 'replication' repo

The End

- questions?
- contact me: apc38
- my website
- [[@cainesap](https://twitter.com/cainesap)](<https://twitter.com/cainesap>)