

Writing reproducible research with R and Markdown

Previously...

Marwick, B., & Birch, S. (2018) A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing. *Advances in Archaeological Practice* 1-19. <https://doi.org/10.1017/aap.2018.3>

Marwick, B. & 47 others (2017) Open science in archaeology. *SAA Archaeological Record*, 17(4), pp. 8-14.

Marwick, B. (2017). Computational reproducibility in archaeological research: Basic principles and a case study of their implementation. *Journal of Archaeological Method and Theory*, 24(2), 424-450. preprint at <https://osf.io/preprints/socarxiv/q4v73/>

Marwick, B., Boettiger, C., & Mullen, L. (2017). Packaging data analytical work reproducibly using R (and friends). *The American Statistician* 70 (1) : 80-88 <https://doi.org/10.1080/00031305.2017.1375986>

Eglen, S. J., Marwick, B., Halchenko, Y. O., et al. (2017). Toward standard practices for sharing computer code and programs in neuroscience. *Nature Neuroscience*, 20(6), 770-773. preprint at <https://www.biorxiv.org/content/early/2017/02/28/045104>

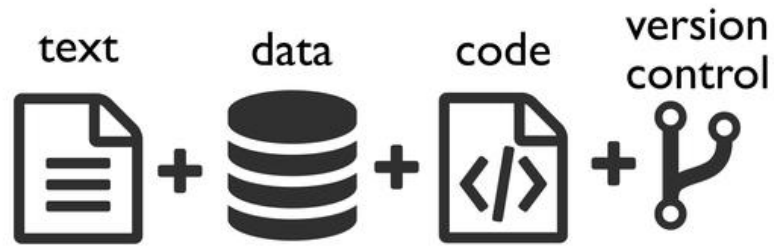
And outside of the ivory tower...

Marwick, B., & Jacobs, Z. (2017). Here's the three-pronged approach we're using in our own research to tackle the reproducibility issue. *The Conversation*.

<https://theconversation.com/heres-the-three-pronged-approach-were-using-in-our-own-research-to-tackle-the-reproducibility-issue-80997>

Marwick, B. (2015). How Computers Broke Science—and What We Can Do to Fix It. *The Conversation*. <https://theconversation.com/how-computers-broke-science-and-what-we-can-do-to-fix-it-49938> translated into French, German, Japanese & Chinese

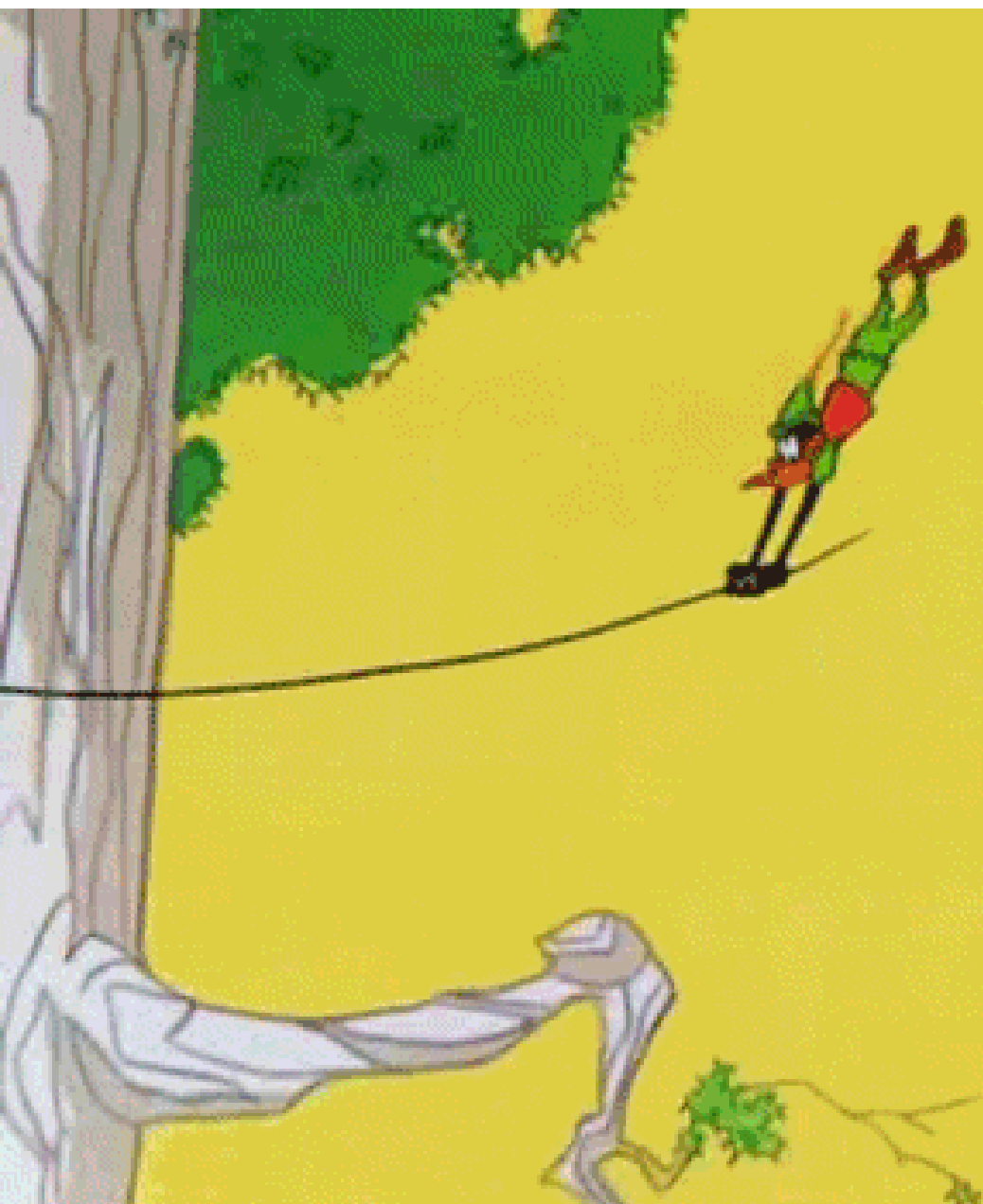
advertising:
text & final
results only



science:
text, code &
data available,
linked & licensed



Adapted with permission from Rodríguez-Sánchez F, Pires-Louze AJ, Barcena L, Varela S (2016) Ciencia reproducible: qué, por qué, cómo. Ecosistemas, 25(2): 83-92. <http://dx.doi.org/10.7818/EECOS.2016.25-2.11>.
See also Marwick, B. (2016). Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. Journal of Archaeological Method and Theory 23(2): 1-27.
<http://dx.doi.org/10.1007/s10816-015-9272-9>. This figure is CC-BY.



How to make this as quick & easy as possible?

💡 Use functions instead of copy-pasting

💡 Use sensible templates & default options

💡 Hide details we don't care much about

A solution: the rrttools package

The aim is to provide instructions, templates, and functions for making a basic compendium suitable for doing reproducible research with R

From zero to reproducible in 5+ steps (minutes)

Ensure you have the current versions of R & RStudio, the development tools (Xcode or Rtools), open RStudio, set your working directory to your desktop, and then:

Install from GitHub (eventually CRAN too):

```
# install.packages("devtools")  
devtools::install_github("benmarwick/rrtools")
```


1. Create a basic R package

This is `devtools::create()` with an additional step to either start the project in RStudio, or set the working directory to the package location, if not using RStudio:

```
rrtools::use_compendium("myprojname") # use a meaningful name :)
```

Why? Using a widely recognised file structure makes it easier for you to navigate, and easier for others to browse and inspect your work.

2. Add a license for our code

This adds a reference to the MIT license in the DESCRIPTION file and generates a LICENSE file listing the name provided as the copyright holder:

```
usethis::use_mit_license(name = "My Name")
```

To use a different license, replace this line with `usethis::use_gpl3_license(name = "My Name")`, or follow the instructions for other licenses

Why? To make it clear that you're happy for others to use your code, and that you're not responsible if they have problems.

3. Use version control for tracking changes and collaboration

Assumning we have Git installed, we go to the R console and introduce ourselves to Git:

```
usethis::use_git_config(user.name = "Ben Marwick",  
                        user.email = "benmarwick@hotmail.com")
```

And let's make sure we have an account at <https://github.com> to back-up our compendium and collaborate with our colleagues

3. Use version control for tracking changes and collaboration

If you are connected to the internet, this initializes a local git repository, connects to GitHub, and creates a remote repository:

```
usethis::use_git()  
# open up the GitHub panel to generate  
# your Personal Authorisation Token (PAT)  
usethis::browse_github_pat()  
# get a token from https://github.com/settings/tokens  
usethis::edit_r_environ()  
# Paste your copied PAT into your .Renviron file as system variable:  
# GITHUB_PAT=XXXXXX  
usethis::use_github(protocol = "https",  
                    private = FALSE)
```

You can also create a private repository. GitHub is what most people are using, but there other good options.

Why? To increase the transparency, efficiency and accessibility of your work.

4. Add instructions for other users and readers

This generates README.Rmd and renders it to README.md, ready to display on GitHub. It contains:

- a template citation to show others how to cite your project.
- license information for the text, figures, code and data in your compendium
- this also adds two other markdown files: a code of conduct for users, and basic instructions for people who want to contribute to your project

```
rrtools::use_readme_rmd()
```

Why? To improve the accessibility of your work by communicating with to others the purpose of your work, how you want it cited. To manage expectations for how others can interact with it.

5. Add compendium file structure and document templates

This creates a research compendium file structure and add template documents such as:

- an R Markdown file for writing a journal article
- a bib file for holding bibliographic citations
- a citation style file
- a template for styling MS Word output

```
rrtools::use_analysis()
```

Why? To write code and text in a reproducible document. To keep your work organised so you can be more efficient, and your project is easier for others to browse.

```
analysis/
├── paper/
│   ├── paper.Rmd          # this is the main document to edit
│   ├── references.bib      # this contains the reference list information
│   └── journal-of-archaeological-science.csl
│                               # this sets the style of citations
│                               # & reference list
├── figures/
├── data/
│   ├── raw_data/          # data obtained from elsewhere
│   └── derived_data/      # data generated during the analysis
└── templates
    ├── template.docx      # used to style the output of the paper.Rmd
    └── template.Rmd
```

You can exclude data/ from git, if you need to keep the data private.

If you're writing a PhD thesis, use [huskydown](#) to get PDF templates and a directory structure suitable for a UW PhD

Let's take R Markdown for a test drive

- 🔧 Code chunks & in-line code

- 🔧 Citations, captions & cross-references

- 🔧 Figures, plots & tables



The compendium plus!

```
rrtools::use_dockerfile()  
rrtools::use_travis()  
usethis::use_testthat()
```

Why? To avoid dependency complications and streamline trouble-shooting

What we have not done

We have not attempted to automate depositing the compendium in a repository.

```
# e.g. something like  
rrtools::deposit_compendium()
```

Why? Because it's not clear to us what the most sensible options are for most people. There are many trustworthy repositories you could use. You may not want to share some or all of your data or intermediate products.

So we leave this up to you, and encourage you to make available as much as possible.

To conclude

☑ We have many encoded best practices in getting organised

🕒 We have saved a lot of start-up time

♻️ We have a sustainable path to working reproducibly

Colophon

Presentation written in R Markdown using xaringan

Compiled into HTML5 using RStudio & knitr

Source code hosting: <https://github.com/benmarwick/>

ORCID: <http://orcid.org/0000-0001-7879-4531>

Licensing:

- Presentation: CC-BY-3.0
- Source code: MIT