Workflow Instructions

Matt Weldy

2020-09-09

Contents

List of Figures

List of Tables

1 Reproducible projects

Recent years have seen an explosion of tools available for use in academic and commecial applications. However, as the Peter Parker principle outlines, "with great power comes great responsibility." In general every scientist I've worked with does their best to operate other this principle. But, let's be honest, most of the readers of this guide are primarily ecologists, not computer scientists, or software engineers. We are in practice terrible hackers trying to tape together the tools we need to use in the best way we can.

Figure 1: The wisdom of Stan Lee. Source: https://www.walmart.com/ip/The-Amazing-Spider-Man-Framed-Marvel-Comics-Pop-Art-Poster-Art-Print-Quote-With-Great-Power-Comes-Great-Responsibility-Size-16-x-16/587466634.

The following guide is my attempt to share a workflow that I currently use. In full disclosure, none of this is original. I have put this together from a number of sources and have found it helps me stay organized, keeps my work portable (i.e. switching computers or passing code to coauthors does not cause a big headache), and helps track and archive changes throughout each project's lifetime. This guide will continue to evolve as I adapt different methods, and I welcome input from others along the way.

For example:

- Croucher et al. (2017)
- Staff (n.d.)
- Harrison (2015)

All this begets the question, why worry about reproducibility and optimizing work flows?

The short anwser is, do it for your future self. The long anwser is, well, long and mostly outside of this articles scope. Many articles, books, and blogposts have outlined the importance of reproducibility in science (see:). There arguments point towards reliable and reproducible results, extension of local results to general theories, avoiding biased inferences, repdoducibility in peer review. These reasons are laudable and important. However, for your own sanity, picking an easy to manage reproducible workflow will slow the growth rate of your bald spot. Analyses typically take months and sometimes years to complete. Personally, I can hardly remember what I had for lunch yesterday, so it is important to me that I can open a research project, after nearly any length of time, and continue working. The goal of this workflow is to demonstrate

a workflow that is insensitive to exploding computers, lost hard drives, sloppy collaborators, and prolonged peer reviews.
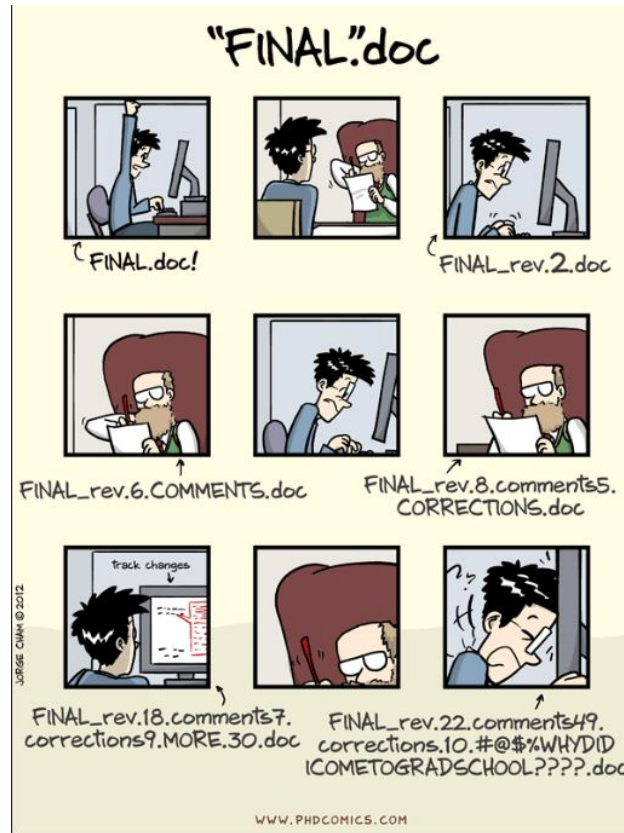
2 Link Rstudio and Github

Figure 2: This is what many of our folder directories look like. Source: "Piled Higher and Deeper" by Jorge Cham on www.phdcomics.com.

```
## Warning: package 'knitr' was built under R version 3.6.3
```

First things first. What is this Git thing I keep hearing about and how do I use it?

Using SSH keys

3 Initialize a new project

4 git ignore

5 Working with a project

This is an RMarkdown document intended to outline my, in progress, solution to juggling project directory madness. Working through this document should outline the procedures I use to:

Figure 3: This is what many of our folder directories look like. Source: Piled Higher and Deeper by Jorge Cham on www.phdcomics.com.

1. Connect Rstudio to github
2. Initialize a new project

- Including importing my project directory template
- Daily workflow
- Daily work documentation procedures

```r
plot(1:12)
```


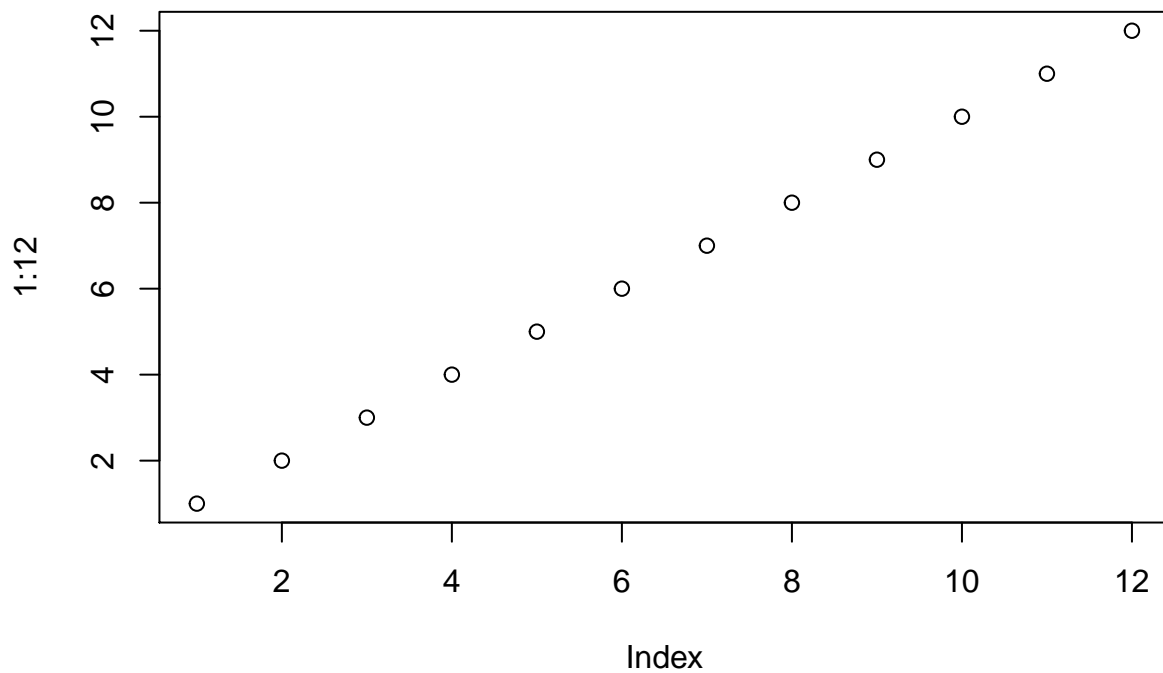Figure 4: figure caption

```r
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```r
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

5.1 Including Plots
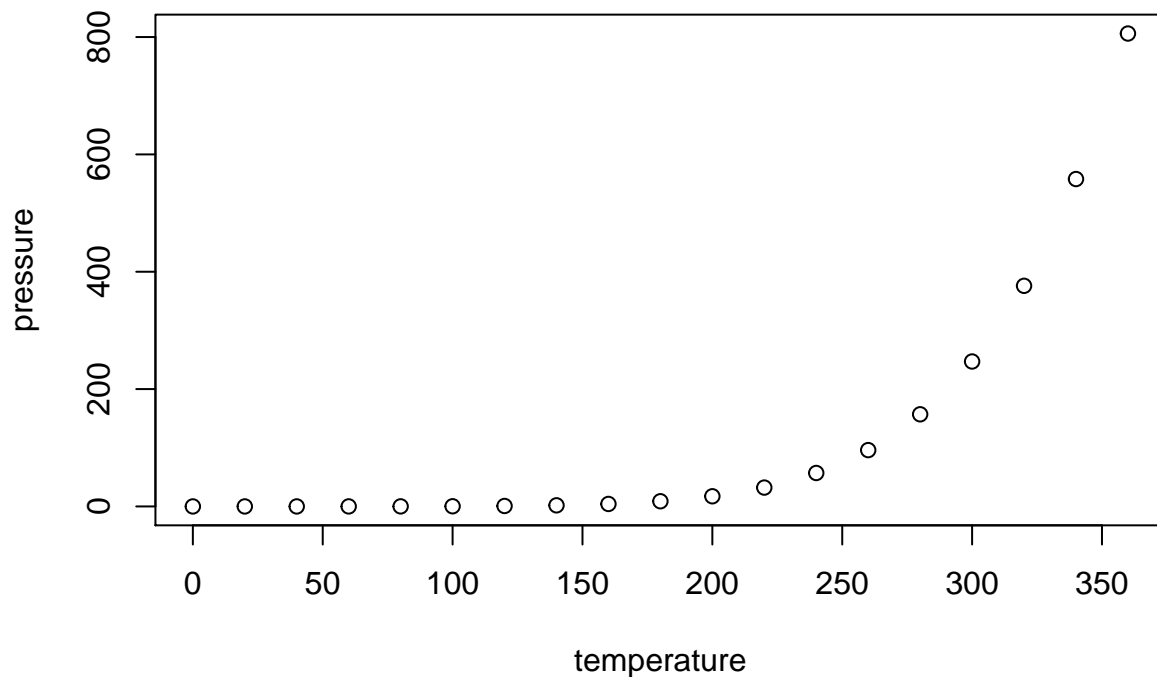
You can also embed plots, for example:

Figure 5: figure caption

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

References

Croucher, Mike, Laura Graham, Tamora James, Anna Krystalli, and Francois Michonneau. 2017. "A Guide to Reproducible Code in Ecology and Evolution." London, United Kingdom: British Ecological Society. https://www.britishecologicalsociety.org/publications/guides-to.

Harrison, Ewen. 2015. "RStudio and Github." https://www.r-bloggers.com/rstudio-and-github/.

Staff, Earth Lab. n.d. "Reproducible Science and Programming Lessons." https://www.earthdatascience.org/tags/reproducible-science-and-programming/.