Reproducible Research Practices for Economists

Mindy L. Mallory

October 24, 2017

Questions for the Audience

How many of your research folders look like this?

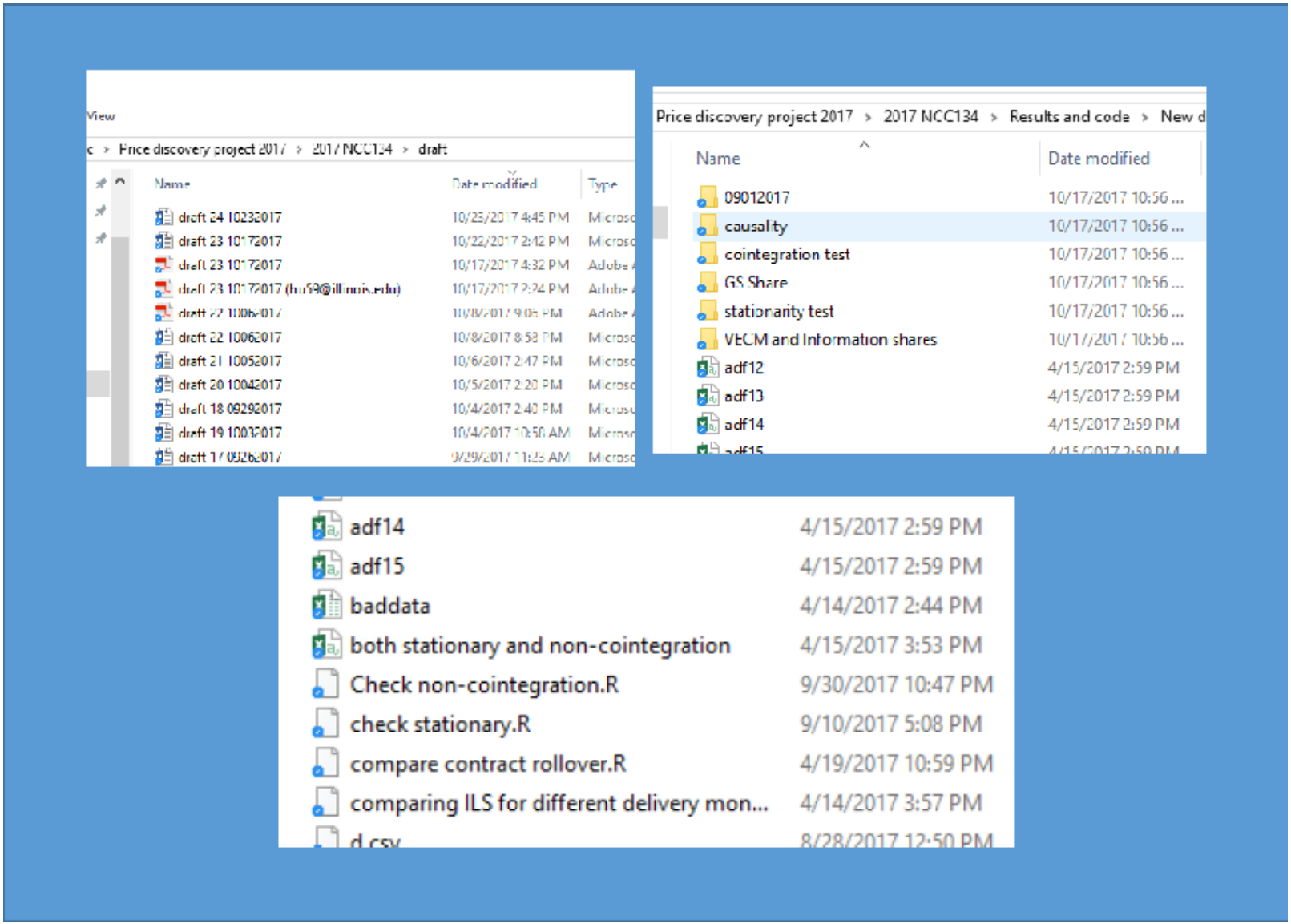
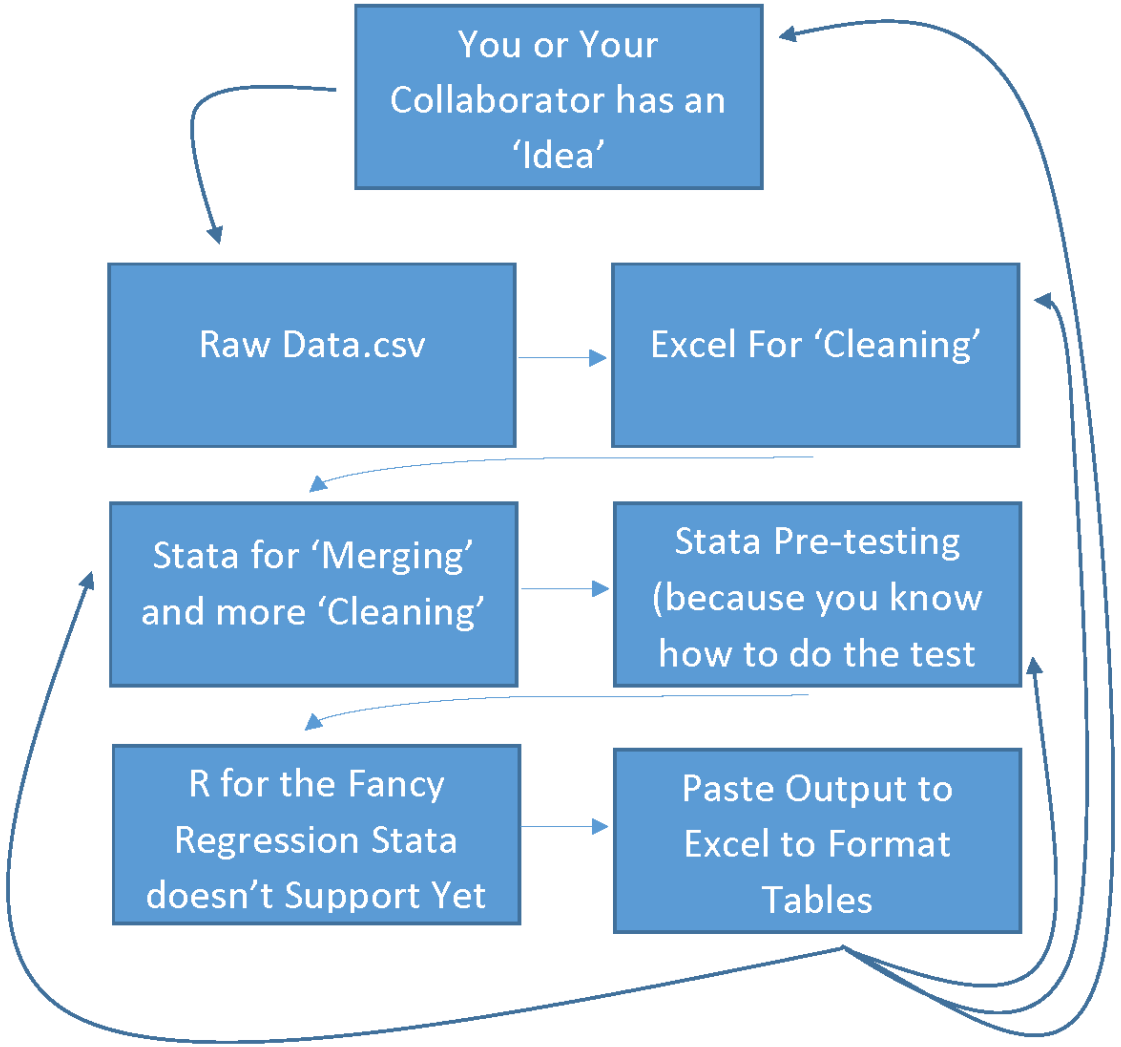


Figure 1: Picking on Zhepeng

How many of you have a research work flow that looks like this?



Questions for the Audience

How many of you would rather die than have to reproduce a table from a paper you published 2 years ago?

Questions for the Audience

Do you wake up in a cold sweat dreaming that Reviewer number 2 asked you to update your data-set (perform robustness test, etc) and you couldn't even reproduce your original results?

Questions for the Audience

Students, have you ever purposely obfuscated your code figuring if your professor can't follow it they can't criticize it?

Questions for the Audience

Have you ever lost data between submission and being asked to revise and resubmit and then you had to go and REPURCHASE!!! said data?

Questions for the Audience

Have you ever lost an entire paper due to the Word file becoming corrupted then you thought you salvaged the paper through document recovery but then it got rejected because you missed some weird characters from the file corruption and reviewer number 2 recommended rejecting your paper because the authors were 'careless' to allow the weird characters to remain the document?

I can say yes to all of these questions!

But I got tired of being nervous all the time!

There is a better way!

Reproducible research with R, RStudio, RMarkdown, Knitr, and Github

- R - is awesome statistical computing software (open source and free!)
- Rstudio - is an awesome integrated development environment (program making it convenient to work with R); also open source and free

Reproducible research with R, RStudio, RMarkdown, Knitr, and Github

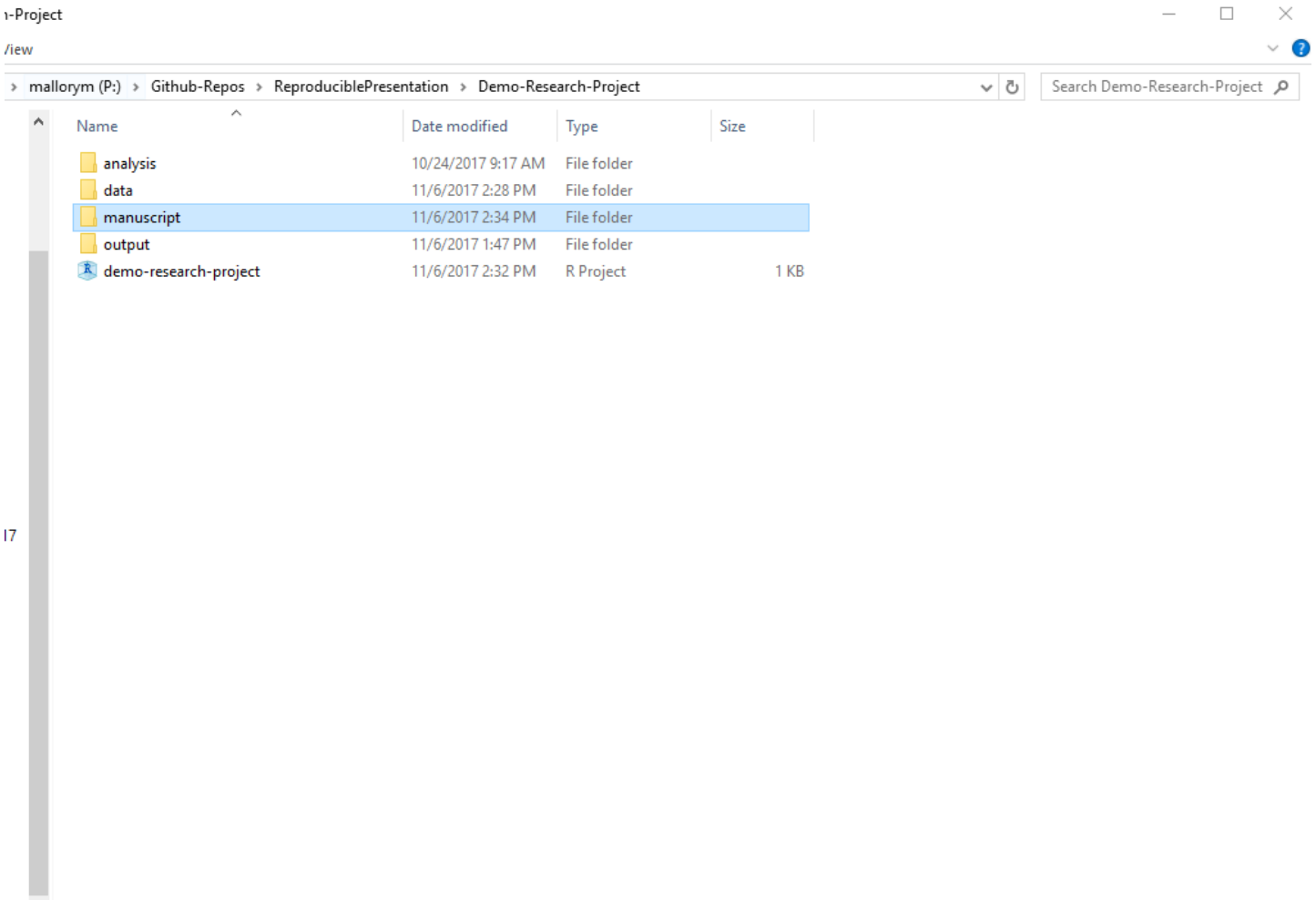
- RMarkdown is a kind of markup language supported by RStudio that uses **Knitr** to weave statistical analysis and results into beautifully formatted documents.
  - Written in plaintext, it understands latex code and documents can be rendered into many different output formats
    - PDF
    - Beamer
    - HTML
    - Word*

Reproducible research with R, RStudio, RMarkdown, Knitr, and Github

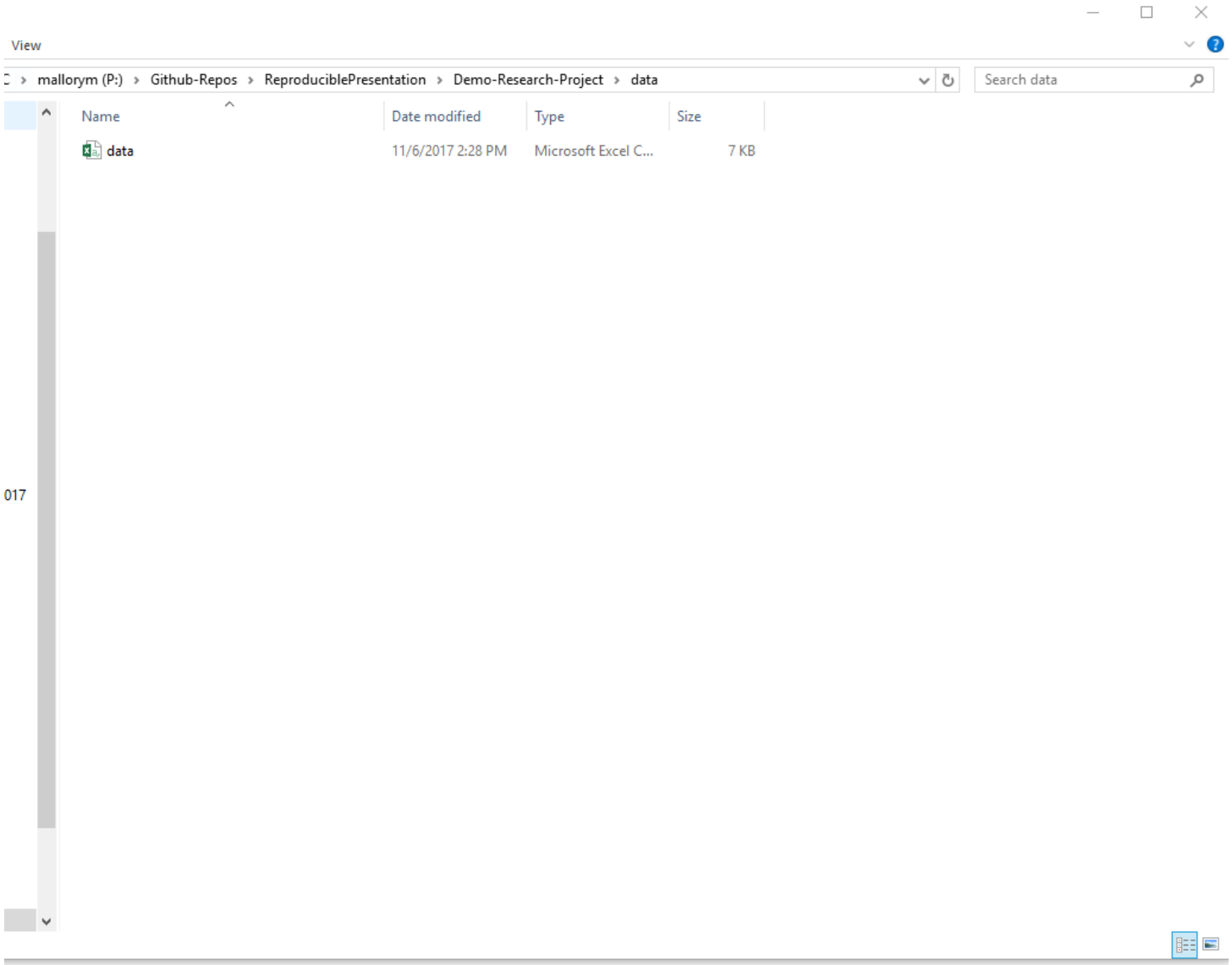
- Github - is a cloud-based repository that is great at versioning (it was designed by and for software developers)

Step 1 - Set up a clean, reproducible project directory

- RStudio Rule #1 - use projects!
- Create 'New Project' in the base of your research project's folder
- Never change the working directory

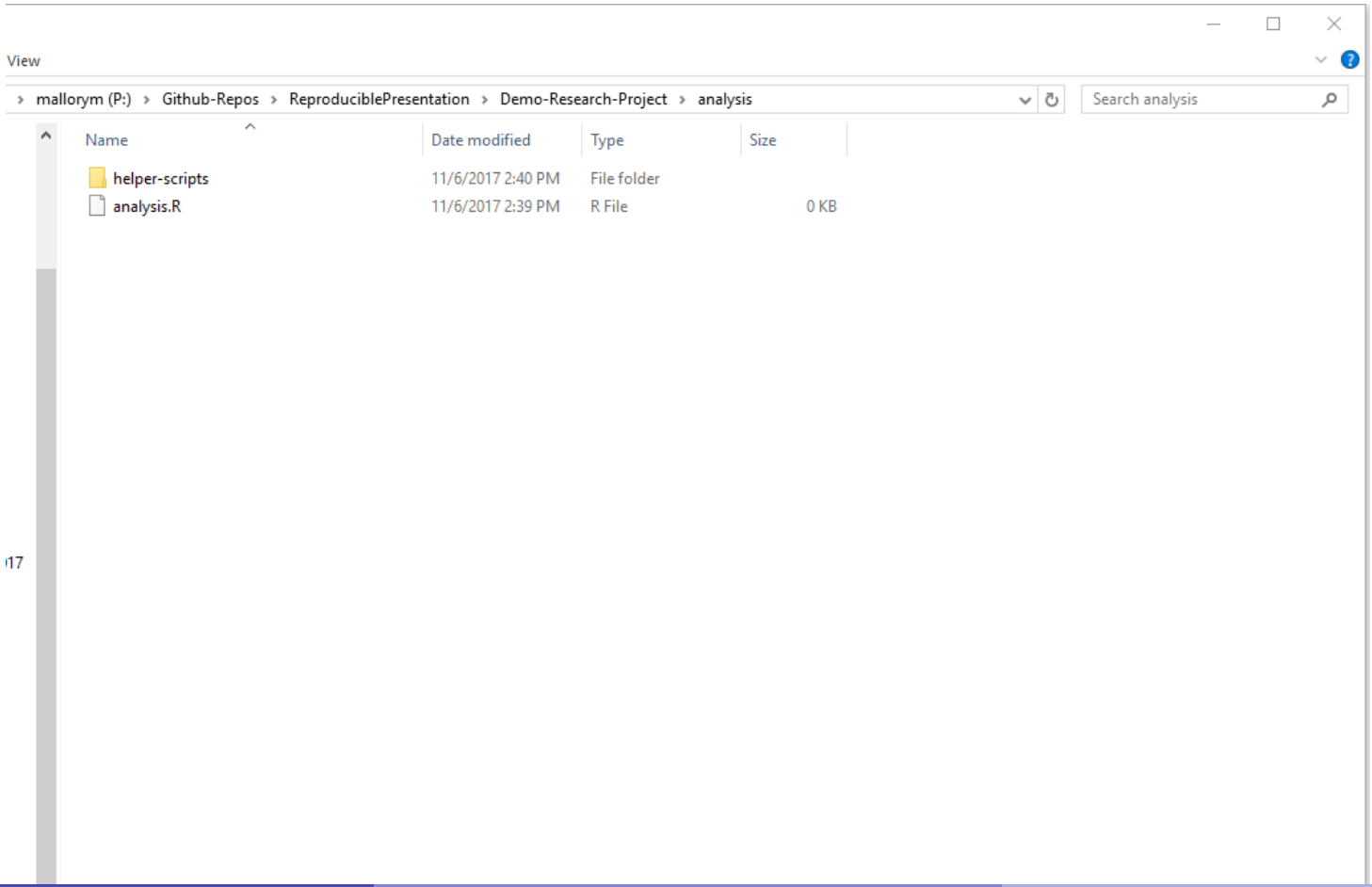


Step 2 - Put your raw data in the 'data' folder and never touch again



Step 3 - Organize Scripts

- Document what each script does
- If your project requires an elaborate 'readme.txt' with instructions about which scripts to run and in what order, you probably need to automate this.



Step 3 - Organize Scripts

- Document what each script does
- If your project requires an elaborate 'readme.txt' with instructions about which scripts to run and in what order, you probably need to automate this.