

Let's 'Git' started: An introduction to version control

Luke & Daiva

2014-11-07

Welcome to our Data-related workshop

Purpose:

To teach a few tips and tricks for more efficiently managing your data, tracking your computer files, understanding appropriate analytical approaches, and speeding up the process from code to tables.

Significance:

Topics we cover will help you get more comfortable with data, reduce the chance of overlooked errors, and give you more control over your work. They are also all important parts of a science movement gaining increasing attention – Reproducible Research.

Overview of some future workshops

- Let's "Git" started! An introduction to version control
- Know your data and how to handle it correctly:
Statistical assumptions
- Fighting chaos: Coding tricks to keep your analysis –
and mind – sane
- Visual exploration (of data): Techniques and code to
better understand your data using plots and graphs

Overview of some future workshops

- Let's "Git" started! An introduction to version control
- Know your data and how to handle it correctly:
Statistical assumptions
- Fighting chaos: Coding tricks to keep your analysis –
and mind – sane
- Visual exploration (of data): Techniques and code to
better understand your data using plots and graphs
- Code Review Clud...?

Reproducible Research

Definition

PLOS Biology by Greg Wilson, founder of Software Carpentry

Version control

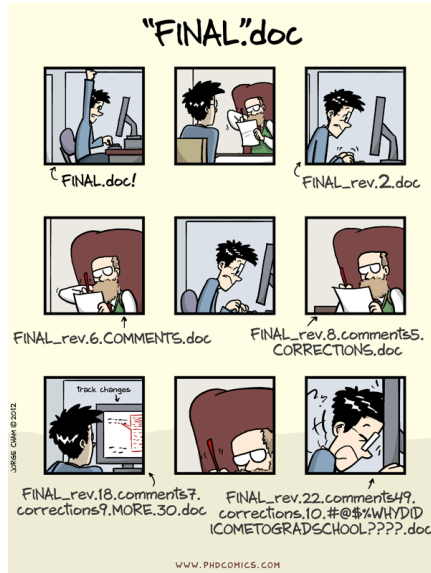


Figure 1:

Brief tangent: Filenaming rules

- 1 Keep names short, meaningful. Remove “the”, “and”, “a”, etc.
- 2 Don't include spaces.
- 3 Use hyphens to separate important parts of the filename.
- 4 Avoid redundancy in file names and folder names (e.g. `folderName/fileName-folderName.txt`, instead use `folderName/fileName.txt`).
- 5 If a number is included, such as for the version number, use two digits not one (e.g. V01, not V1).
- 6 When including a date, include it at the end of the filename and in the international standard format YYYY-MM-DD.

Slide 5

What happens when your files aren't organized? (Can we include a funny screenshot of a super cluttered folder with crazy file names – e.g. Thesis_Proposal_Draft, Thesis_Proposal_FINAL, Thesis_Proposal_FINAL2, Thesis_Proposal_FINAL_FINAL)

-Maybe add another slide after this with points about what could happen (errors, difficulty keeping track of edits, more files to back-up. . .)

Slide 6

Solution: Version Control Define VC Explain Git (brief history of how it came to be, how it is used, Git Repo)

Slide 7

Git Some common git codes and their functions, can verbally explain analogies to familiar tasks (e.g. "cd")

Slide 8

GitHub Have students make an account State that this can be used to back-up files Go through mini exercise of pushing/pulling a txt file – will have to explain why we are tracking txt files and why it is a good idea to get into the habit of using txt files.

Slide 9

Main Exercise