Git/GitHub + Quarto

January 15, 2026

Plan for Today

1 Git and GitHub

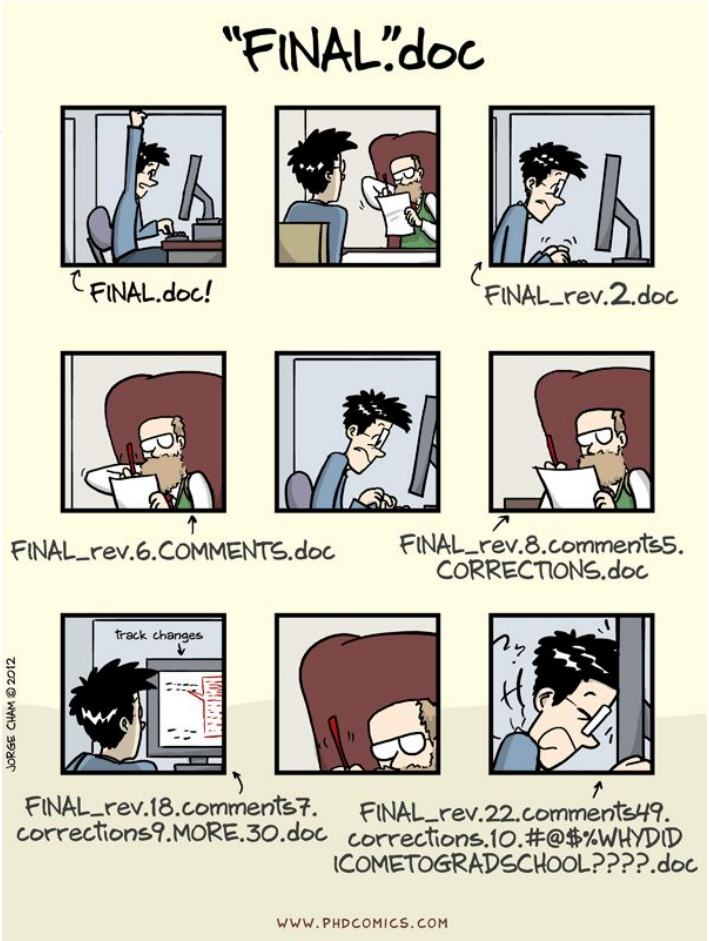
+ Cloning *our* class GitHub repository
+ Creating *your* class GitHub repository
+ GitHub workflow: git pull, add, commit, push

2 Quarto

Introduction to Git/GitHub

What is git?

- A version control system
- Stores data as a series of snapshots
- If files have not changed, it will simply access the file from a previous commit instead of saving it again
- Allows access to all the committed steps along the way



Git vs. GitHub

Local Git Repository

- You have a local version of the folder on your computer
- History stored in .git file
- Only you can see the changes made in the local version

Remote GitHub Repository

- A remote version of the folder is hosted on the GitHub website
- Everyone can see these changes (if repository is public)

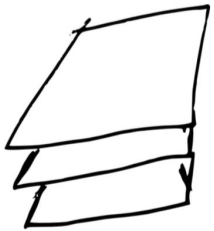
Why do we need Git/GitHub?

- Imagine working on a project with several collaborators...
- Using Git/GitHub allows everyone to have their own local version of the project while still maintaining a “main” version of the project, hosted remotely on GitHub
- You can make changes freely without people seeing what you are doing
  - You can thoroughly test your changes before adding to the master copy
- Version control!!
  - Especially great if your changes create bugs because you can backtrack/revert

Typical Git/GitHub Pipeline

(2) make local changes

(e.g., create file called filename.txt)



LOCAL REPOSITORY

(1) git pull

(to retrieve the most recent version from the server)

REMOTE REPOSITORY

(3) git add filename.txt

(changes are staged/waiting to be committed)

(4) git commit -m "[description of changes]"

(commit when you have made some changes and want to be able to save your current checkpoint as a snapshot)

(5) git push

(make changes available to everyone with access to the repo)

Warning: remember to "git pull" before "git push" to mitigate potential merge conflicts

Getting started with Git/GitHub for this class

Tools for today: Git and GitHub

Make sure that you have installed **git** on your computer: https://git-scm.com/book/en/v2/Getting-Started-Installing-Git

- In terminal, type: `git --version`

If you haven't already, **please sign up for GitHub** (https://github.com)

- Sign up for the student pack (https://education.github.com/) to get unlimited private repositories. You are a "student" and want an "individual account".

While optional, I highly recommend downloading:

- **GitKraken:** https://www.gitkraken.com/
  - A nice GUI for interacting with git/GitHub

Introduction to Quarto

What is quarto?

- A powerful open-source tool for creating **dynamic documents**, reports, presentations, and websites using a simple **markdown-based format**
  - Akin to R Markdown but has multi-language support (R, Python, Julia, JavaScript, ...)
- Designed to enhance **reproducible research** and data science workflows by allowing users to **combine code, text, and visualizations** in a single document
- How to install?
  - Download quarto: https://quarto.org/docs/download/index.html
  - To check installation, type in terminal: `quarto --version`

Recap + What's Coming Up

Recap

- **Git** is a version control system. **GitHub** hosts git repositories remotely + does more.
- **Typical GitHub Workflow:** git pull, add, commit, push
- **Quarto:** for reproducible documentation

Supplemental Resource

- Missing Semester of your CS Education: https://missing.csail.mit.edu

Coming up...

- What do **problem formulation** and **data collection** look like in reality?
- We will discuss **case studies**, focusing on these beginning stages of the data science life cycle