Learning goals:

- Understand Jupyter
- Understand Datahub
- git review
- Lab & Q&A

Discussion 1

COGS 108 Winter 2021

Discussion slides and materials adapted from Sam Lau (TA: WI20)

Technical Discussion Sections

Section Philosophy

- Attendance is not required
- Reasons to never miss the discussion sessions:
  - Demos to help you set up and start working
  - Hands-on experience and personalized guidance
  - Time to practice technical aspects of the course
  - Discussion is not supposed to be a monologue – Please ask questions
  - Condensed information >> searching the infinite www

Zoom format

- Review -> Demo -> Lab & Q/A.
  Please mute your microphone, but feel free to unmute and ask questions!
- If you’re more comfortable with text -> Keep asking questions on chat
- We will take up assignment specific questions towards the end of Discussion or during OH.

Programming

- This course assumes basic programming knowledge…but not much
- Resources:
  - codeacademy
  - Start Here: https://github.com/COGS108/Tutorials/blob/master/01-Python.ipynb
  - Python in detail: https://jakevdp.github.io/PythonDataScienceHandbook/
  - Pandas: https://www.dataschool.io/python-pandas-tips-and-tricks/
  - Git: https://guides.github.com/activities/hello-world/
- Cheatsheets
  - Google: 'python cheatsheet', 'pandas cheatsheet', 'git cheatsheet' (find one that's good for you)

Resources will be added throughout the quarter: https://github.com/COGS108/Resources

Anaconda

The Data Science Toolkit – contains Python and data science libraries (including jupyter notebooks)

- Download: https://www.anaconda.com/products/individual
- Installation: https://docs.anaconda.com/anaconda/install/
- Verify installation: https://docs.anaconda.com/anaconda/install/verify-install/
- Make sure anaconda is added to the system path: For mac: export PATH="/usr/local/anaconda3/bin:$PATH"

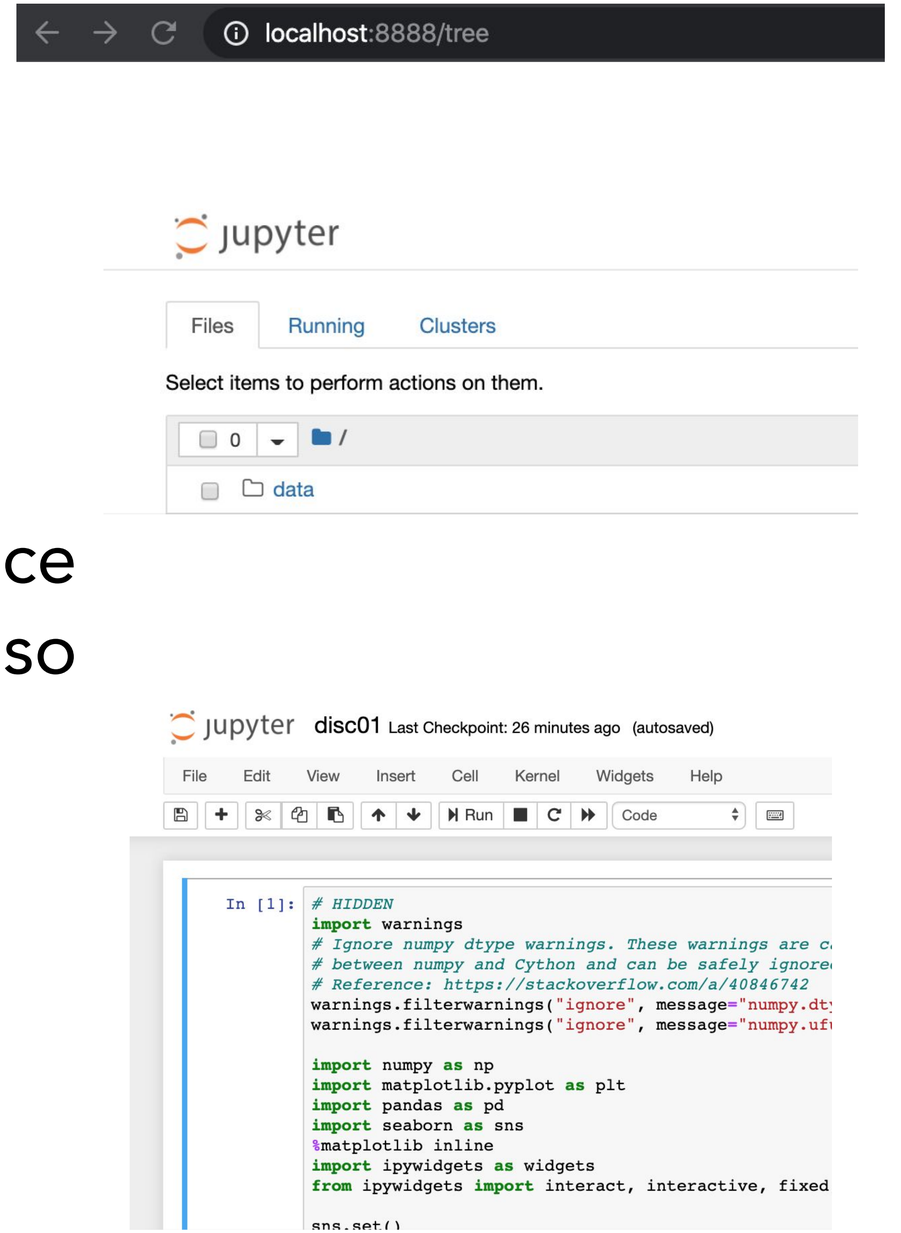
git

Version control system!

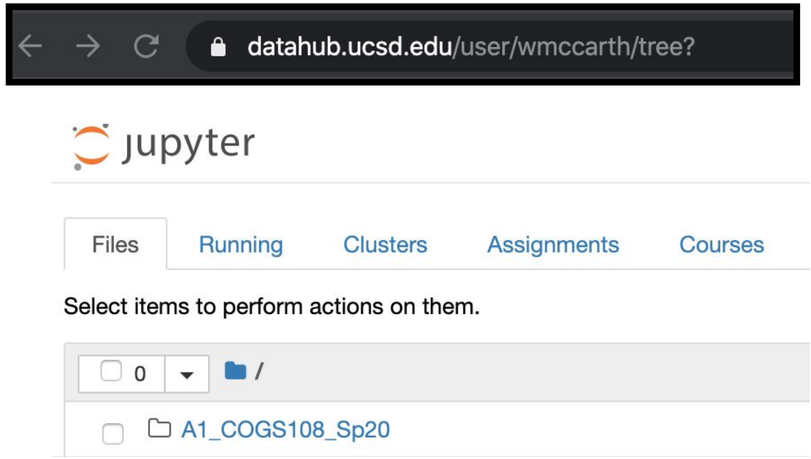
- Go to https://git-scm.com/downloads
- Choose your Operating System (Windows/OS X/Linux)
- Follow the steps specific to your OS
- Verify installation: In terminal type “git –-version”

Jupyter

- Python code is run on a python interpreter
- Jupyter is a program that creates an interface for typing python code in a browser, that also runs that code in a python interpreter
- What does this mean?!
  - Jupyter is a way of running python programs from a browser (like chrome) (hooray!)



datahub



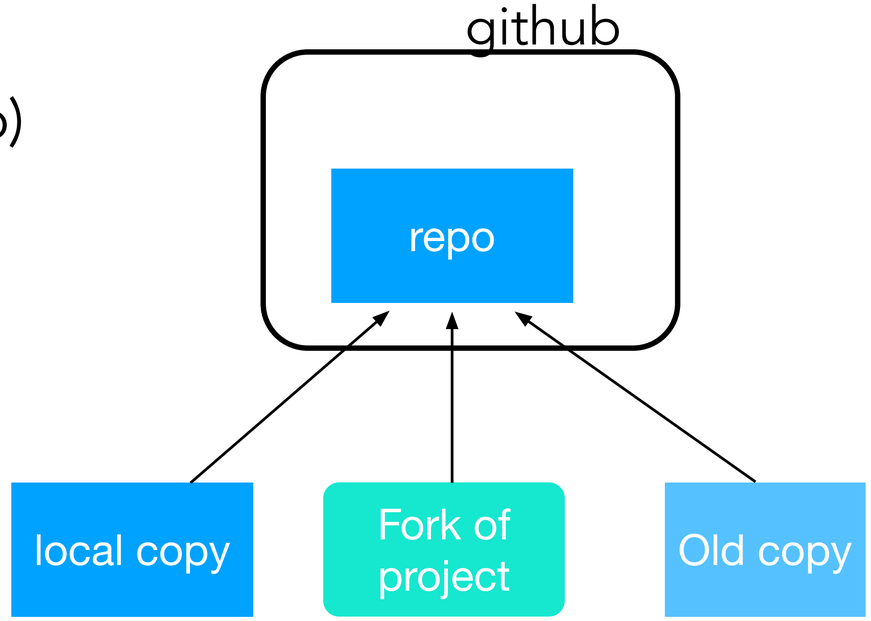
- Jupyter runs python code in a browser.
  - But Jupyter is itself just a program that's running on a computer somewhere.
- datahub lets you interact with Jupyter that's running somewhere else.
- What does this mean?!
  - You don't need to worry about installing Jupyter
  - You can use datahub to create and run python programs (online)
  - You can use this interface to fetch and submit assignments

Checking your work (in general on datahub)

- The tests built into the notebook are (very) minimal.
  - To write your own tests, add a cell with assert statements below your code.
- Make sure to click Validate before turning in your notebook!
  - This replicates what our autograder will do.
  - Hidden tests on the autograder aren't validated.
- Make sure to click submit once you're ready to submit.
  - And don't click submit after the deadline unless you mean to submit late.

What is git + GitHub?

- Somewhere online to store a copy of a project (Github)
- Plus a tool to interact with this copy (Git)
  - Command line and desktop versions
- A way of keeping track of changes you make to this project



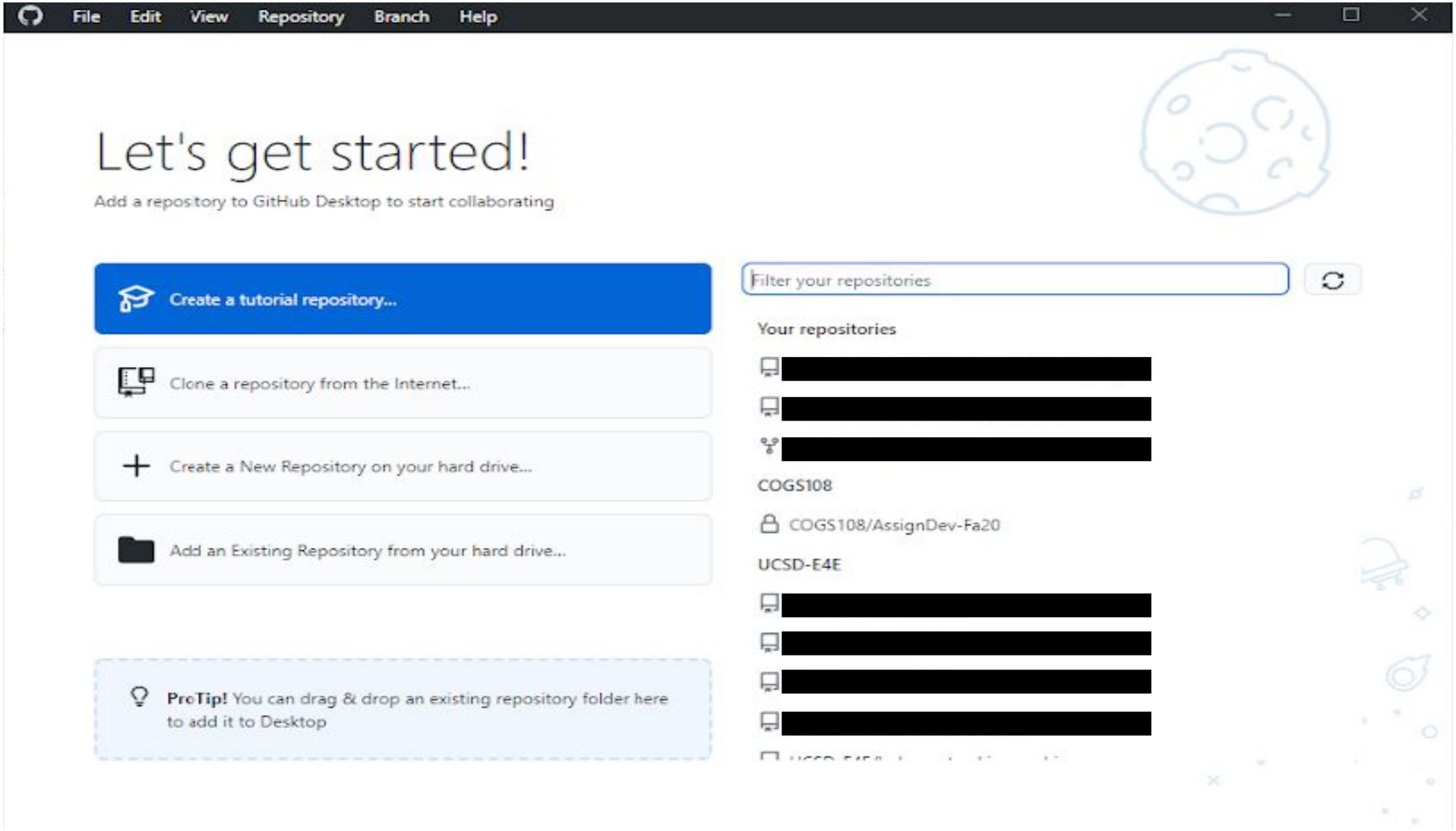
Why use git + GitHub?

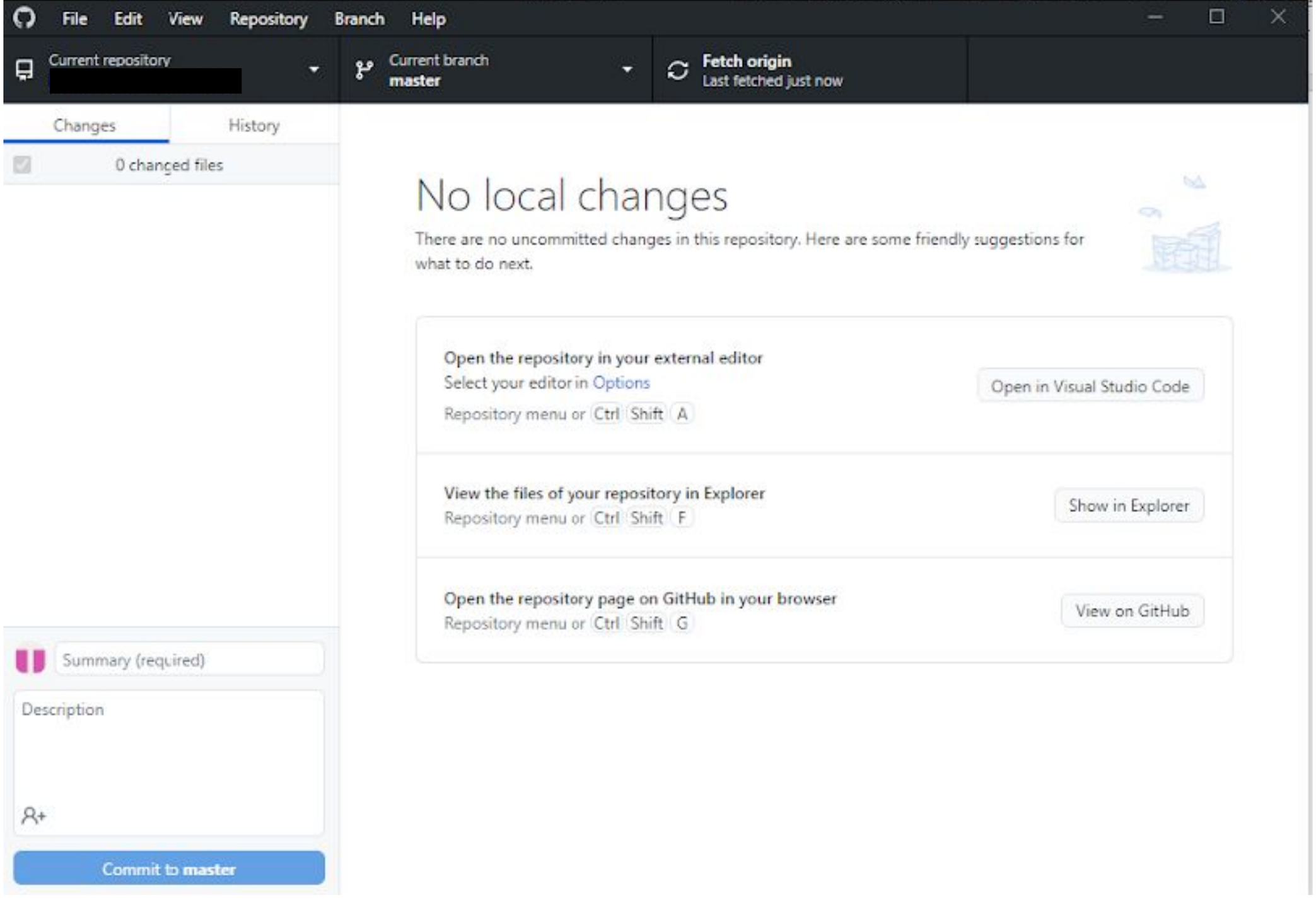
- Git allows you to work on code projects with other people. It's the preferred tool for many projects, like:
  - Python: https://github.com/python/cpython
  - Jupyter: https://github.com/jupyter/
  - COGS 108: https://github.com/COGS108/
- Backup
- Version control (*undo* on a large scale)
- Code reuse

Part 1 Walkthrough

Git Stage, Commit and Push (Demo)

- Cloning a repo on your local machine
- Working on you remote repo (making changes to files)
- Stage, commit and push these changes to your Github repo
- Commands you should know:
  - Git clone
  - Git status (not really needed – but really helpful)
  - Git add
  - Git commit
  - Git push


File
Edit
View
Repository
Branch
Help
Let's get started!
Add a repository to GitHub Desktop to start collaborating
Create a tutorial repository...
Clone a repository from the Internet...
Create a New Repository on your hard drive...
Add an Existing Repository from your hard drive...
ProTip! You can drag & drop an existing repository folder here to add it to Desktop
Filter your repositories
Your repositories
[redacted]
[redacted]
[redacted]
COGS108
COGS108/AssignDev-Fa20
UCSD-E4E
[redacted]
[redacted]
[redacted]
[redacted]


File Edit View Repository Branch Help
Current repository
Current branch
master
Fetch origin
Last fetched just now
Changes
History
0 changed files
No local changes
There are no uncommitted changes in this repository. Here are some friendly suggestions for what to do next.
Open the repository in your external editor
Select your editor in Options
Repository menu or Ctrl Shift A
Open in Visual Studio Code
View the files of your repository in Explorer
Repository menu or Ctrl Shift F
Show in Explorer
Open the repository page on GitHub in your browser
Repository menu or Ctrl Shift G
View on GitHub
Summary (required)
Description
Commit to master

Checking your work for Part 1

- Check that your COGS108_Repo has a README and a .gitignore file. (Repo must be on your account.)
- Go to COGS108/MyFirstPullRequest, click Pull requests.
  - Make sure your PR shows up in that list.
  - Make sure your PR has the right title.
  - Make sure your PR has the right file (correct name).

Working on your labs/assignments

- Log into datahub.ucsd.edu
- Go to Assignments tab
- ‘fetch’ assignments you have access to -> Submit after completion

Today’s Discussion Lab: `D1_python`

Resources

\Most of the course related information:
https://github.com/COGS108/Overview
(It'll be worth bookmarking this page. You'll use it a lot)

For a long list of interesting datasets:
https://tinyletter.com/data-is-plural

Next week: A1 help, git walkthrough