HAS Tools:

Overview of tools and the modern Python data science landscape

August 28, 2024

Recall the two platforms

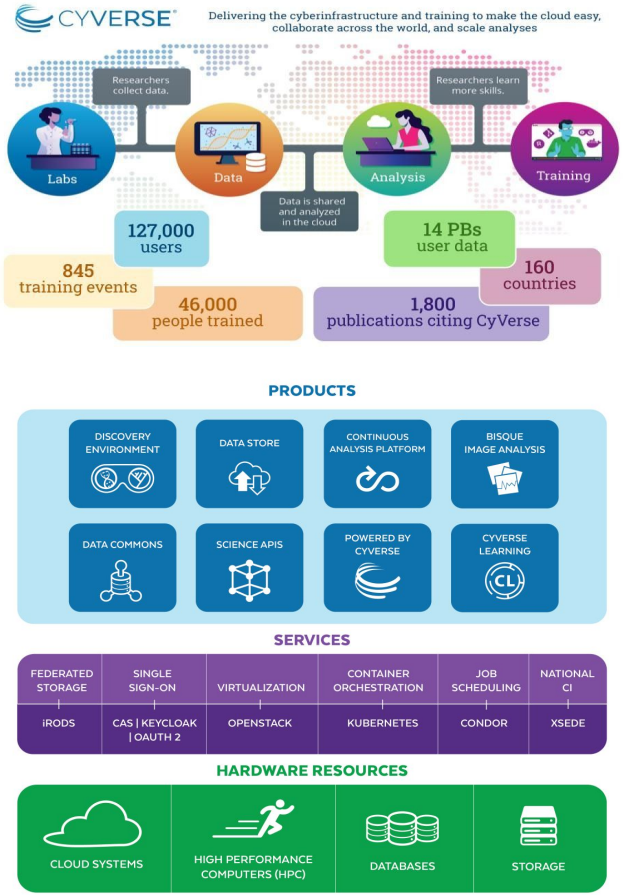
*was going to be

This ~~will~~ be our computational platform and coding interface

This will be our content distribution platform and version control system

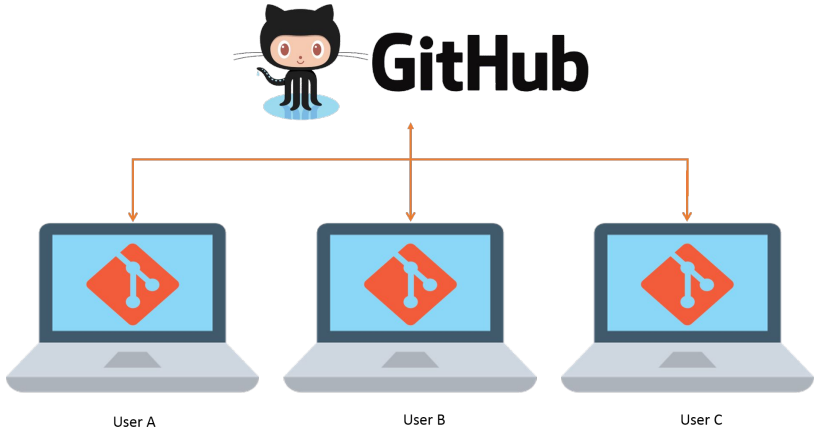
CYVERSE®

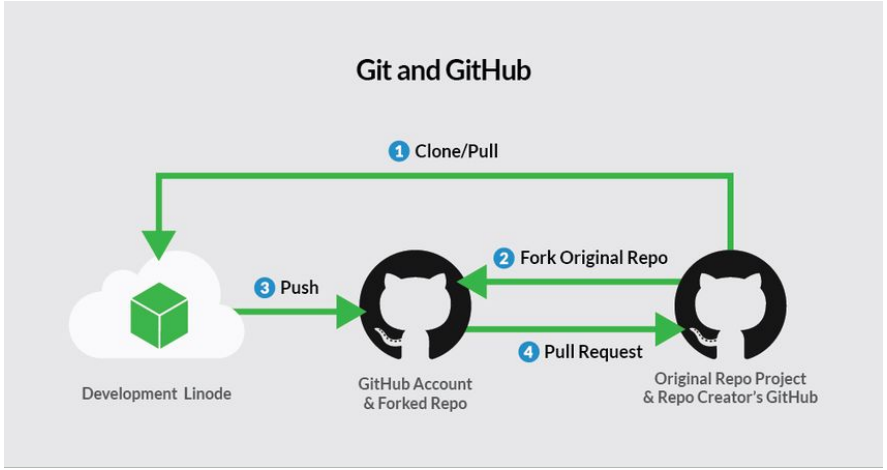
- CyVerse is an open platform and project led by the University of Arizona
- Provides a platform to enable large datasets and computation
- Also allows for standardized “apps” that can be used by many users
- We will use it as our main entrypoint for the class



GitHub

- GitHub is a web platform that enables software developers to manage and share their code
- Also has great integrations for running automation, building websites, and increasing AI integration
- We will use it for course centralization and homework management via GitHub education





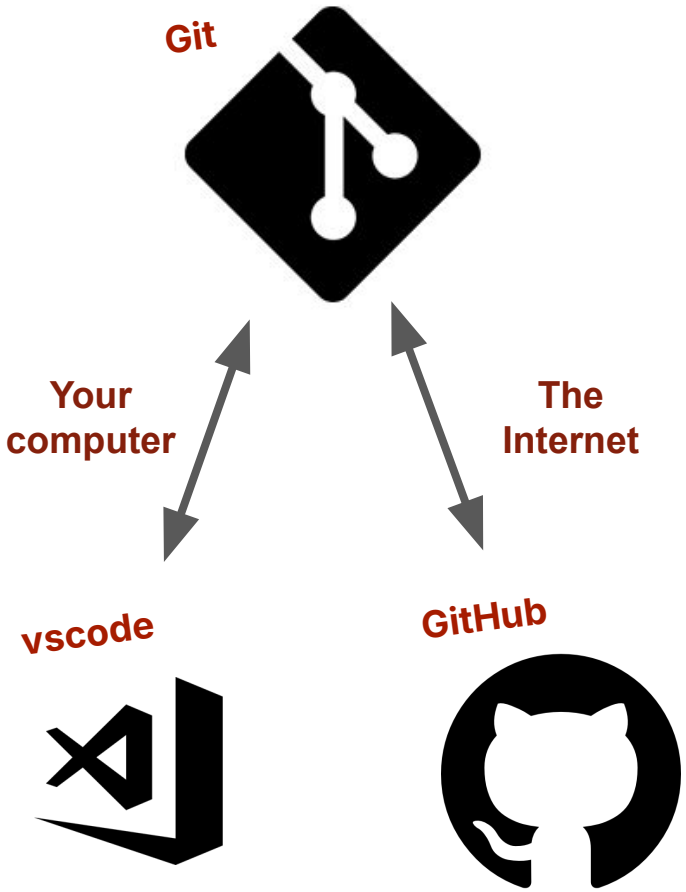
Some other tools used to build this course

- We will be doing all of our work in Python
  - High level, dynamic, and prominent programming language with massive community support
- We use conda for environment management
  - Provides access to community developed python packages - more on this towards the end of the semester
- Under the hood we use jupyter as an interactive computing framework
  - Makes it easy to run “snippets” of code and view their outputs. The standard framework for a lot of exploratory scientific computing
- VSCode will provide the user interface and code editing experience
  - A modern, extensible, and popular code editing application

Please bear with me for a few minutes, but I swear this will all make sense eventually

Let’s talk about git.

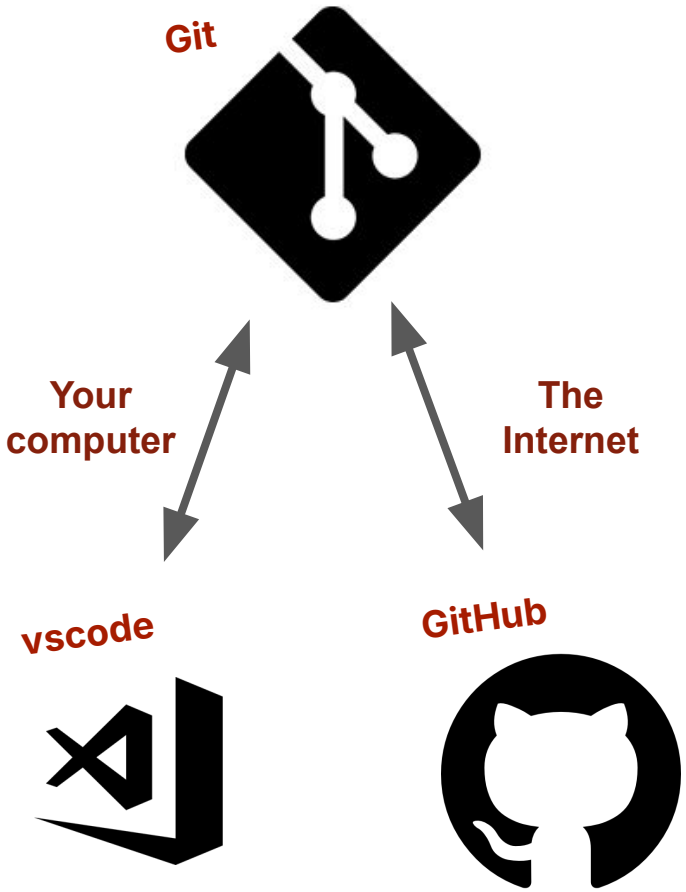
Git is a version control system. It allows teams to work collaboratively on the same pieces of code (like track changes for MS Word but much more sophisticated)



Let’s talk about git.

Git is a version control system. It allows teams to work collaboratively on the same pieces of code (like track changes for MS Word but much more sophisticated)

GitHub is a website that is essentially a social platform for code, built around git as a version control system.

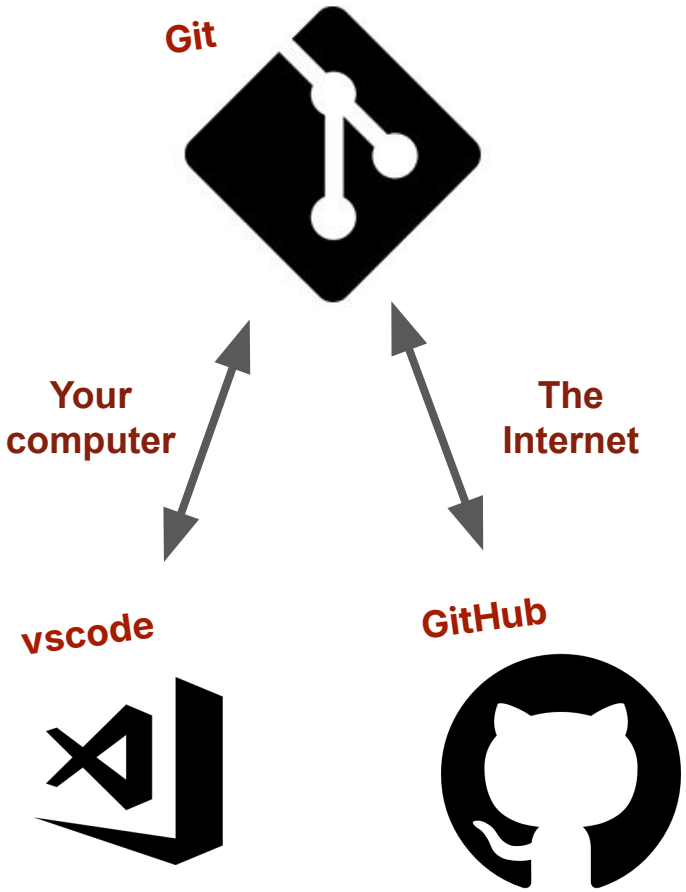


Let’s talk about git.

Git is a version control system. It allows teams to work collaboratively on the same pieces of code (like track changes for MS Word but much more sophisticated)

GitHub is a website that is essentially a social platform for code, built around git as a version control system.

You will mostly interact with git inside of vscode, and use it to communicate with GitHub



The first thing to know about git:

JARGON

repository	The overall “container” for your work (synonym: repo) [noun]
clone	Make a copy of the repository on your local machine [verb]
commit	A snapshot in time of your work, or to save changes [noun|verb]
checkout	How to get to a particular snapshot [verb]
branch	A label to a commit, usually denoting a separate stream of work [noun]
merge	Combine changes from different branches [verb]
remote	A computer somewhere else with a repo on it (GitHub) [noun]
local	The computer you are working on (laptop) [noun]
fetch	Download information about history from a remote [verb]
pull	Do a fetch, but also update the code status with a checkout [verb]
push	Sync changes on your local to a remote [verb]
conflict	Occurs when two commits contradict each other [noun?]

Basic concepts in git: Nodes and trees

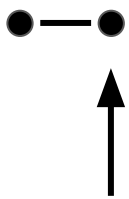
All repos start from an initial commit

You are here

```
# Your Codefile
print(‘hello world’)
```

Basic concepts in git: Nodes and trees

All repos start from an initial commit
As you do work, you can make changes and add commits

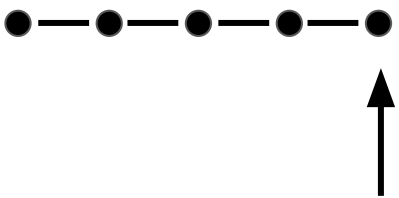


You are here

```
# Your Codefile
name = ‘HAS Tools’
print(‘hello ’ + name)
```

Basic concepts in git: Nodes and trees

All repos start from an initial commit
As you do work, you can make changes and add commits



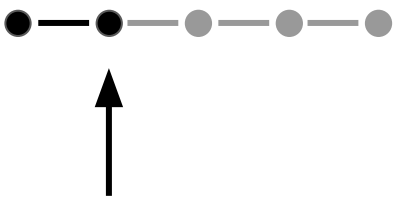

```
# Your Codefile
import numpy as np
name = ‘HAS Tools’
print(‘hello ’ + name)

x = np.arange(50)
y = np.sin(x / np.pi)

# more fun stuff here
```

Basic concepts in git: Nodes and trees

All repos start from an initial commit
As you do work, you can make changes and add commits
You can also move back and forward

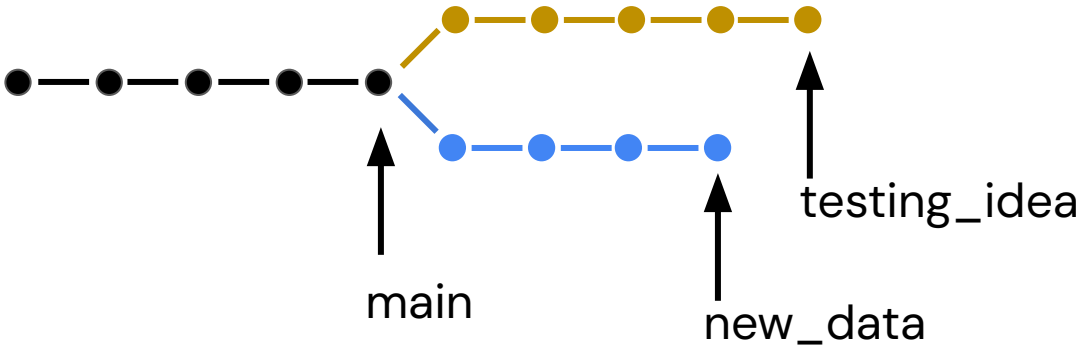


You are here

```
# Your Codefile
name = ‘HAS Tools’
print(‘hello ’ + name)
```

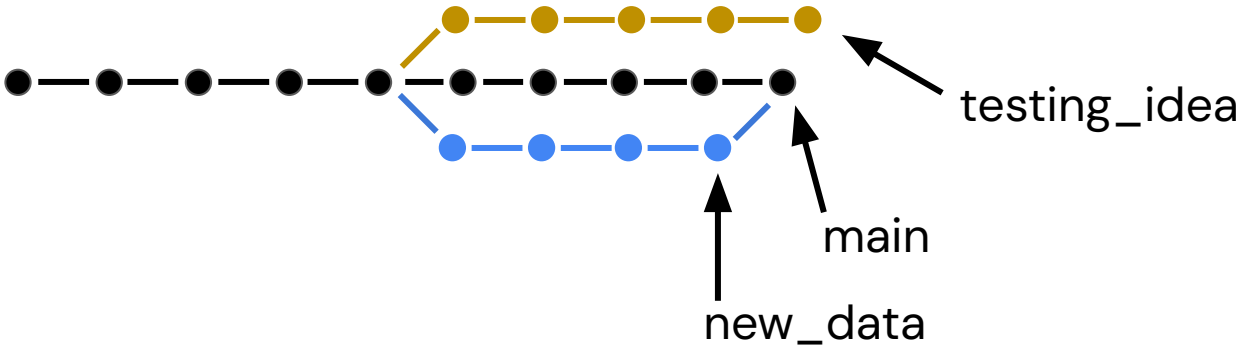
Basic concepts in git: Nodes and trees

You can also start splits called branches

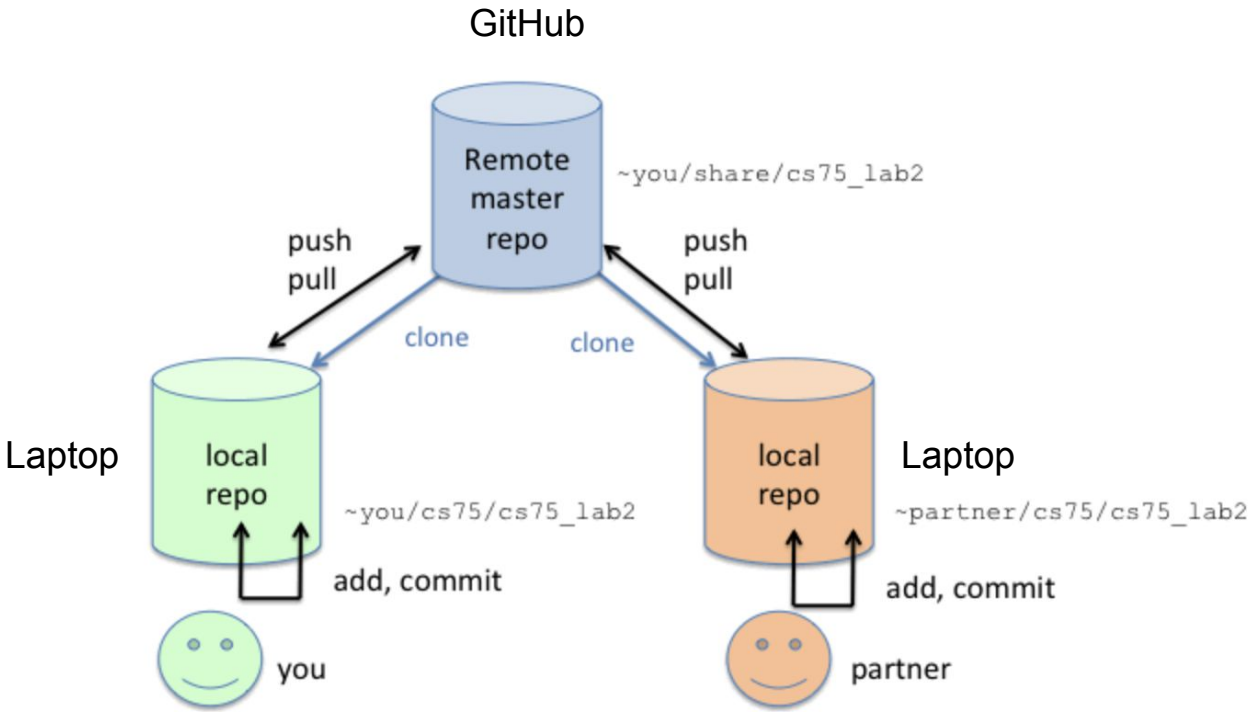


Basic concepts in git: Nodes and trees

You can also start splits called branches
And when you finish an idea you can merge them together

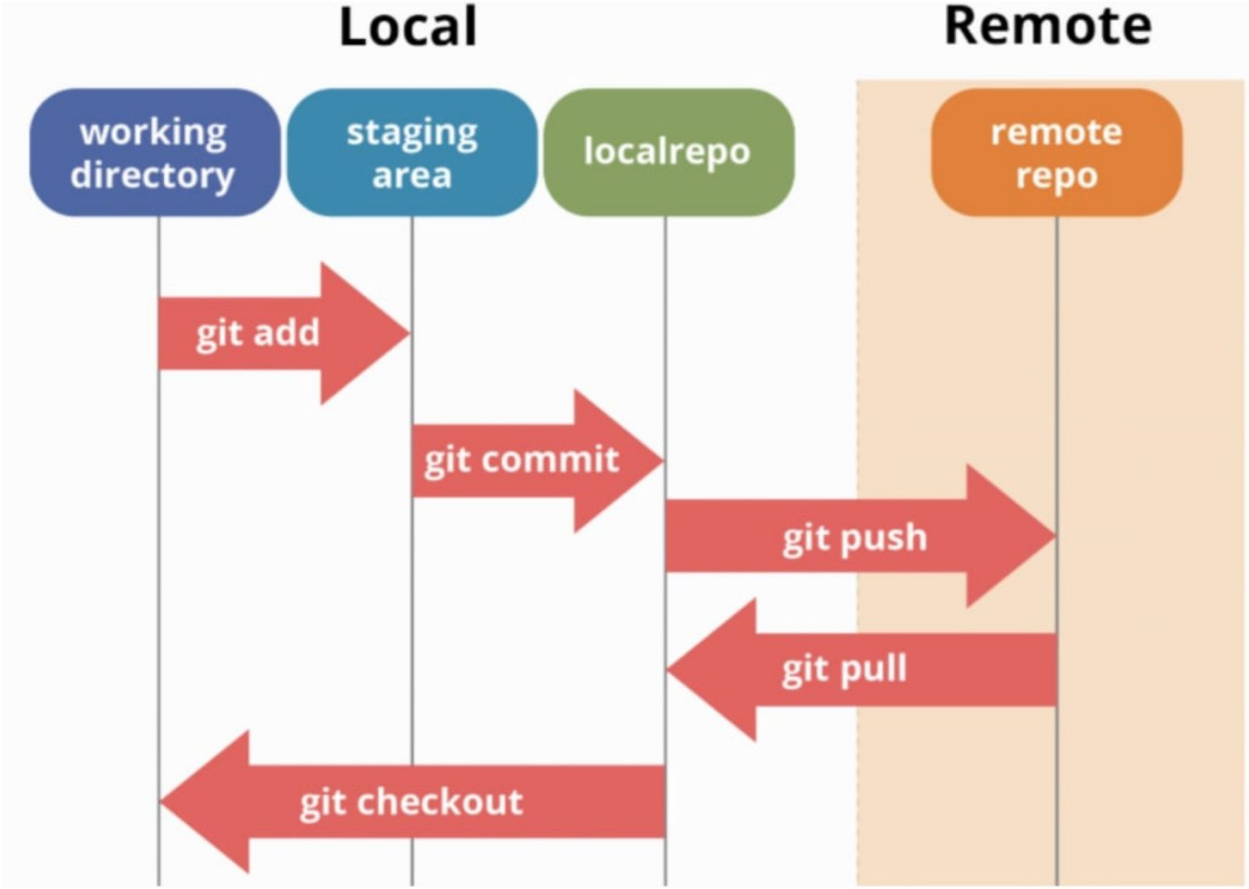


Basic concepts in git: Local and remote



https://www.cs.swarthmore.edu/~newhall/unixhelp/git_create.php

GitHub workflow

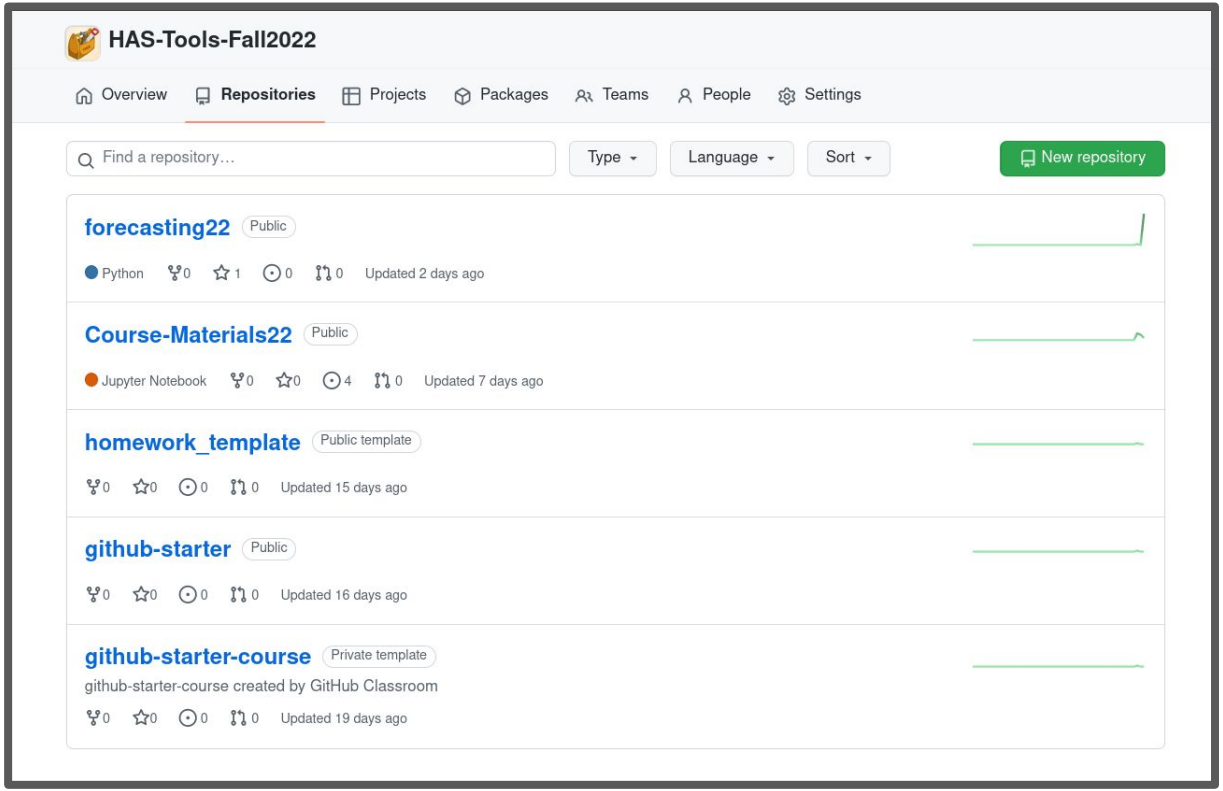


https://dev.to/mollynem/git-github--workflow-fundamentals-5496

https://www.reddit.com/r/git/comments/99ul9f/git_workflow_diagram_showcasing_the_role_of/

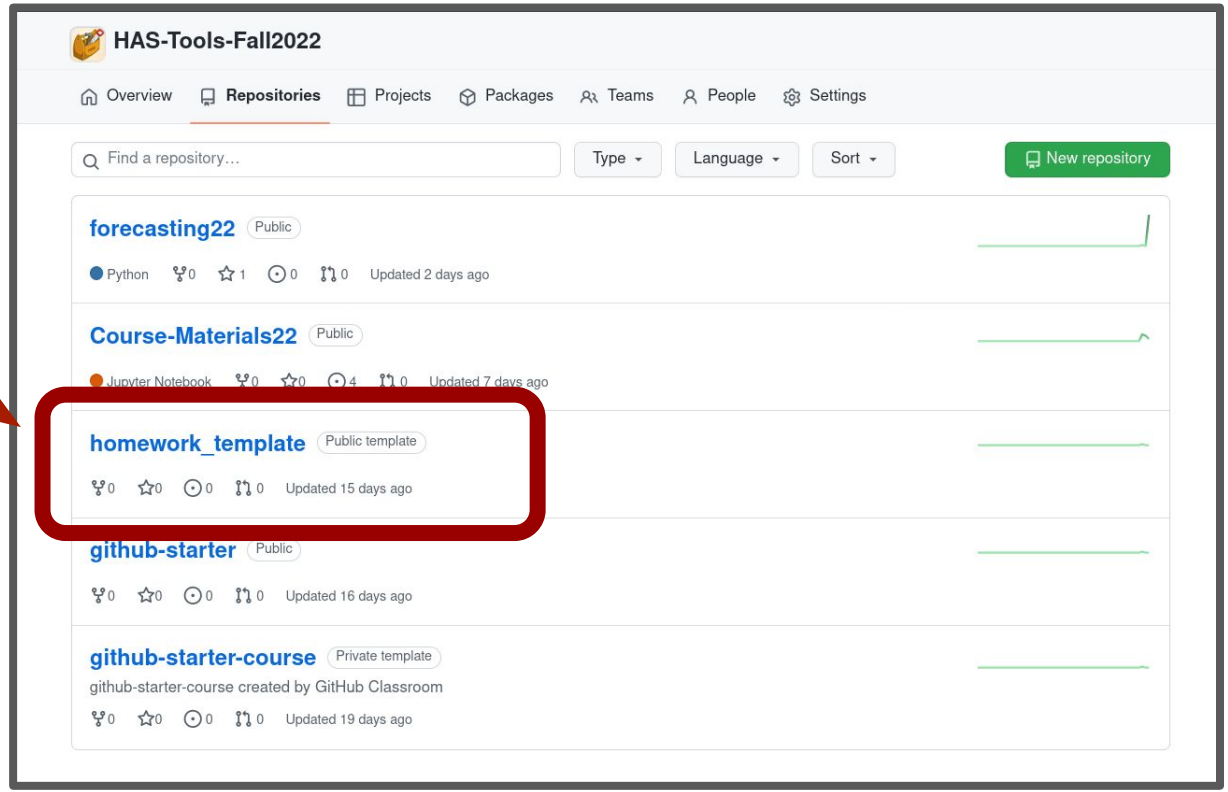
To get familiar with the setup let’s start with a short tour

The homework repo

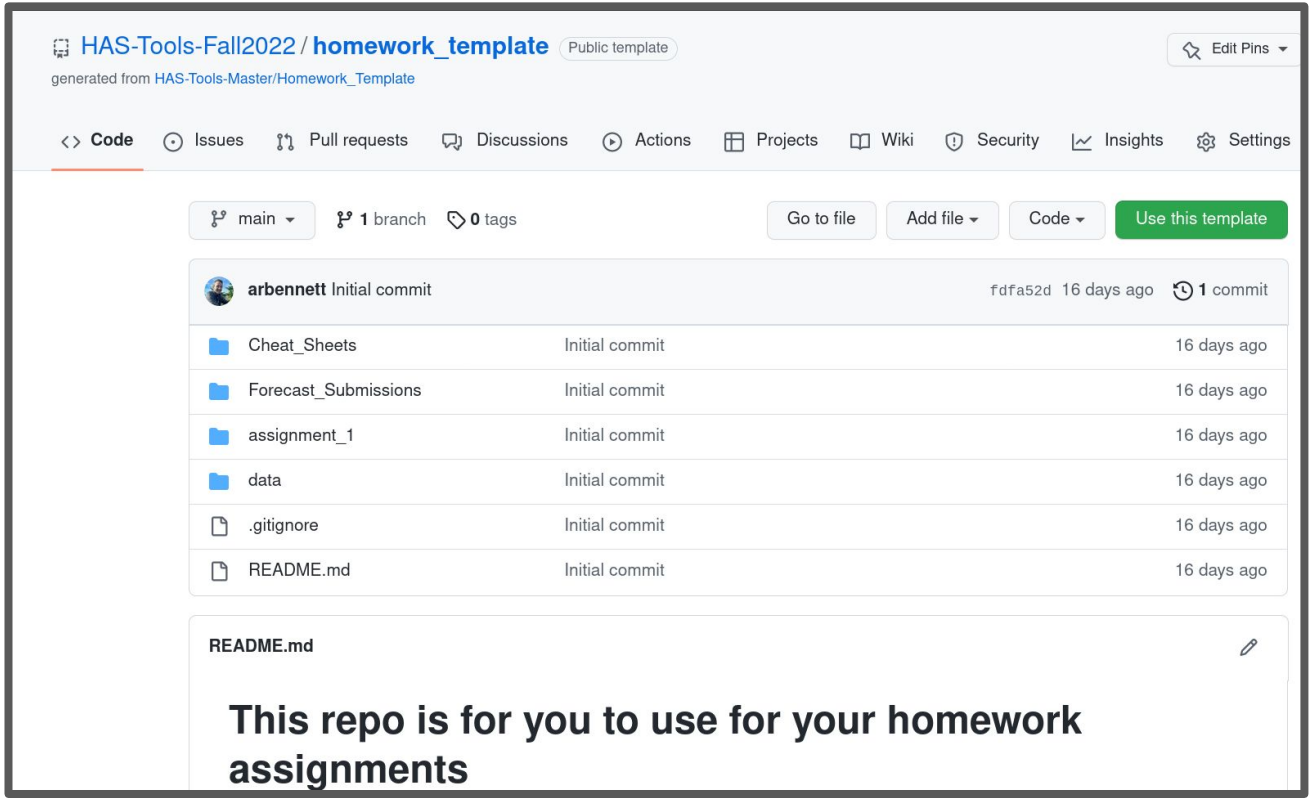


The homework repo

CLICK HERE

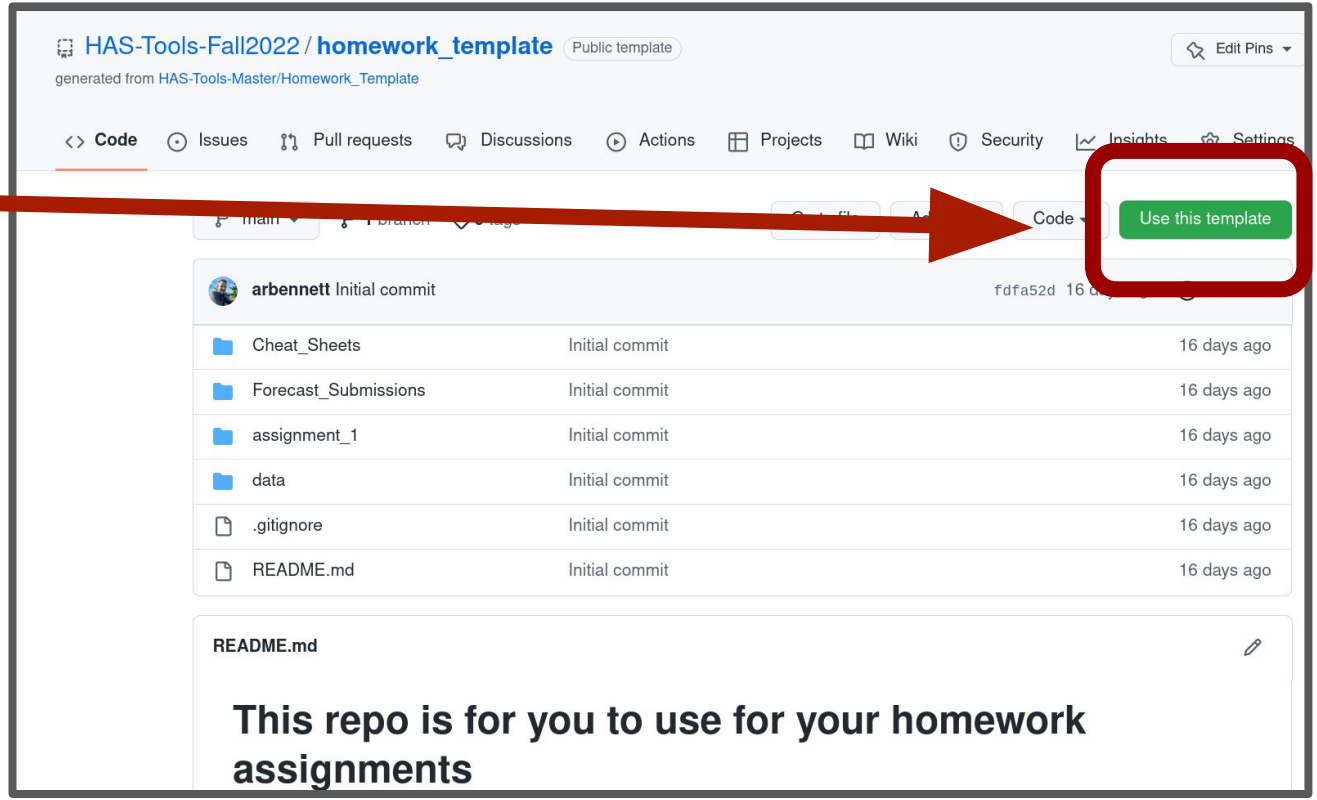


The homework repo



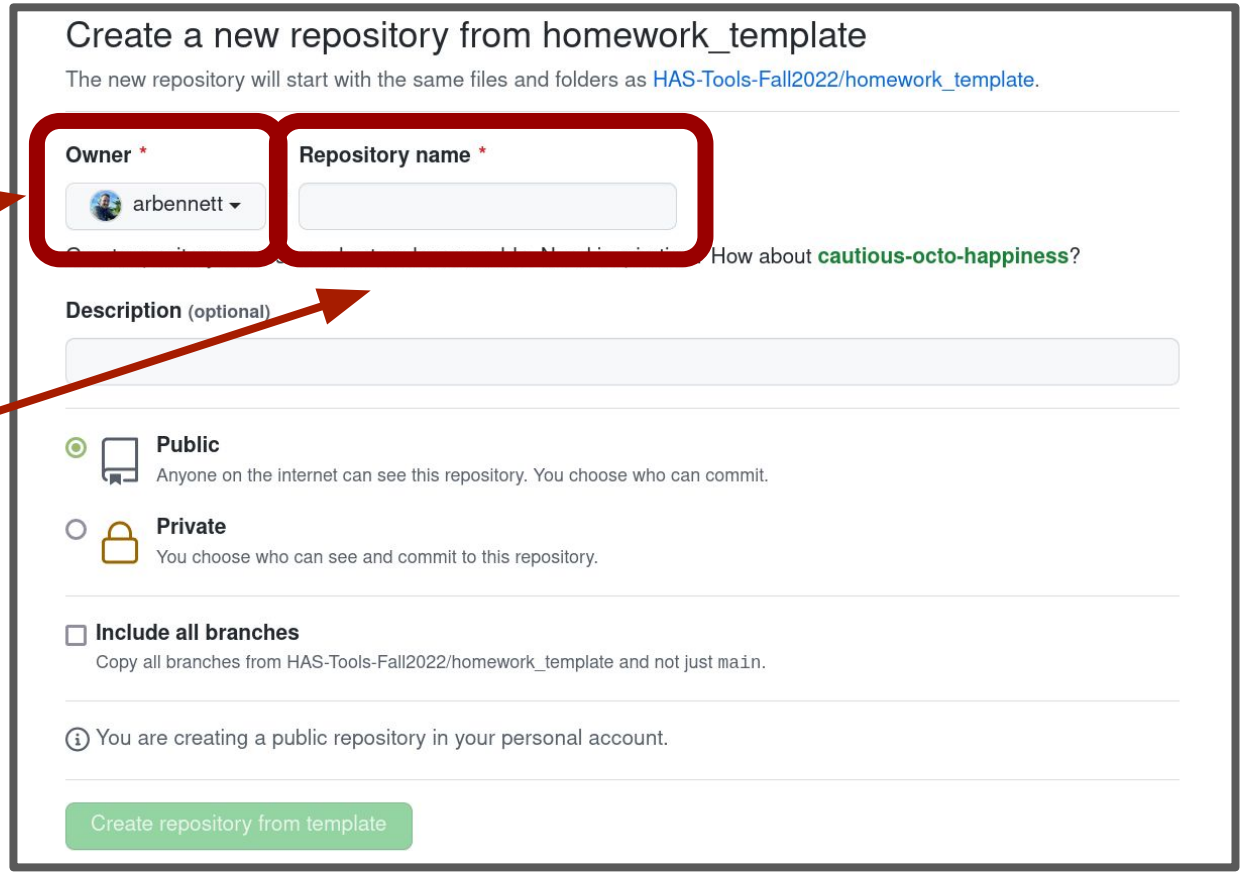
The homework repo

CLICK HERE



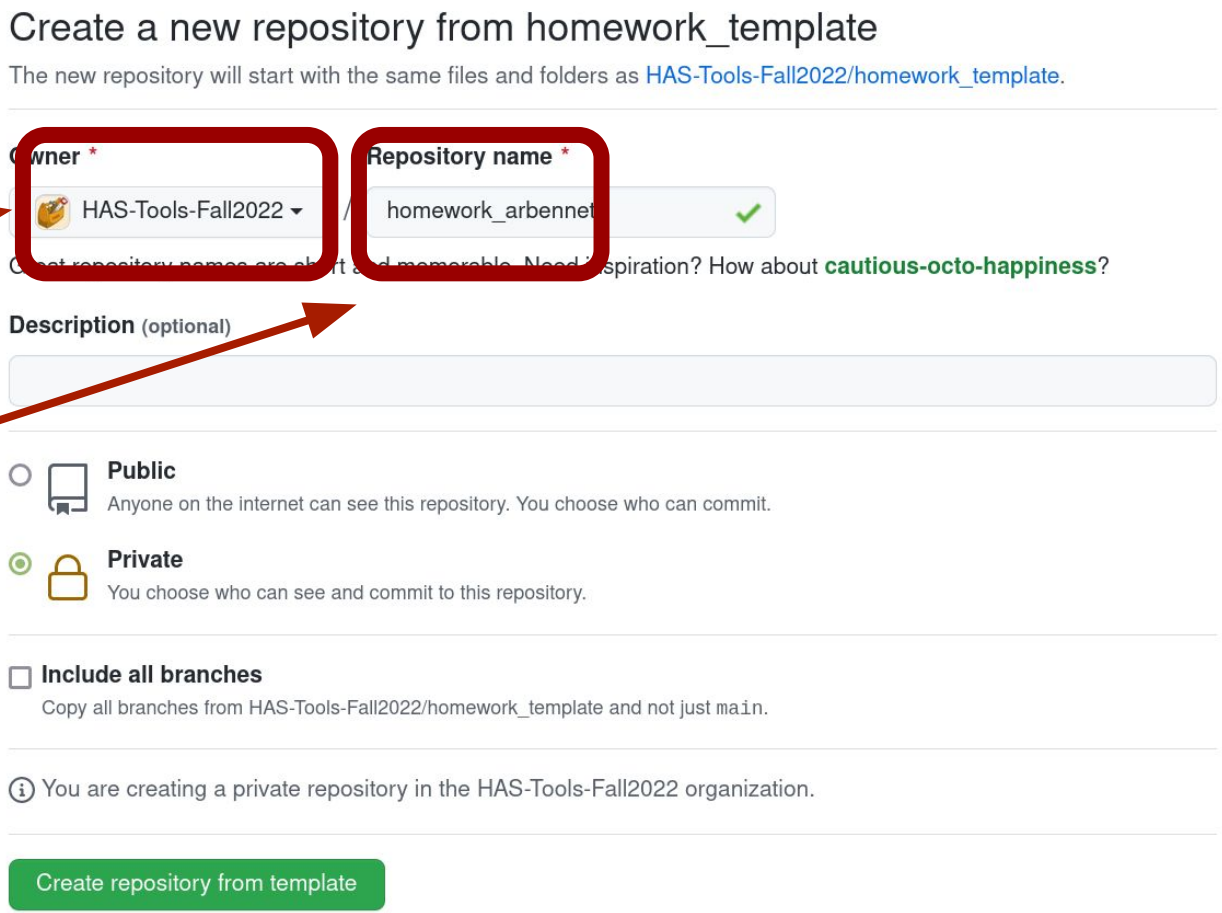
Change to HAS-Tools-Fall2022

Call it homework_YOURNAME



Change to HAS-Tools-Fall-2024

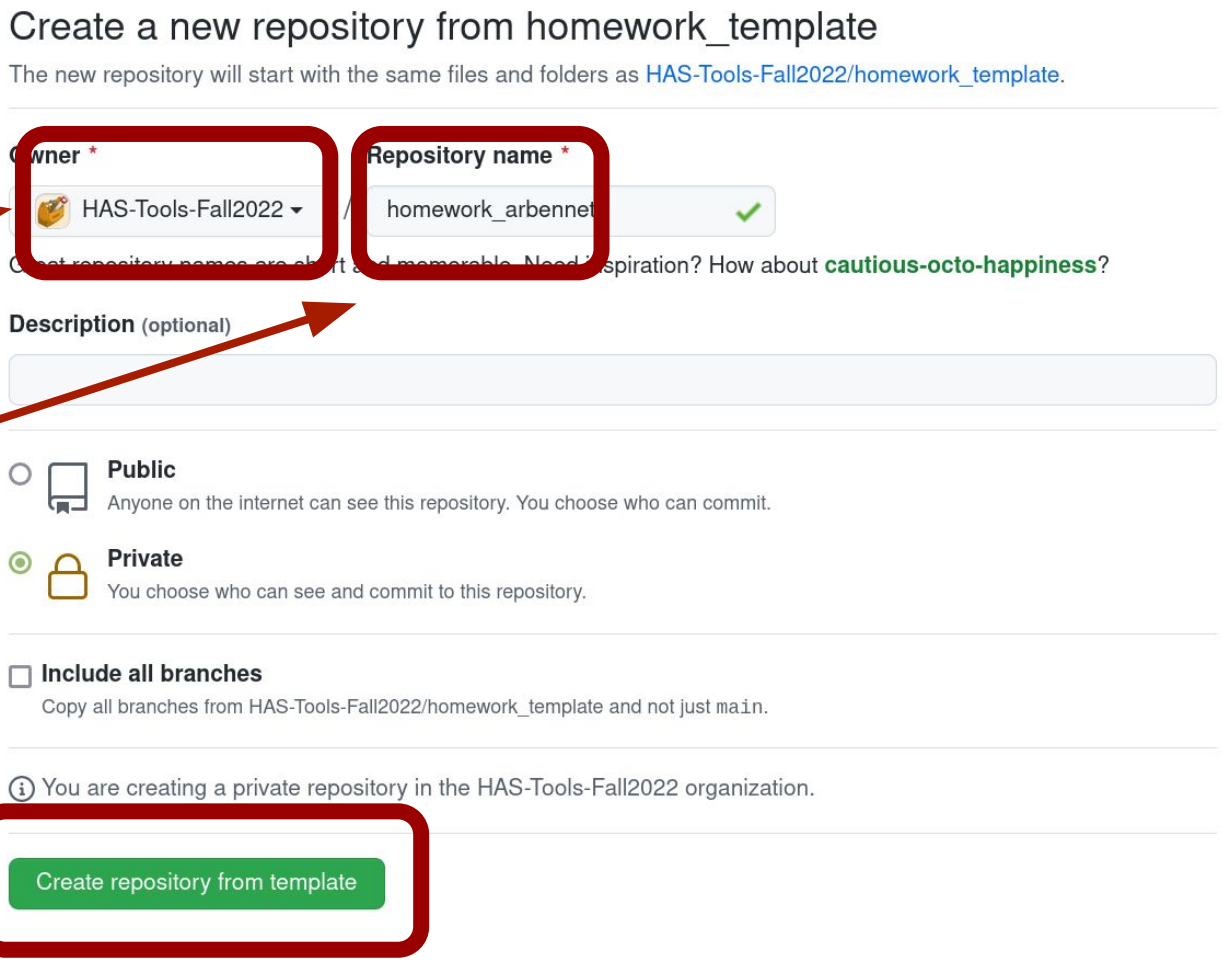
Call it homework_USERNAME



Change to HAS-Tools-Fall-2024

Call it homework_USERNAME

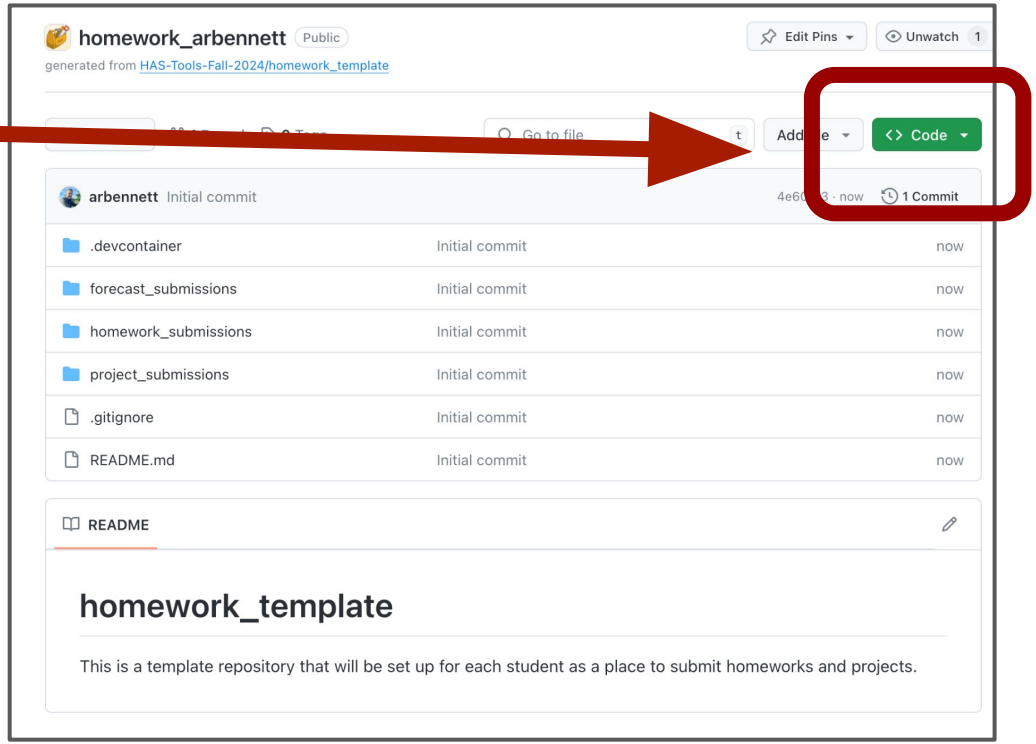
SMASH that create button



Your first codespace

CLICK HERE

Then select the “Codespace” tab and finally hit the “+” button



Whiteboard example time while the codespace boots up

Whiteboard example time while the codespace boots up

Once codespaces boots up, a quick tour of vscode and git

That’s all for today. Next class, we start python!

- Please make sure your GitHub account is in the class organization
- Also please make sure you get your GitHub education account set up
- Your first assignment: Make a modification to the README.md file in your homework repository to add the following line:
    - “The first assignment was to modify the readme to include this statement”
- Make sure that this change is “committed” and “pushed” from your “local” codespace back to the “remote” on GitHub so that you and I can see the change