



DSA101.1

Introduction to Data Analytics and Python

Prof. Alexandria Deliz Liang

Prof. Meredith C. Schuman, meredithchristine.schuman@uzh.ch

Dr. Domitille Coq-Etchegaray

Who am I?

Meredith C. Schuman (Merry)

Born in Minnesota, USA

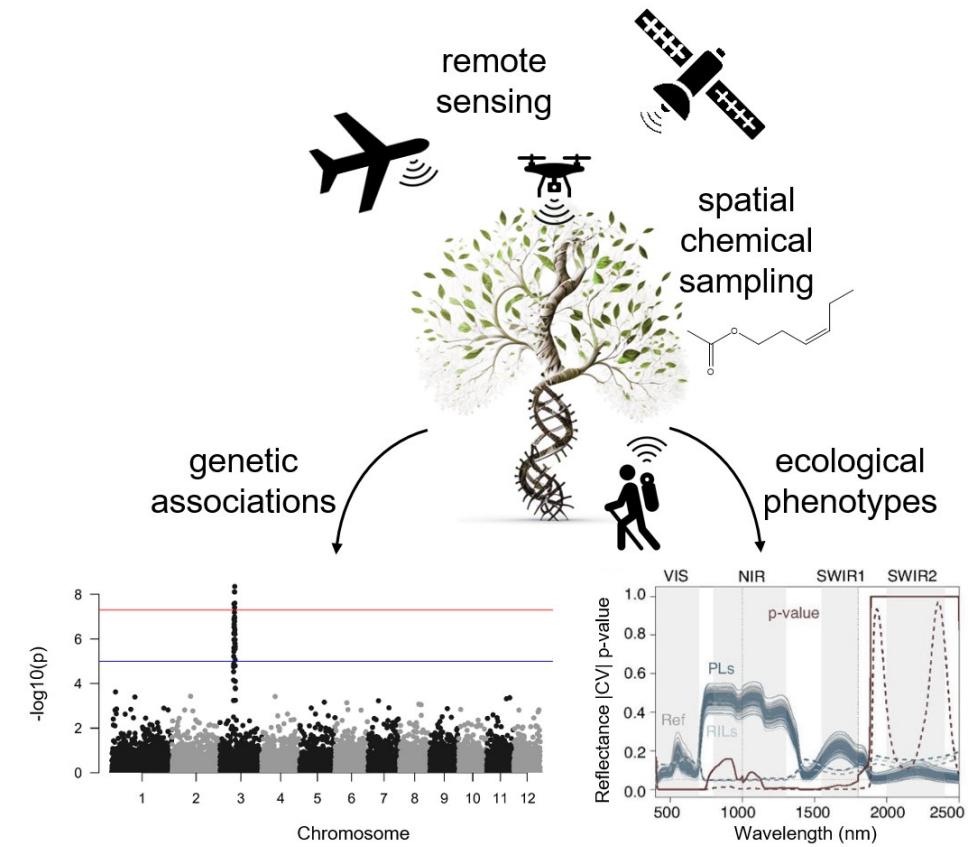
2012 PhD in Chemical Ecology, FSU,
Jena, DE

2012-2019 Group Leader, Max Planck
Institute for Chemical Ecology, Jena, DE

Since 2019 Assistant Professor for
Spatial Genetics, UZH, CH



We want to know *why* living things differ and
where those differences matter for ecology and for people.



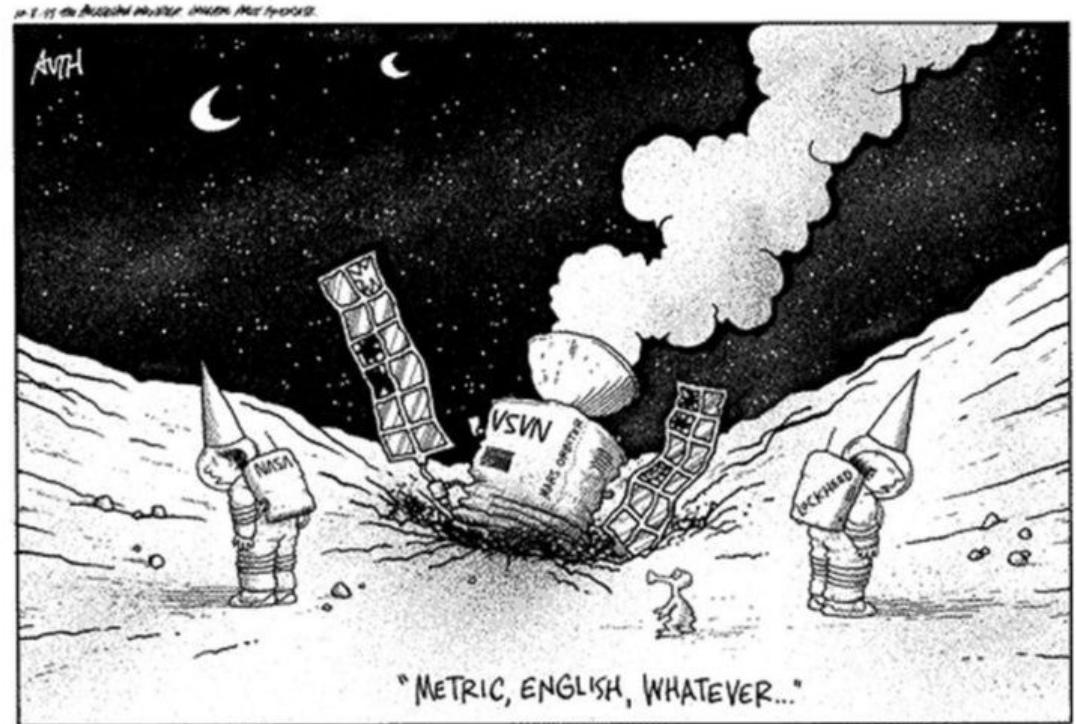
"tree DNA white background": gencraft, spectra: C. Li et al. (2023) CC-BY 4.0

Who are you?

Motivation



A Boeing Delta II 7425 expendable launch vehicle lifts off with NASA's Mars Climate Orbiter on Dec. 11, 1998
<https://www.simscale.com/blog/nasa-mars-climate-orbiter-metric/>



Motivation

The retracted paper analyzed excitatory synapse number and neurotransmitter release in mouse hippocampal neurons missing a presynaptic cell-adhesion molecule called neurexin-2. **Südhof**, professor of molecular and cellular physiology and neurosurgery at Stanford University, received the Nobel Prize for some of his previous work on neurexin proteins.

The authors retracted the study “because re-analysis of the original raw data for Figs. 2, 4 and 6 revealed that, although our analyses of the original data are supportive of the conclusions of the paper, unresolvable differences exist between these raw data and the published data source file that cannot be corrected by a simple erratum,” according to the 5 March [retraction notice](#).

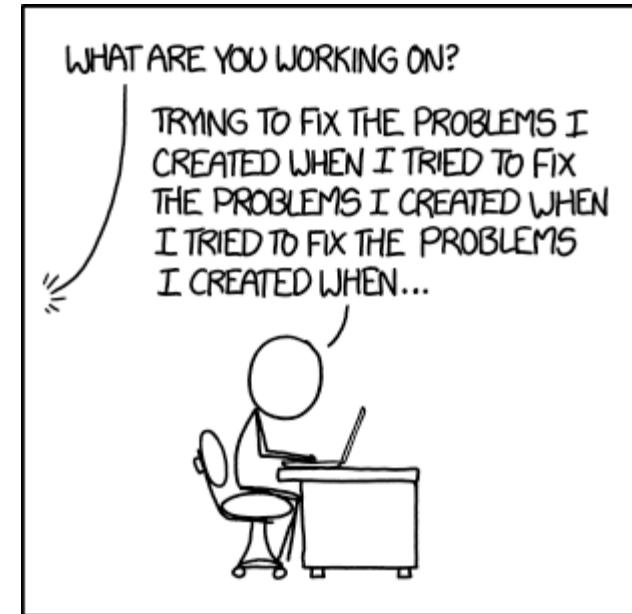
We want to solve a problem using a computer.

There will be many similar, but not identical problems.

Did we solve the original problem correctly?

What if...

- The data change?
- The assumptions change?
- We have 10x as much data? (Or 0.1x as much data?)
- We pass this on to a colleague, or to ourselves six months or six years from now?

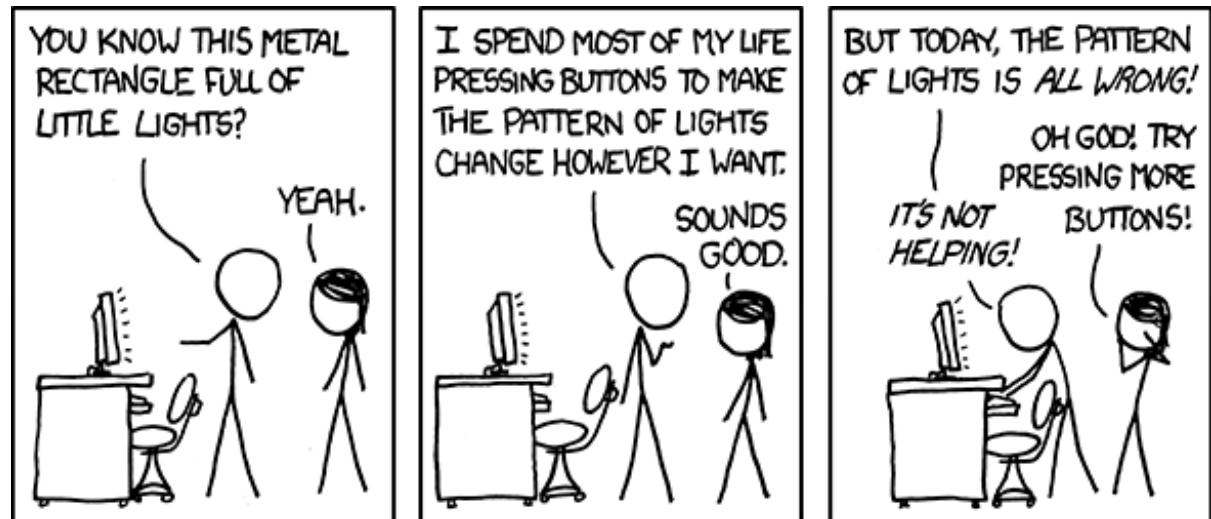


<https://www.xkcd.com/1739/> 

We want to solve a problem using a computer.

If we made a mistake, what happens?

- Program crashes?
- Everything is *increeeedibly* slow?
- It still works for *this specific case*?
- The output is a *little bit* wrong?
- The output is wrong *sometimes*?



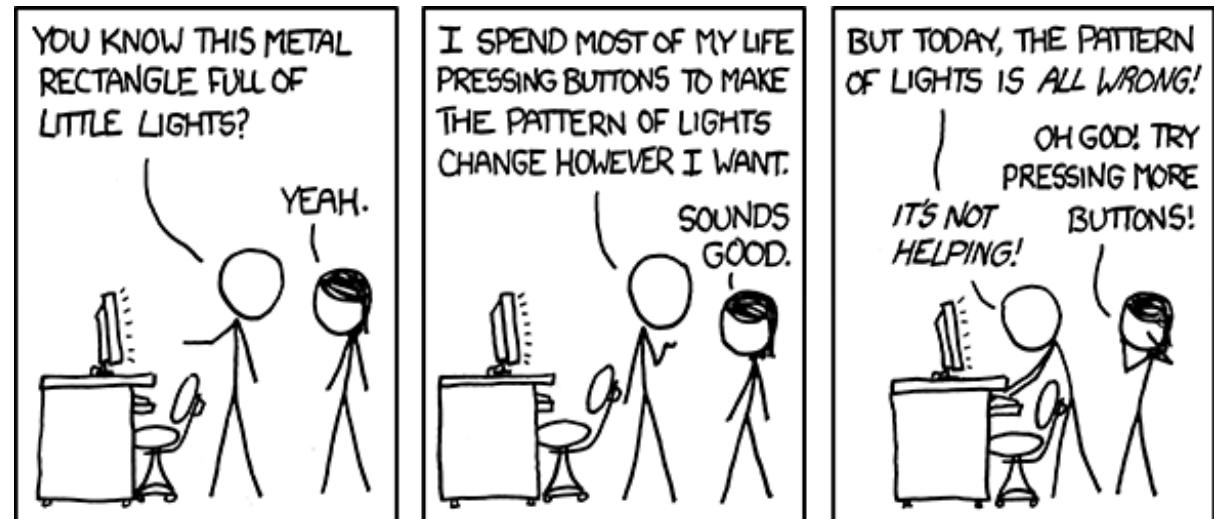
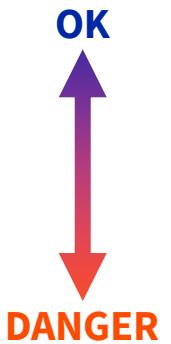
<https://xkcd.com/722/>

We want to solve a problem using a computer.

If we made a mistake, what happens?

Clear data science ranking:

- Program crashes?
- Everything is *increeeedibly* slow?
- It still works for *this specific case*?
- The output is a *little bit* wrong?
- The output is wrong *sometimes*?



<https://xkcd.com/722/> 🐍

LIKELIHOOD YOU WILL GET CODE WORKING
BASED ON HOW YOU'RE SUPPOSED TO INSTALL IT:



What is Git?

Git is a program for version control.

What is version control and why is it important?

- You change a file → change can be tracked
- A collaborator changes a file → changes can be merged
- You can go back and forth in time of your project
- You can branch out, test something and either integrate or revert changes
- **You can hardly break anything**



The program



GitHub

Free individual accounts,
North American interface
used worldwide

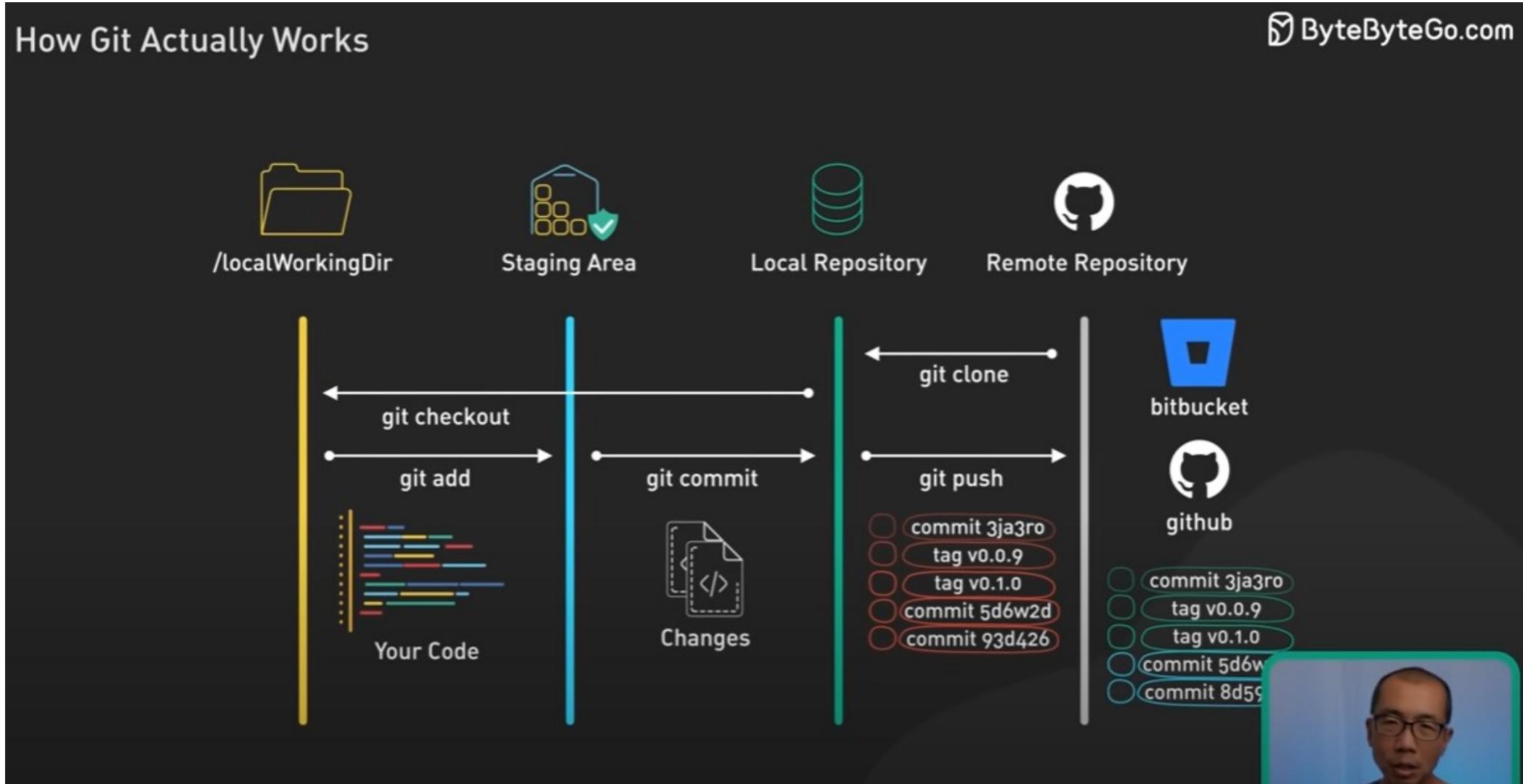


GitLab

Institutional accounts,
European interface

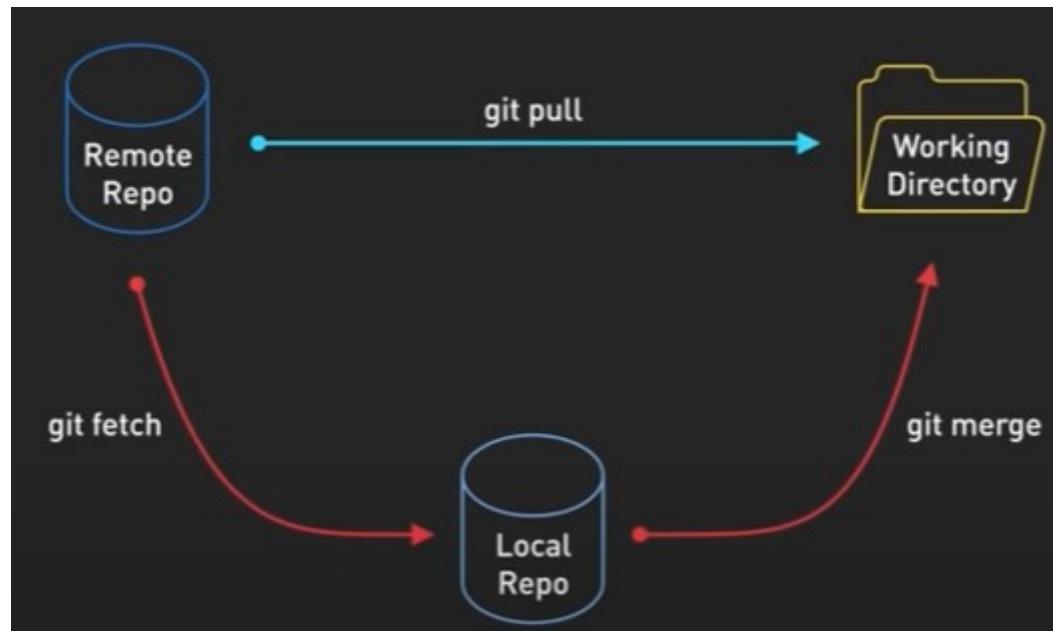
<https://docs.gitlab.com/ee/topics/git/>

How git actually works



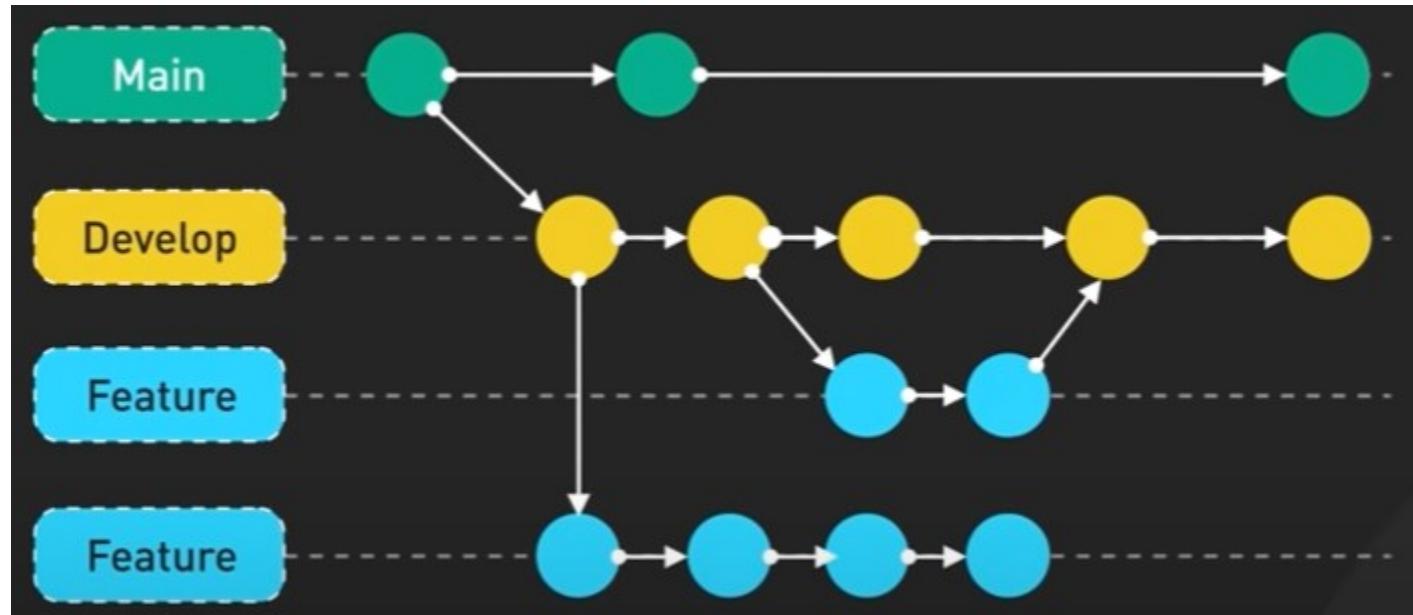
https://youtu.be/e9lNsKot_SQ?si=g0d_pQD2eUPpP8Ae

How git actually works



https://youtu.be/e9lnsKot_SQ?si=g0d_pQD2eUPpP8Ae

How git actually works



https://youtu.be/e9InsKot_SQ?si=g0d_pQD2eUPpP8Ae

Git resources

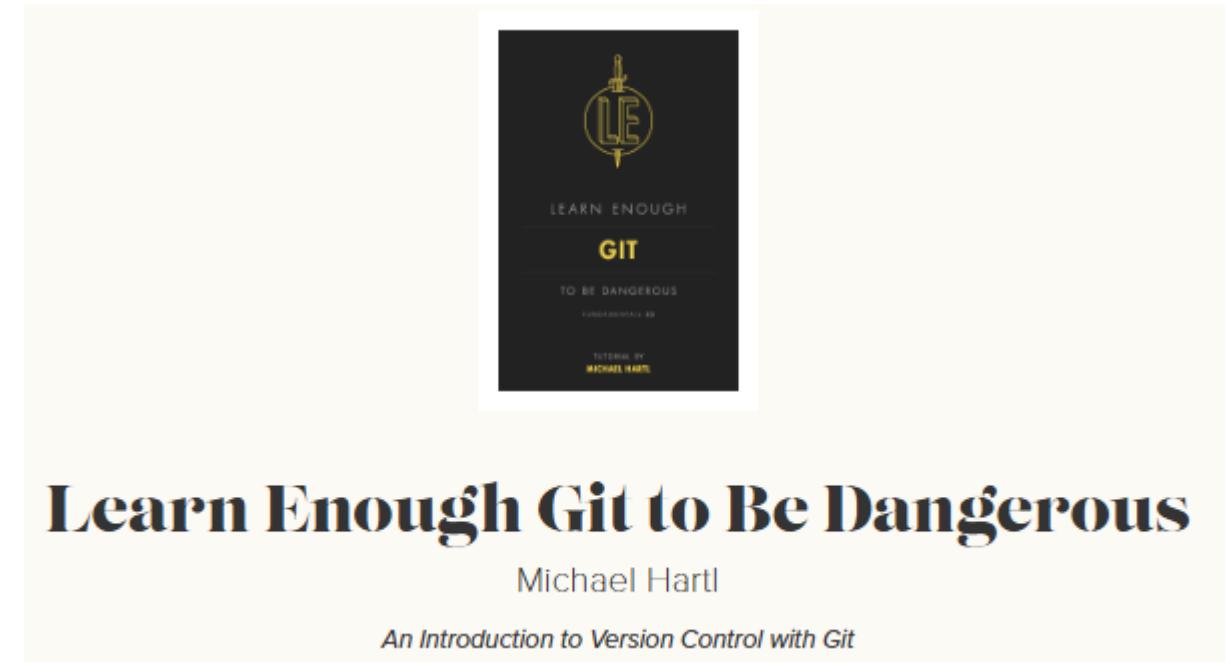
The Git Cheat Sheet

Basic guides

- Git - the simple guide, <https://rogerdudler.github.io/git-guide/>
- Git tutorial and quiz, <https://www.w3schools.com/git/>
- Github hello world guide, <https://docs.github.com/en/get-started/quickstart/hello-world>

Advanced guides

- Learn Enough Git to Be Dangerous,
https://www.learnenough.com/git-tutorial/getting_started



We will be working from a GitHub repo

The screenshot shows a GitHub repository page for 'mcschuman / DSA101'. The repository is public and contains 25 commits. The main branch is 'main'. The repository is described as 'Repository for DSA101 exercises (UZH DSA minor)'. It includes files like 'README', 'LICENSE', '.gitignore', 'exercise1', 'exercise2', 'exercise3', 'solutions', and 'collider'. The 'Code' tab is selected. The 'About' section provides details about the repository, including its license (AGPL-3.0), activity, and reporting options. The 'Releases' and 'Packages' sections are currently empty. The 'Languages' section shows Jupyter Notebook as the primary language (97.9%) and Python as a secondary language (2.1%).

mcschuman / DSA101 Public

Code Issues Pull requests 1 Actions Projects Security Insights

main 1 Branch 0 Tags Go to file Code

mcschuman Updated main ReadMe file f65016e · 35 minutes ago 25 Commits

collider Re-organized repository to allow different courses narratives 38 minutes ago

exercise1 Add files via upload last week

exercise2 Re-organized repository to allow different courses narratives 38 minutes ago

exercise3 Re-organized repository to allow different courses narratives 38 minutes ago

solutions Re-organized repository to allow different courses narratives 38 minutes ago

.gitignore Create .gitignore last week

LICENSE Initial commit last week

README.md Updated main ReadMe file 35 minutes ago

README AGPL-3.0 license

DSA101

Repository for DSA101 exercises (UZH DSA minor)

There are some exercises to get started and to introduce specific concepts, and a simulation to try out collaborative

About

Repository for DSA101 exercises (UZH DSA minor)

Readme AGPL-3.0 license Activity 1 star 0 watching 17 forks Report repository

Releases

No releases published

Packages

No packages published

Languages

Jupyter Notebook 97.9% Python 2.1%

<https://github.com/mcschuman/DSA101>

Let's start!

(1) Make a GitHub account
if you don't have one

The screenshot shows the GitHub signup page (github.com/signup) with a dark theme. On the left, there's a large "Create your free account" heading and a "See what's included" link. On the right, there's a "Sign up for GitHub" form with fields for Email*, Password*, Username*, and Your Country/Region*. There are also "Email preferences" checkboxes and a "Create account >" button. A small purple cat icon is visible at the bottom.

github.com/signup

Research communic... Think Check Attend - NotebookLM Adobe Acrobat

Already have an account? [Sign in →](#)

Create your free account

Explore GitHub's core features for individuals and organizations.

See what's included ▾

Sign up for GitHub

Continue with Google

Continue with Apple

or

Email*

Email

Password*

Password

Password should be at least 15 characters OR at least 8 characters including a number and a lowercase letter.

Username*

Username

Username may only contain alphanumeric characters or single hyphens, and cannot begin or end with a hyphen.

Your Country/Region*

Switzerland

For compliance reasons, we're required to collect country information to send you occasional updates and announcements.

Email preferences

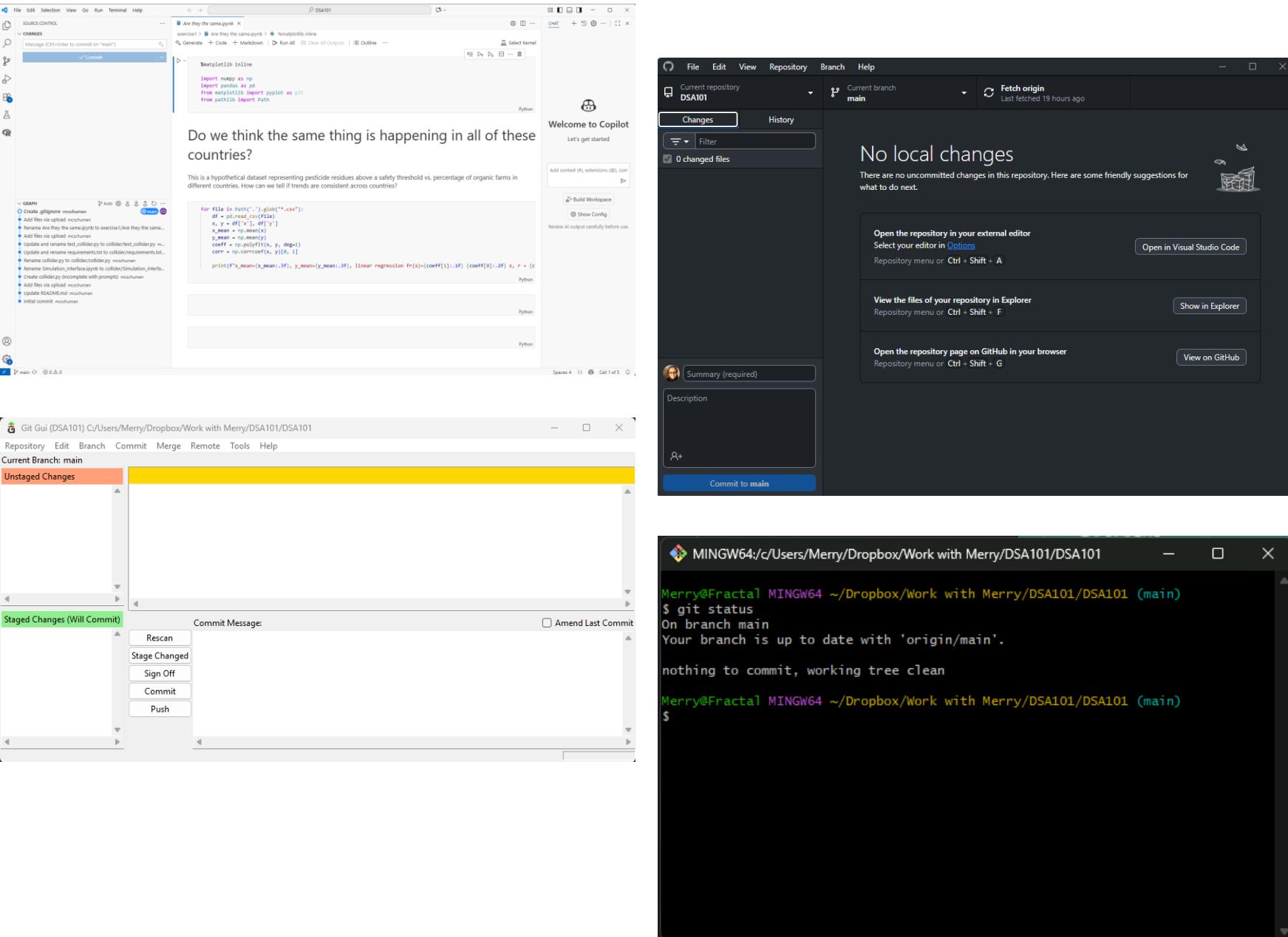
Receive occasional product updates and announcements

Create account >

By creating an account, you agree to the [Terms of Service](#). For more information about GitHub's privacy practices, see the [GitHub Privacy Statement](#). We'll occasionally send you account-related emails.

Let's start!

- (1) Make a GitHub account if you don't have one
- (2) Install Git: there are many options!

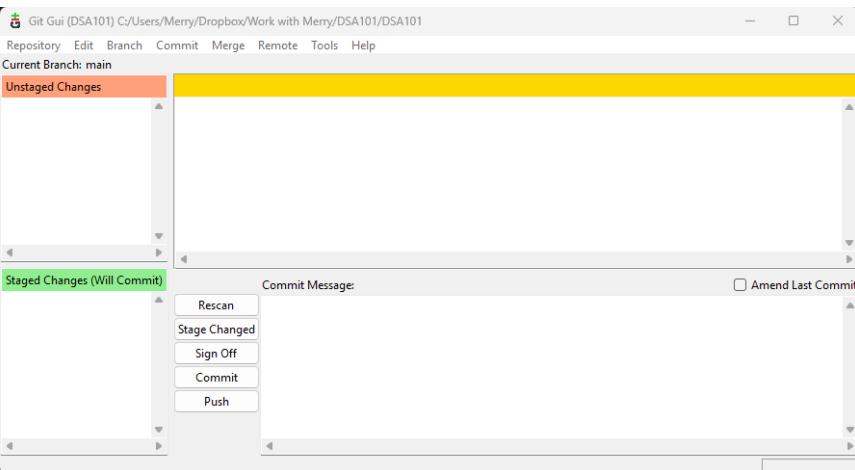
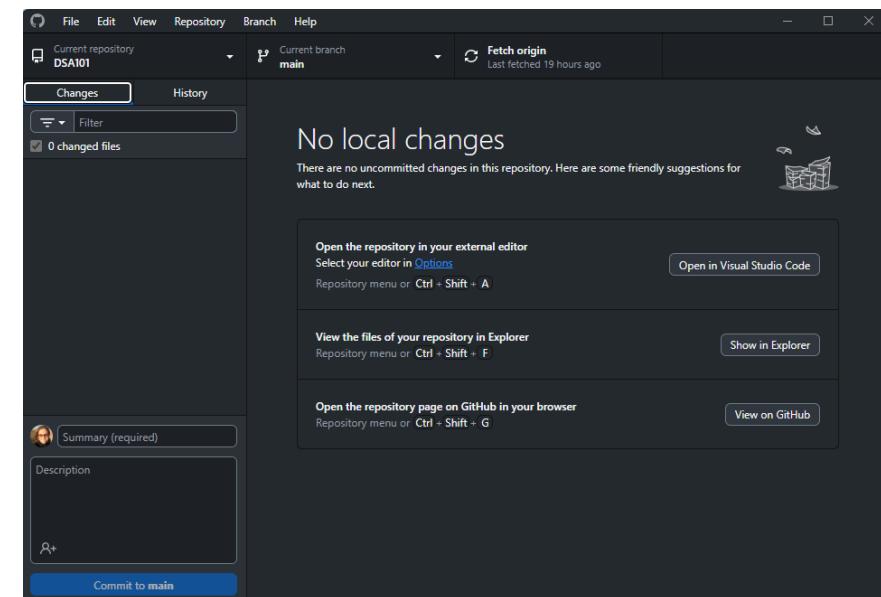
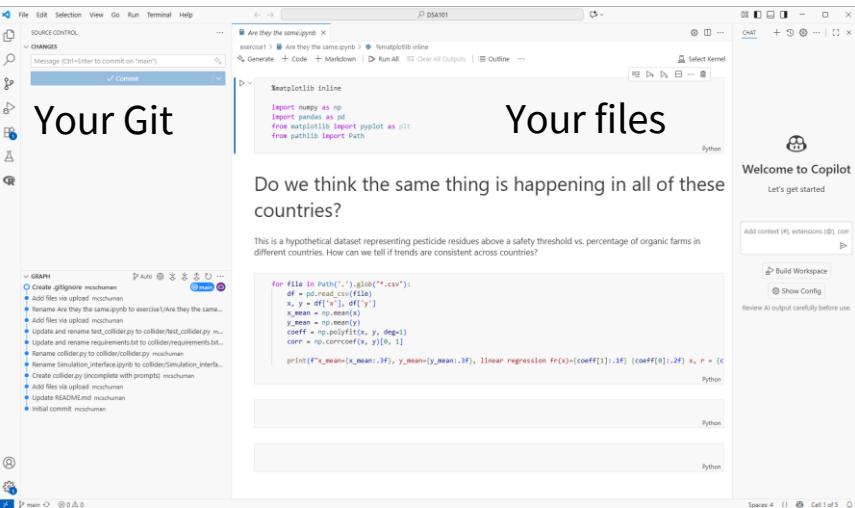


Let's start!

- (1) Make a GitHub account if you don't have one
- (2) Install Git: there are many options!

- **Easiest:** an IDE like Visual Studio Code: interface to interact with code and Git

Visual Studio Code or another IDE



A screenshot of a terminal window titled 'MINGW64:c/Users/Merry/Dropbox/Work with Merry/DSA101/DSA101'. The prompt is 'Merry@Fractal MINGW64 ~/Dropbox/Work with Merry/DSA101/DSA101 (main)\$. The user runs the command '\$ git status' which outputs:

```
$ git status
On branch main
Your branch is up to date with 'origin/main'.

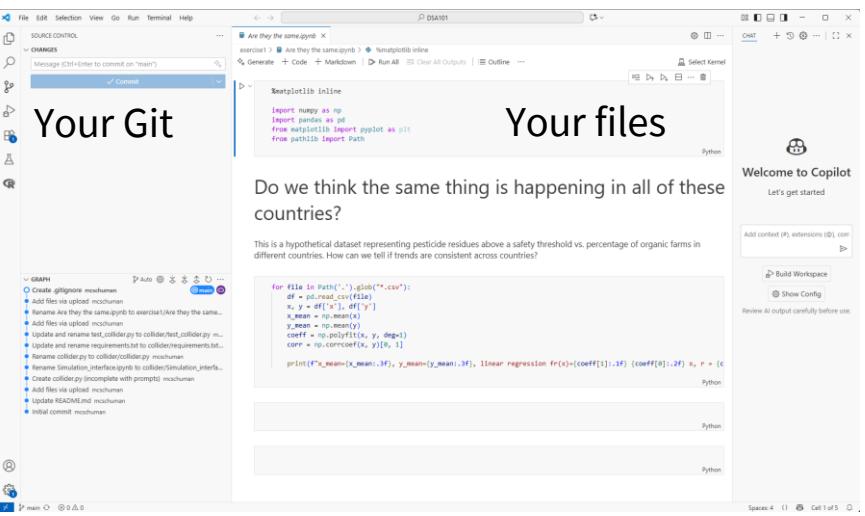
nothing to commit, working tree clean
```

Let's start!

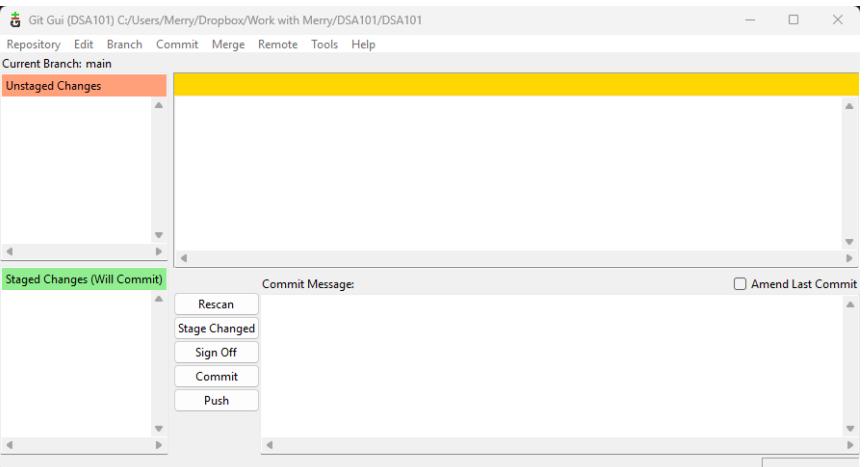
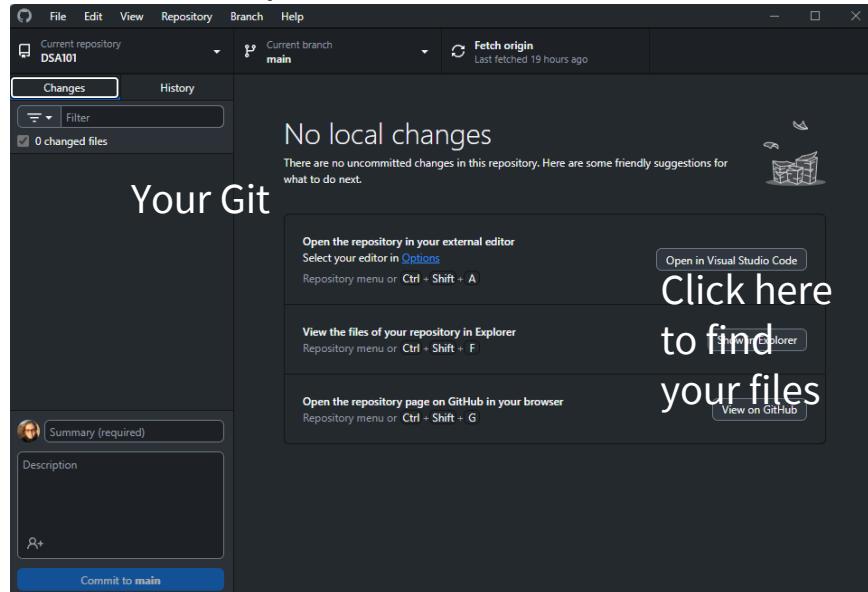
- (1) Make a GitHub account if you don't have one
- (2) Install Git: there are many options!

- **Easiest**: an IDE like Visual Studio Code: interface to interact with code and Git
- **Second easiest**: GitHub Desktop, a GUI that prompts you to use Git correctly

Visual Studio Code or another IDE



GitHub Desktop



A screenshot of a terminal window titled 'MINGW64:c/Users/Merry/Dropbox/Work with Merry/DSA101/DSA101'. The command '\$ git status' was run, and the output is:
Merry@Fractal MINGW64 ~/Dropbox/Work with Merry/DSA101/DSA101 (main)
\$ git status
On branch main
Your branch is up to date with 'origin/main'.
nothing to commit, working tree clean
Merry@Fractal MINGW64 ~/Dropbox/Work with Merry/DSA101/DSA101 \$

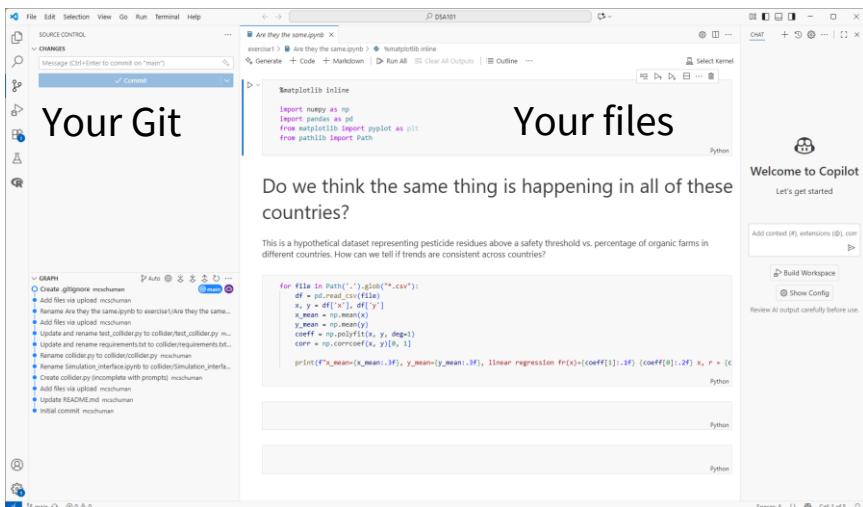
Let's start!

(1) Make a GitHub account if you don't have one

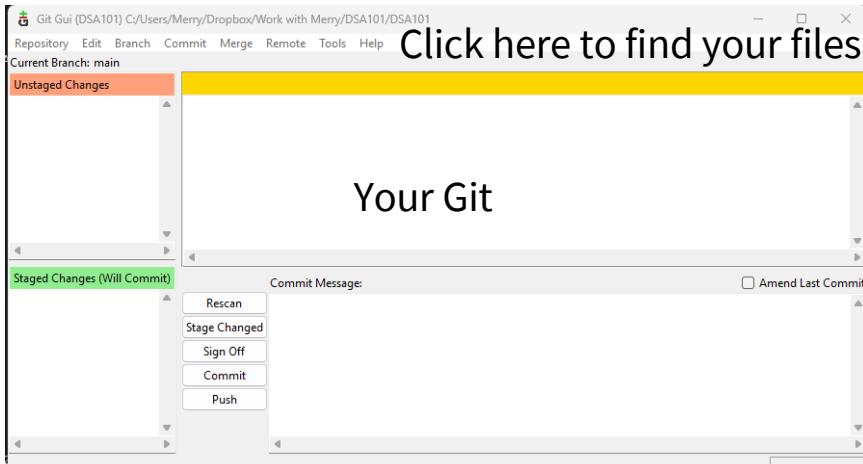
(2) Install Git: there are many options!

- **Easiest**: an IDE like Visual Studio Code: interface to interact with code and Git
- **Second easiest**: GitHub Desktop, a GUI that prompts you to use Git correctly
- **Develop your computer skills (still fairly easy)**: install with Git installer, then interact with Git using Git GUI or Git Bash (or Terminal)

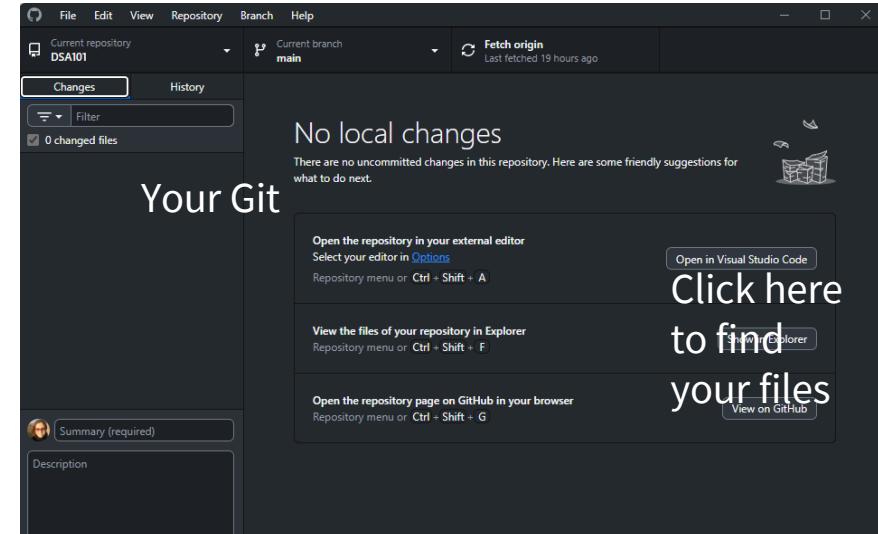
Visual Studio Code or another IDE



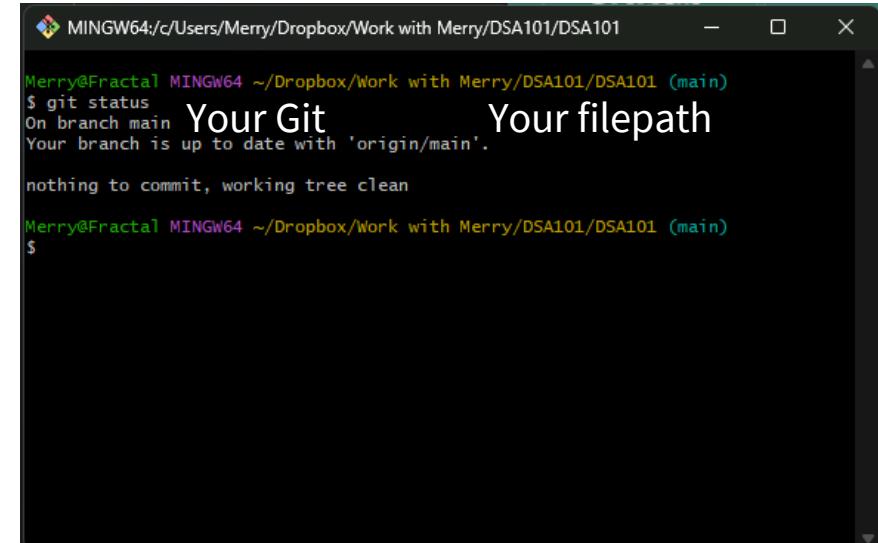
Git GUI



GitHub Desktop



Git Bash

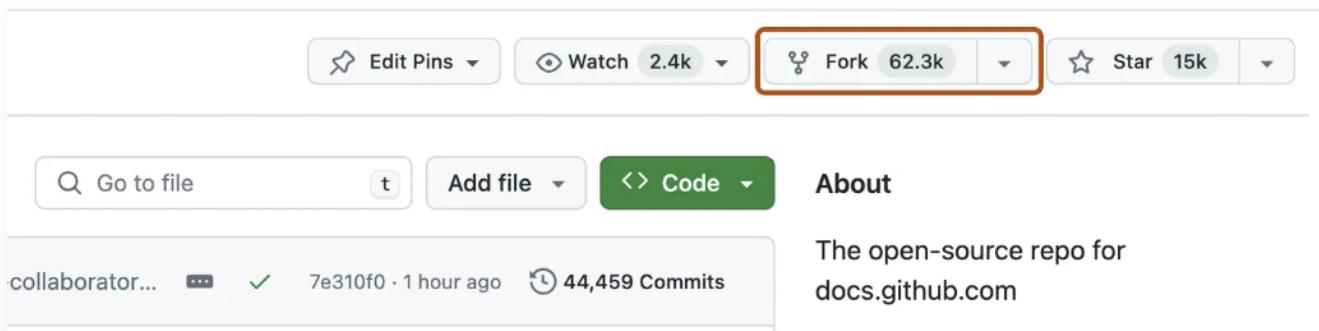


Let's start!

(1) Make a GitHub account
if you don't have one

(2) Install Git: there are
many options!

(3) Fork the DSA101
repository



- 3 Under "Owner," select the dropdown menu and click an owner for the forked repository.
- 4 By default, forks are named the same as their upstream repositories. Optionally, to further distinguish your fork, in the "Repository name" field, type a name.
- 5 Optionally, in the "Description" field, type a description of your fork.
- 6 Optionally, select **Copy the DEFAULT branch only**.

For many forking scenarios, such as contributing to open-source projects, you only need to copy the default branch. If you do not select this option, all branches will be copied into the new fork.

- 7 Click **Create fork**.

<https://docs.github.com/en/pull-requests/collaborating-with-pull-requests/working-with-forks/fork-a-repo>

Let's start!

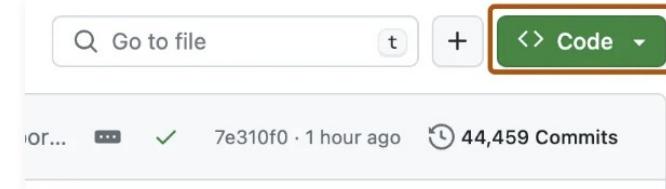
(1) Make a GitHub account
if you don't have one

(2) Install Git: there are
many options!

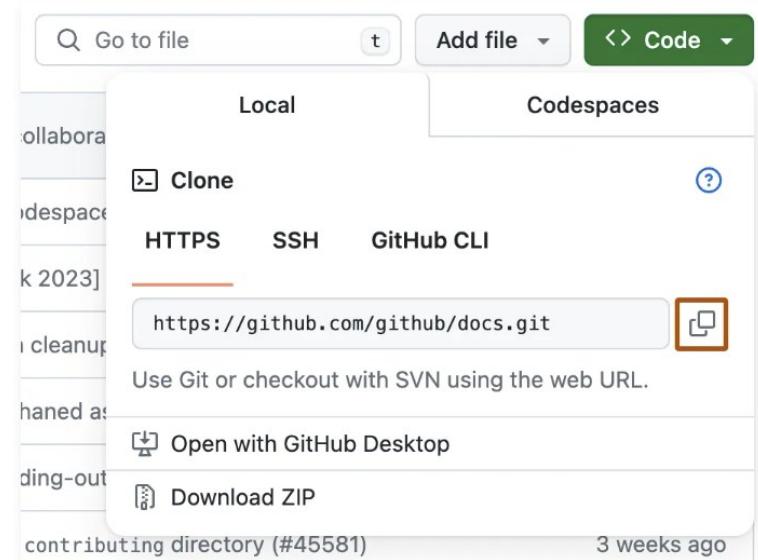
(3) Fork the DSA101
repository

(4) Clone your fork

I expect you to
understand how Git
works at this level for the
exam.



- ③ Copy the URL for the repository.
- To clone the repository using HTTPS, under "HTTPS", click .
 - To clone the repository using an SSH key, including a certificate issued by your organization's SSH certificate authority, click **SSH**, then click .
 - To clone a repository using GitHub CLI, click **GitHub CLI**, then click .



<https://docs.github.com/en/pull-requests/collaborating-with-pull-requests/working-with-forks/fork-a-repo>

Next: try exercise1

Are they the same?

We have a fictitious dataset from 12 different countries documenting the relationship between proportion of organic farms, and proportion of test sites in which pesticide residues are above a regulatory threshold.

You will characterize this relationship for each country and answer the question, is the relationship essentially the same across all countries in our dataset?

I expect you to understand how to look at data to answer specific questions for the exam.



<https://4.bp.blogspot.com/-1gcbXO2ePh4/UY8Vzp5khqI/AAAAAAAABk/i6HUCVq43GE/s1600/pesticide+spraying+man.jpg>