Intro to coding and Git

Raphael Eisenhofer

2022_03_03

Outline for today:

- 1. What is coding?
- 2. Why is coding useful?
- 3. What is Git/GitHub?
- 4. Getting started with Git/GitHub

1. What is coding?

Q: What is coding to you?

Instructions for computers

- Machine code: 01010001 01100110 10100011 10011010

Instructions for computers

- Machine code: 01010001 01100110 10100011 10011010

- Human-readable programming languages (R, Python, etc.) are translated to machine code

Instructions for computers

- Machine code: 01010001 01100110 10100011 10011010

- Human-readable programming languages (R, Python, etc.) are translated to machine code

- In this way, coding is a form of communication with computers!

2. How is coding useful?

What computers are good at

- **Processing** things VERY fast (GHz = 1,000,000,000 / second)

What computers are good at

- **Processing** things VERY fast (GHz = 1,000,000,000 / second)

- **Automation** (able to perform exact tasks many times without errors)

What computers are good at

- **Processing** things VERY fast (GHz = 1,000,000,000 / second)
- **Automation** (able to perform exact tasks many times without errors)
- **Storing and retrieving data**
  - >1,000 books per GB of computer storage

Examples of how coding can be useful

- Moving/editing/renaming hundreds/thousands of files

Examples of how coding can be useful

- Moving/editing/renaming hundreds/thousands of files

- Automating workflows/analyses/repetitive tasks

Examples of how coding can be useful

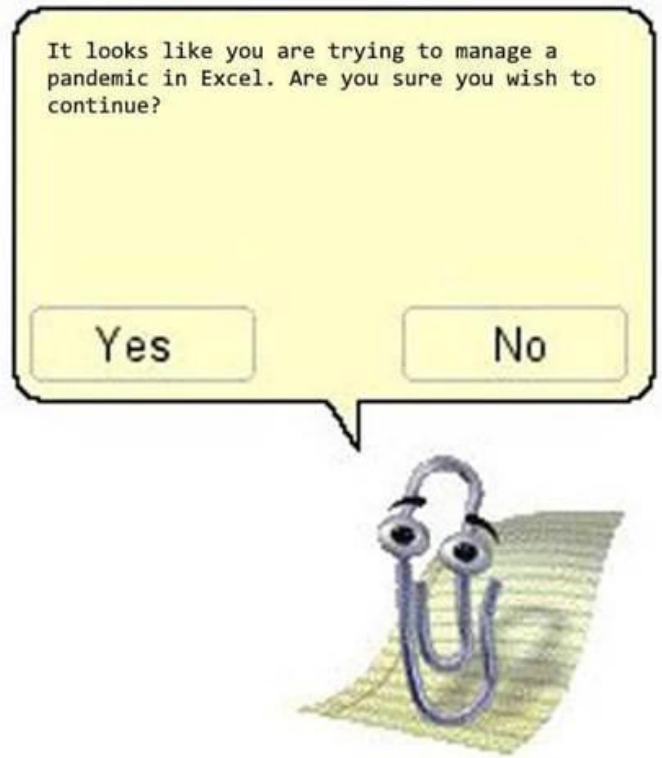
- Moving/editing/renaming hundreds/thousands of files
- Automating workflows/analyses/repetitive tasks
- Ensuring analyses can be reproduced by others

Examples of how coding can be useful

- Moving/editing/renaming hundreds/thousands of files
- Automating workflows/analyses/repetitive tasks
- Ensuring analyses can be reproduced by others
- Dealing with large datasets (or excel files!)

Perils of excel for large data

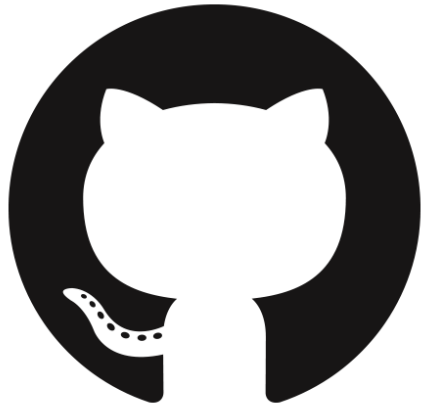
- Study looked at >10,000 genetics papers published between 2014-2020 – found **30%** of studies had data that was falsely autocorrected by excel! (https://doi.org/10.1371/journal.pcbi.1008984)



3. What is Git/GitHub?

Git and GitHub

- Git is a open-source version control system



- Records:
  - What?
  - When?
  - Why?

- GitHub is a online service provider of git (+ extras)

- A really good way of managing software development projects (or research projects in general!)

Examples:

- Reproducible code for a scientific paper: https://github.com/EisenRa/2020_SHNW_Faecal_16S
- Hosting/development of software: https://github.com/wwood/CoverM
- Misc. projects/groups: https://github.com/EisenRa/2022_Adelaide_Code_Club

4. Getting started with Git/GitHub

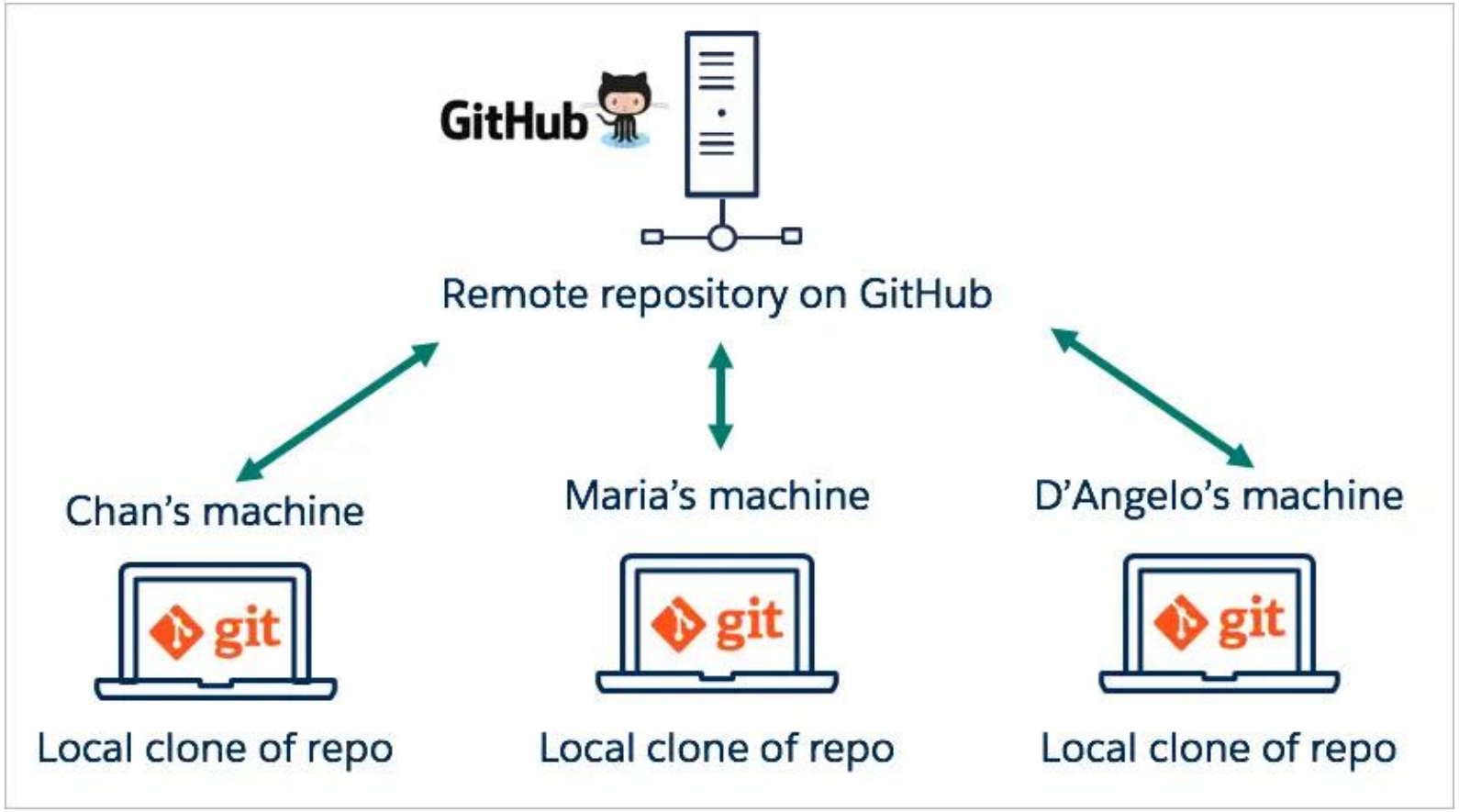
The command line

Manoeuvring around the command line

- **pwd** (‘**p**ath to **w**orking **d**irectory’ prints where you are o the computer)
- **ls** (**l**ist**s** the contents of the directory/folder you’re in)
- **cd** (‘**c**hange **d**irectory’ changes your current directory

Repositories

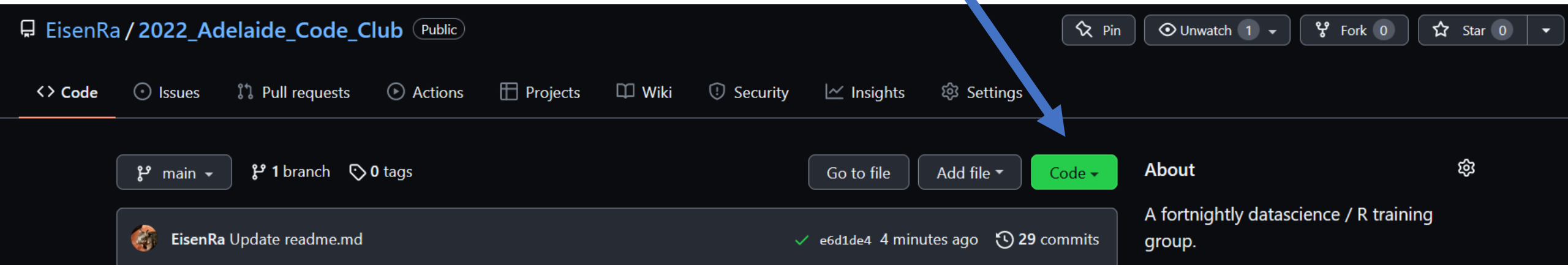
- A repository is where a given project/software is stored (analogous to a folder)



Cloning a repository

- To clone it, type the following:

git clone https://github.com/***user/repositoryname***.git



Creating your own repositories

- Plenty of great guides out there, so if you're interested, follow one of these:

- From scratch: https://docs.github.com/en/get-started/quickstart/create-a-repo

- Already have files/folders/project that you want to push to GitHub remote: https://docs.github.com/en/get-started/importing-your-projects-to-github/importing-source-code-to-github/adding-an-existing-project-to-github-using-the-command-line

Summary

- Knowing the basics of coding can save you a lot of time/pain

- Git/GitHub are fantastic tools for managing projects/research
  - Collaborating with others
  - Ensuring reproducibility (yourself and for others)