

Intro to RStudio and GitHub

Week 1 (10/1/25)



Artwork by Allison Horst

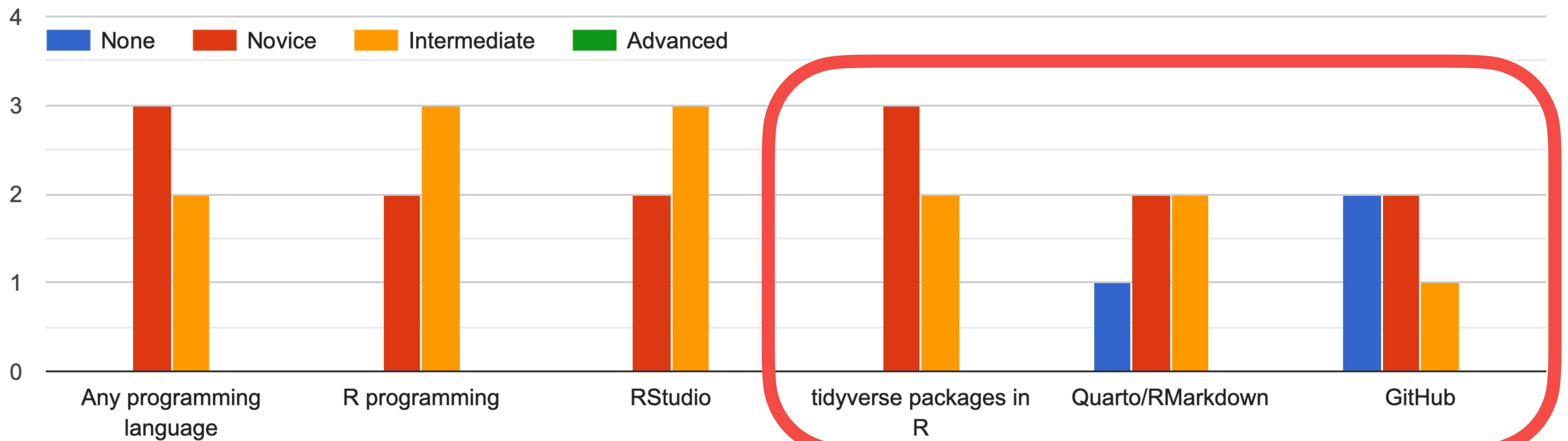
Stepfanie M. Aguillon

Outline of today's class

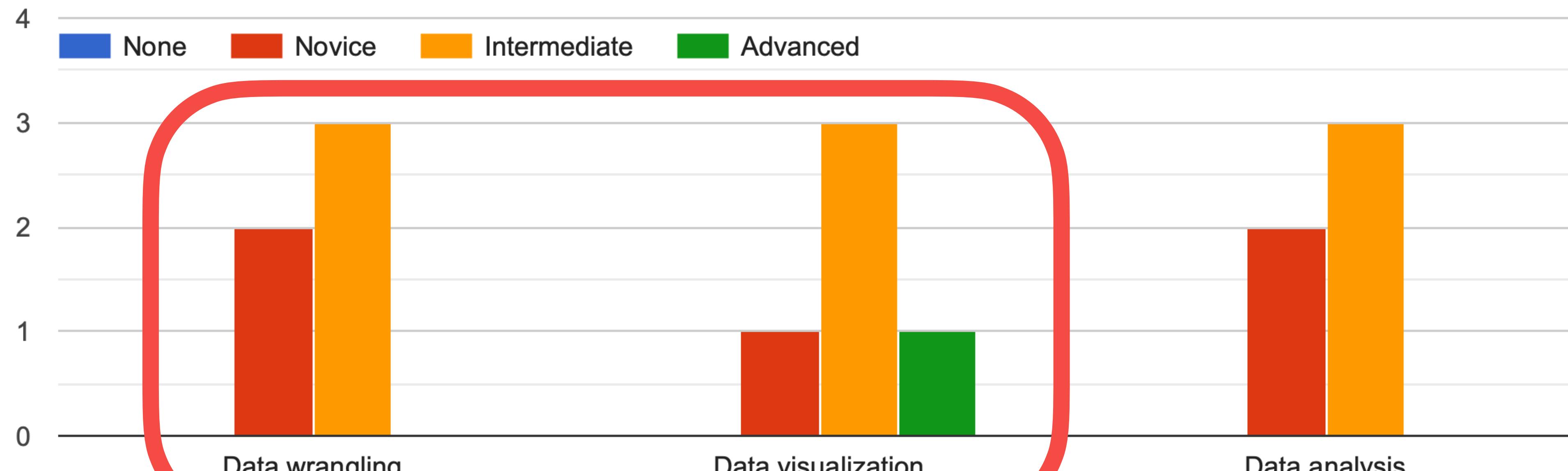
- Course logistics
- What is a reproducible workflow?
- Intro to RStudio and the tidyverse
- Week 1 problem set (*this week is a bit unusual)

Course Logistics

Please describe your experience level with...



Please describe your experience level with the following R tasks...



rewarding

absorbing impatient

intimidating

motivated

automatic

curious
frustrating
confusing
opaque useful
doable

Course logistics

- Each class session will be split up into: lecture and coding activities
- You'll then get to practice these skills in the weekly problem sets
- The course website on GitHub will have slides, pre-class readings, and post-class problem sets
- The Slack channel for the course will be our primary mode of communication (#eeb201-fall2025)

Course logistics

To receive a satisfactory grade...

1. **Attendance** - attend 9/10 lectures
2. **Problem Sets** - submit 7/10 problem sets on time (Wed 10pm of the following week)

LLM/AI policy

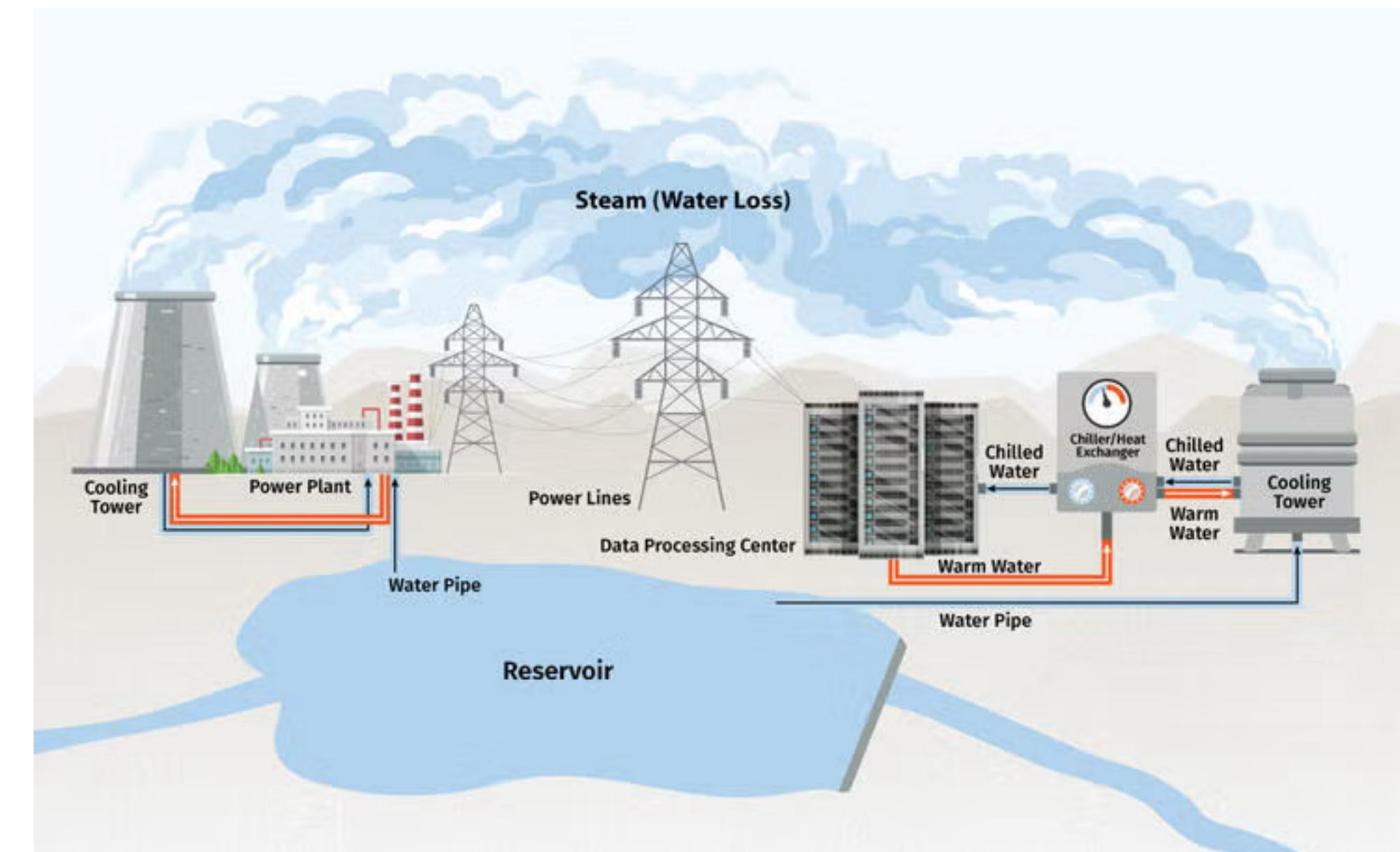
As the goal of our course is for you to practice your coding skills and build a solid foundation in R programming, I ask that you do not use LLMs/AI for any aspect of this course.

Generative AI is not a search engine.

They're predictive models that guess what word is next. That means they're wrong... A LOT.

Generative AI is bad for the environment!

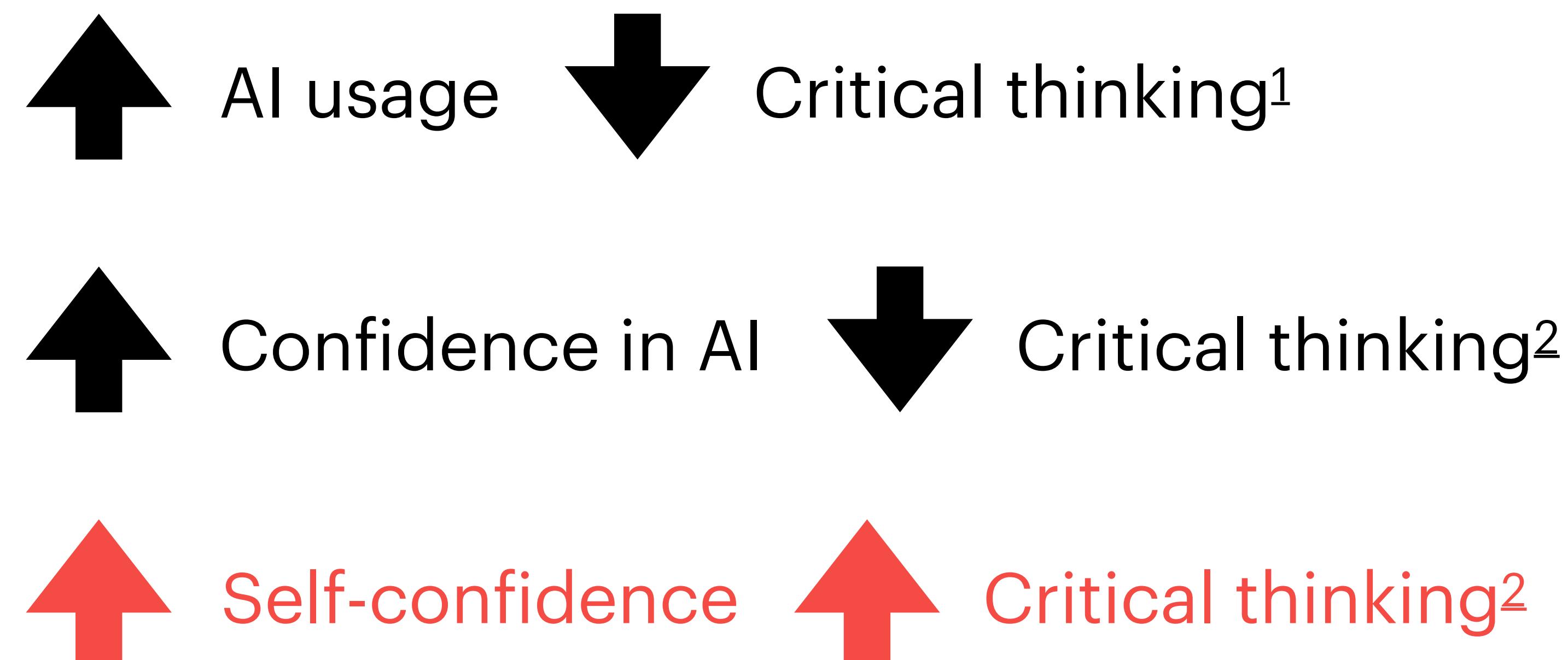
This is particularly true for water resources¹ and air quality in low-income communities².



LLM/AI policy

As the goal of our course is for you to practice your coding skills and build a solid foundation in R programming, I ask that you do not use LLMs/AI for any aspect of this course.

**Generative AI
impedes your
critical thinking
skills!**



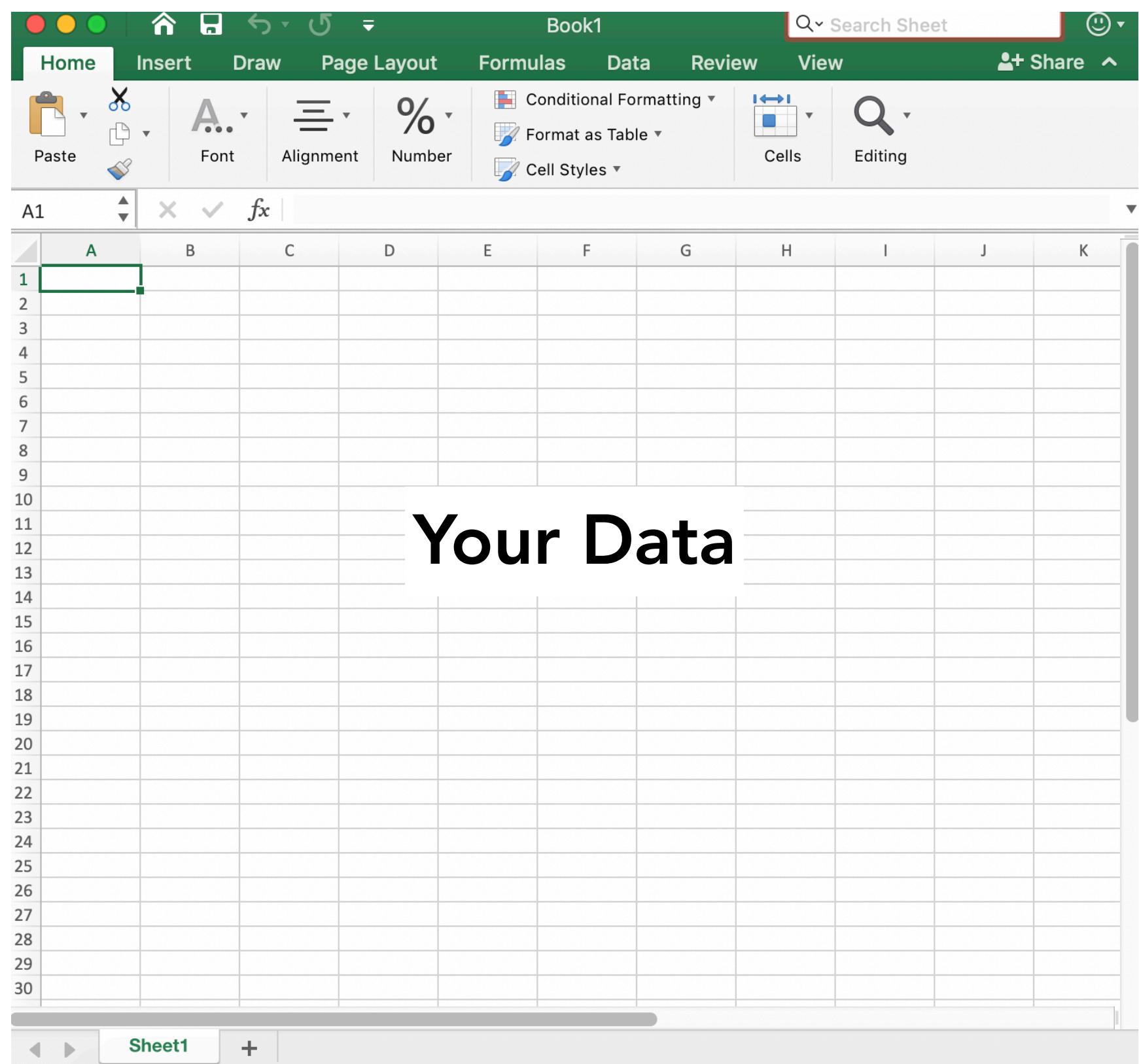
Reproducible Workflows

What is a reproducible workflow?

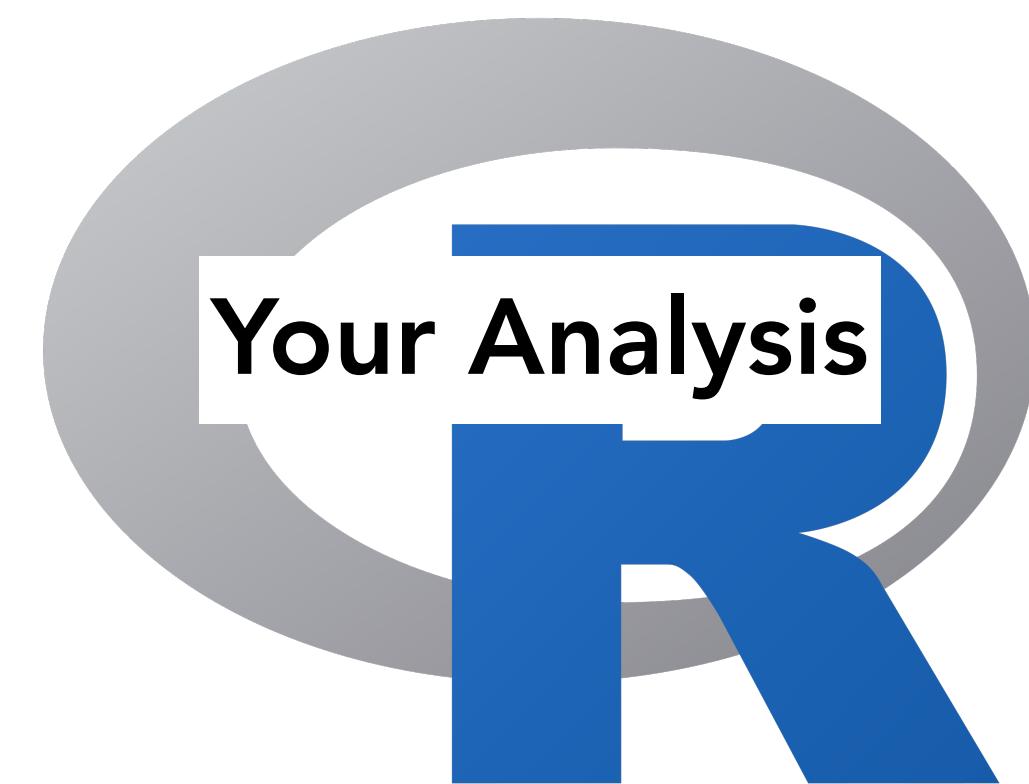
“Reproducible workflows consist of three components: a **fully scriptable statistical programming environment** (such as R or Python), **reproducible analysis** (first described as literate programming), and **version control** (commonly implemented using GitHub).”

Why make a reproducible workflow?

Spreadsheet Program



Statistics Program

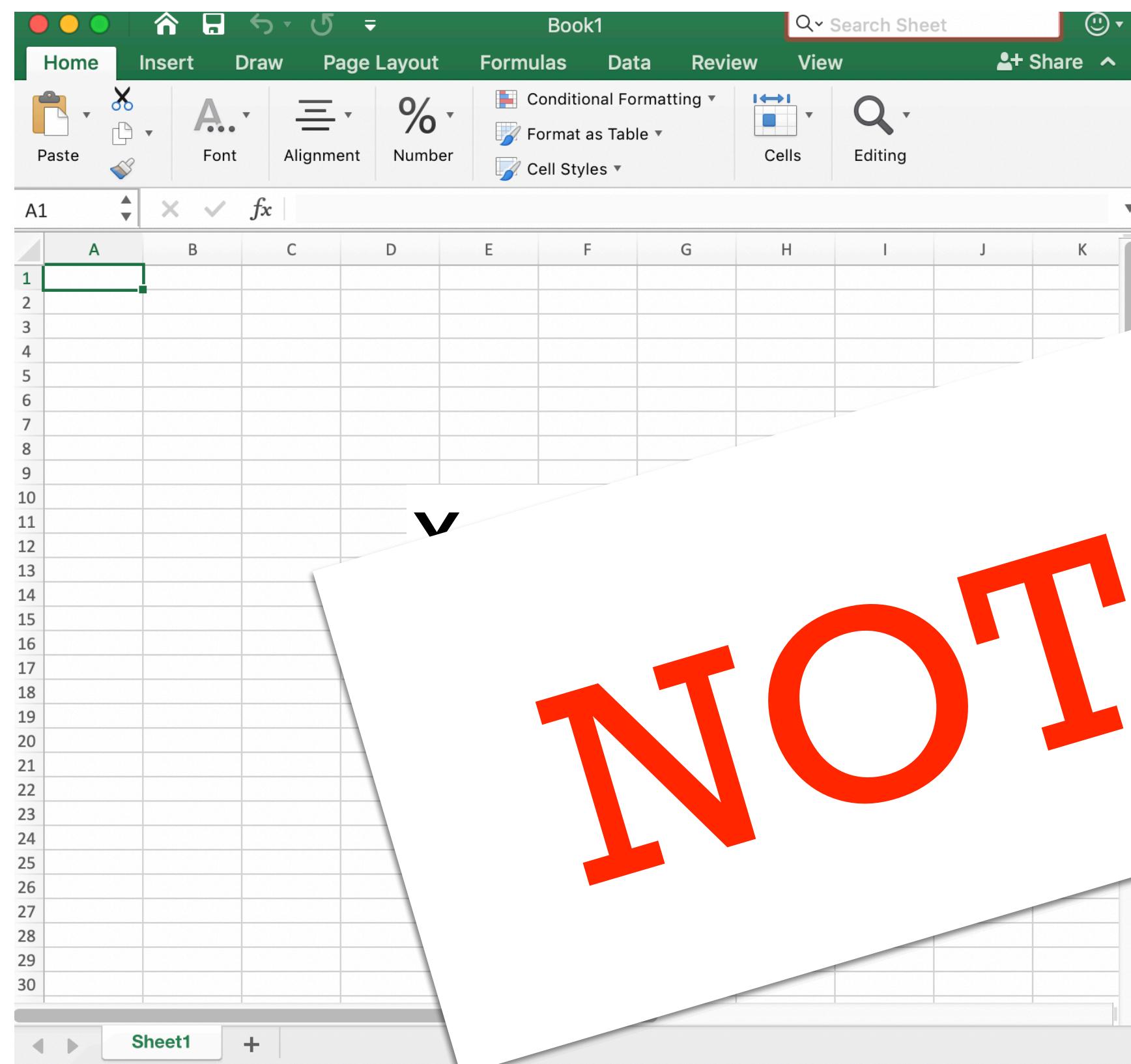


Published Paper



Why make a reproducible workflow?

Spreadsheet Program



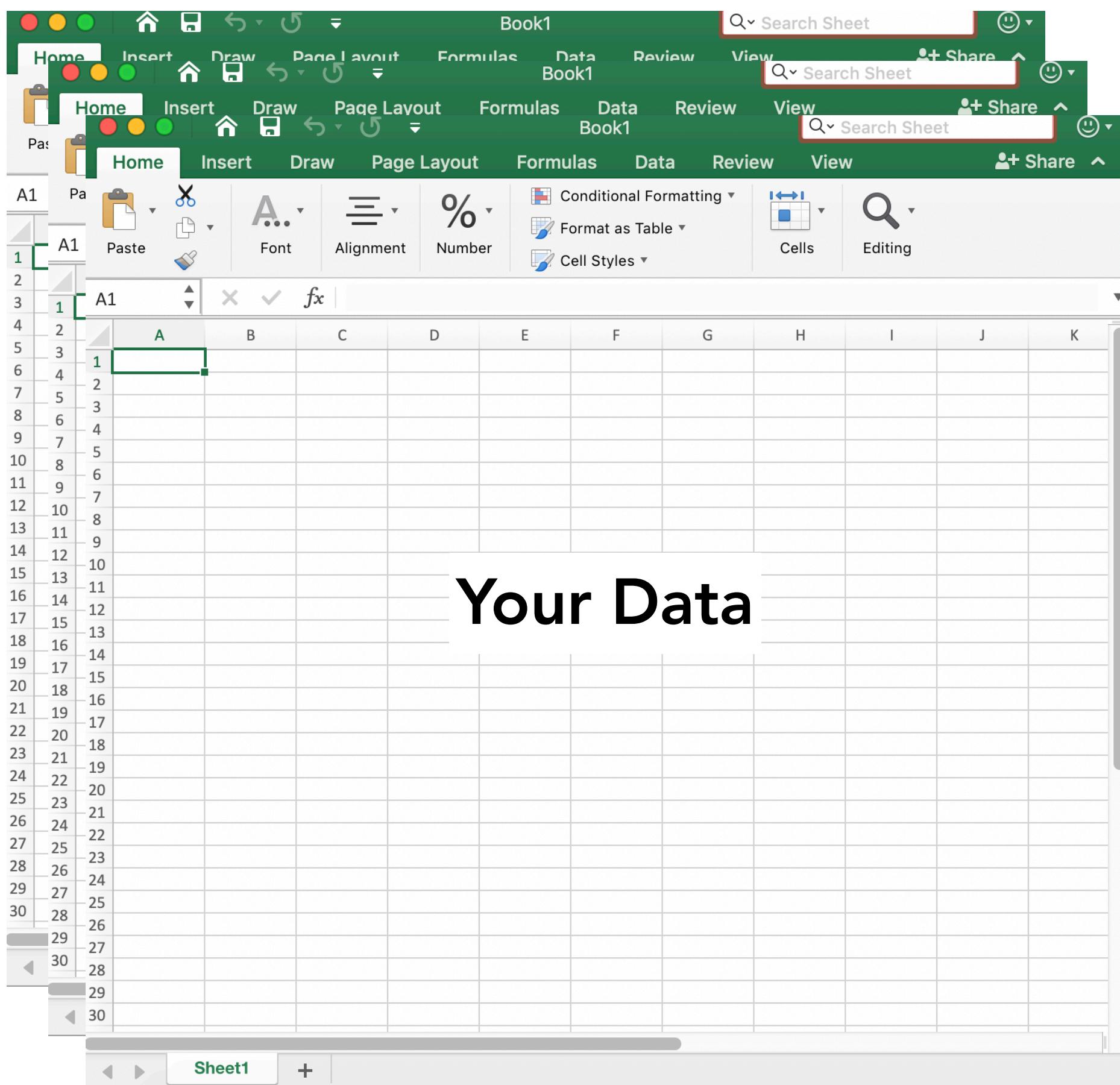
NOT REALISTIC!

Paper

Your Results

Why make a reproducible workflow?

Spreadsheet Program



Published Paper



What is a reproducible workflow?

Your Goal: Make an analysis workflow
that ~~someone else~~ could use to produce
the same result.



YOU, 6 months from now!

This is all part of “Open Science”

Making scientific research and data available and freely accessible to everyone.

>WHY OPEN CODE IN RESEARCH?



RESEARCHERS STUDENTS

- Learning new computational methods
- Sharing verified resources
- Collaborating efficiently
- Reproducing statistical analyses
- Improving employability

FUNDERS

- Reducing costs
- Improving research efficiency
- Increasing transparency and integrity
- Facilitating project evaluations
- Showcasing best projects

JOURNALS PUBLISHERS

- Increasing impact of papers
- Improving journal's metrics/ranking
- Improving peer-review process
- Ensuring integrity
- Becoming role models in open science

PRACTITIONERS AGENCIES

- Facilitating industry applications
- Encouraging collaborations
- Leading change to best practice
- Supporting R&D information transfer
- Repurposing code

GENERAL PUBLIC

- Building trust in science
- Encouraging public access
- Delivering robust, science-based solutions
- Promoting public computational literacy
- Reusing code in citizen-science

© by Society for Open Reliable and Transparent Ecology and Evolutionary Biology sorteef.org 2021 CC-BY-SA

OPEN DATA IN RESEARCH



WHAT?

OPEN DATA is data that may be used, re-used, and redistributed freely by anyone



FAIR

- Findable
- Accessible
- Interoperable
- Reusable

WHY?

Increase your:

- Citations by 25%¹
- Public trust
- Exposure
- Transparency
- Reproducibility



HOW?

- 1 **Prepare your data**
Tidy data + meta-data
Editable and open fileformats
- 2 **Pick platform**
OSF, GitHub, Zenodo, Dryad, Figshare, PANGAEA and more!
- 3 **Make it public and citeable!**


1. Colavizza, G. et al. (2020). The citation advantage of linking publications to research data. *PLOS ONE*, 15(4), e0230416.

© by Society for Open Reliable and Transparent Ecology and Evolutionary Biology sorteef.org 2021 CC-BY-SA

What will you learn in this course?

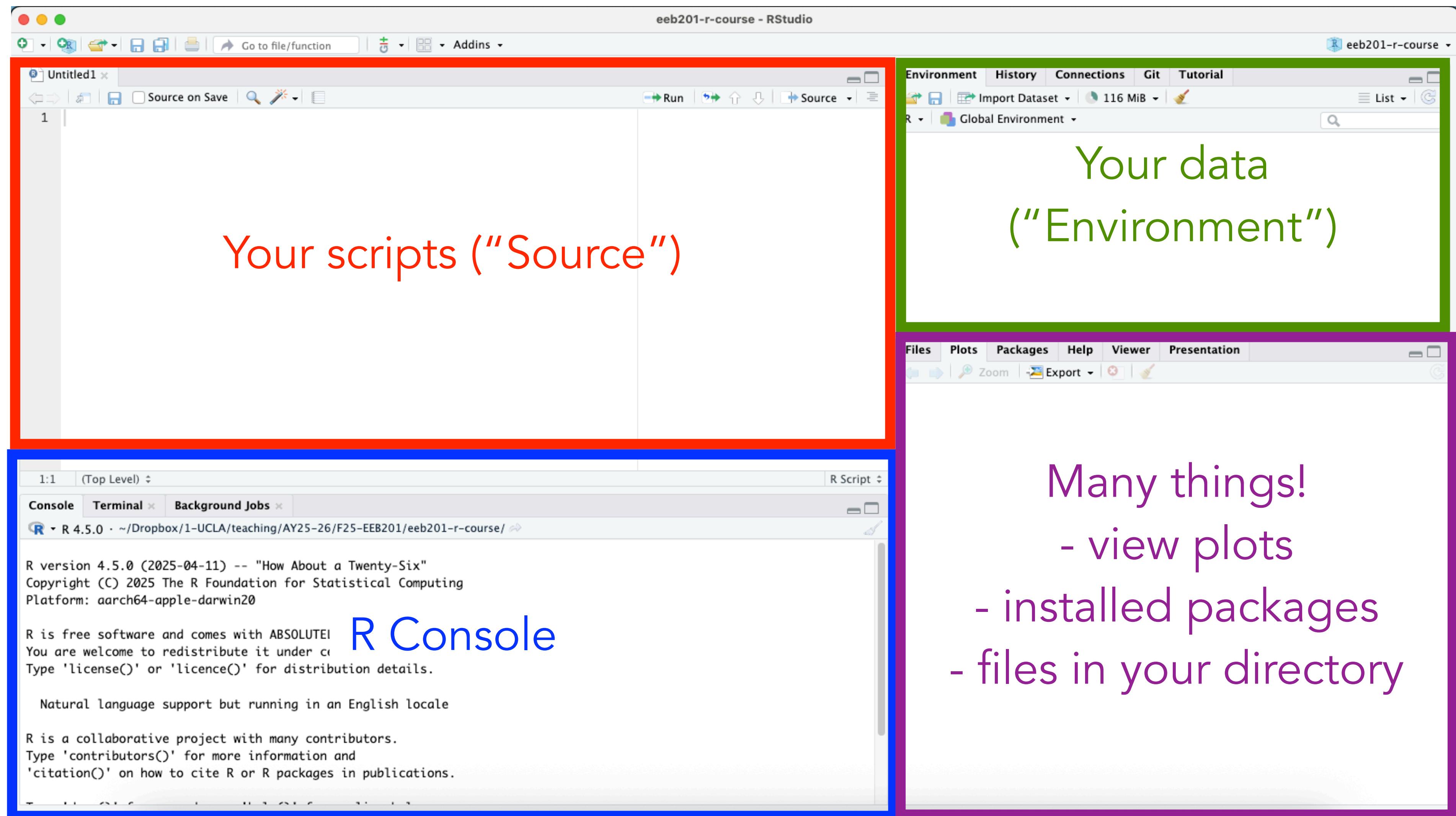
“Reproducible workflows consist of three components: a **fully scriptable statistical programming environment** (such as R or Python), **reproducible analysis** (first described as literate programming), and **version control** (commonly implemented using GitHub).”





Practice, practice, practice...

Getting started in RStudio



Getting started in RStudio

```
## SM Aguillon
## 10/1/25
## EEB 201, Week 1 test script

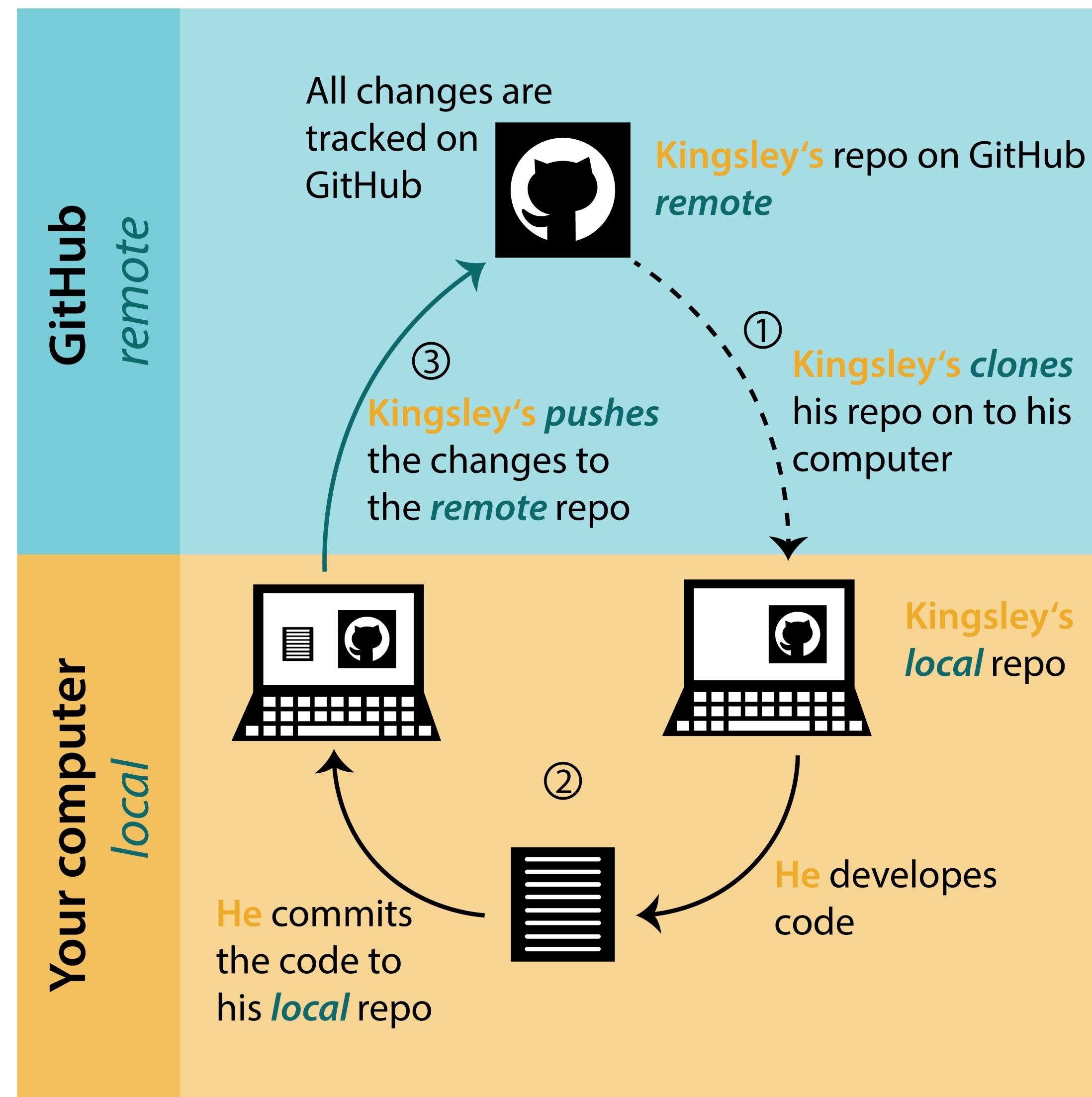
# in R, you can assign values to variables using <-
x <- 3
y <- 4

# NOTE: you can do this in a script or directly in the console
# if working in a script, you can run individual lines of code by...
# highlighting the line and using command+enter (Mac) or control+enter (PC)

# you can also do simple math using operators like: + - * / ^
x + y

# there are also many basic functions for math
sum(x,y)
log(x)
```

Week 1 Problem Set



Week 1 Problem Set

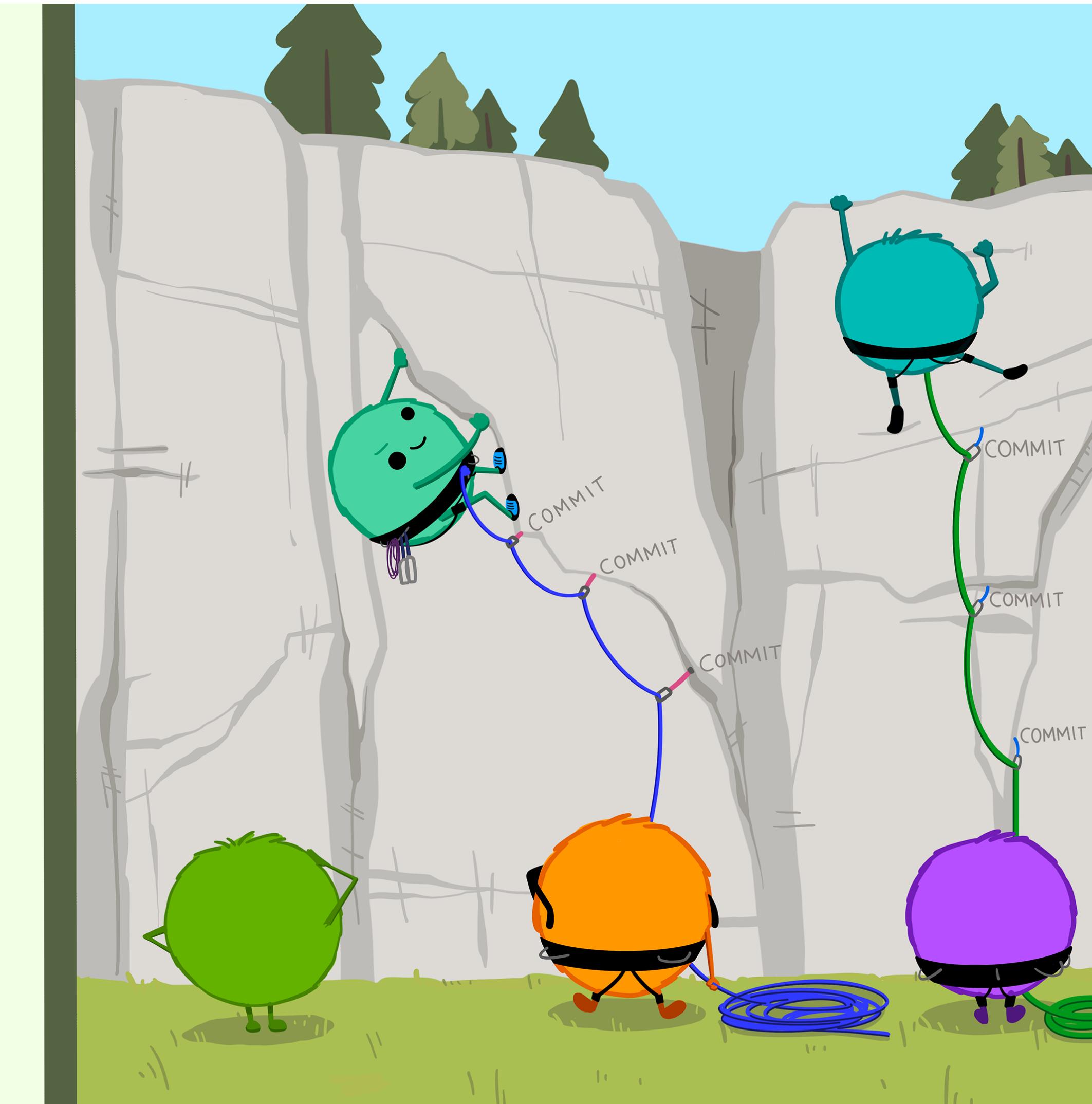
“

Using a Git commit is like using anchors and other protection when climbing...**if you make a mistake, you can't fall past the previous commit.**

Commits are also helpful to others, because **they show your journey, not just the destination.**

- HADLEY WICKHAM & JENNY BRYAN

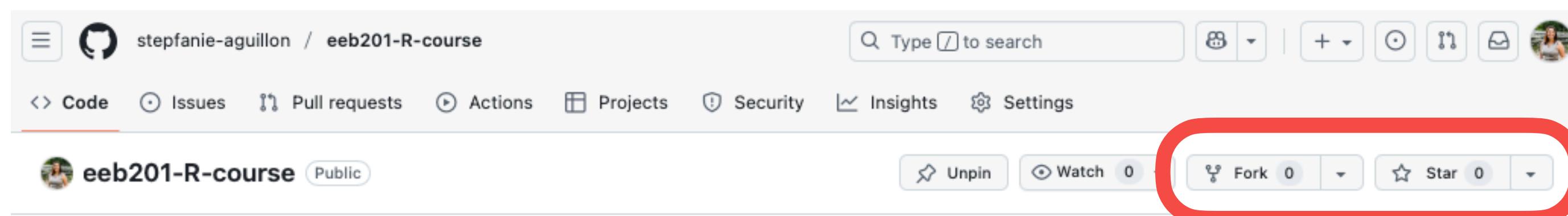
Wickham & Bryan, R Packages (<https://r-packages.org/preface.html>)



Artwork by Allison Horst

Week 1 Problem Set

<https://github.com/stepfanie-aguillon/eeb201-R-course>



The screenshot shows the GitHub repository page for 'eeb201-R-course'. At the top, there's a navigation bar with links for Code, Issues, Pull requests, Actions, Projects, Security, Insights, and Settings. Below the navigation bar, the repository name 'eeb201-R-course' is displayed, along with its status as 'Public'. To the right of the repository name are buttons for Unpin, Watch (0), Fork (0), and Star (0). A red box highlights the 'Fork (0)' button. The main content area shows a list of recent commits made by 'stepfanie-aguillon'. On the right side of the page, there's an 'About' section with details about the repository, such as 'Repository for the UCLA course EEB 201', and sections for Releases, Packages, and Activity.

stepfanie-aguillon / eeb201-R-course

Type to search

Code Issues Pull requests Actions Projects Security Insights Settings

eeb201-R-course Public

Unpin Watch 0 Fork 0 Star 0

main 1 Branch 0 Tags

Go to file

About

Repository for the UCLA course EEB 201

Readme

Activity

0 stars

0 watching

0 forks

Commits

stepfanie-aguillon update README

8220927 · 4 minutes ago 16 Commits

images add images to repo last month

problem-sets adding Week 1 PS 7 minutes ago

README.md update README 4 minutes ago

eeb201-syllabus-F25.pdf add syllabus to course repo last month

install-guide.md fixing typo in install page 3 hours ago

README

EEB 201, Fall 2025

Introduction to R for Ecology and Evolutionary Biology

Course Description

From data analysis to visualization to simulations, coding has become essential to communicate scientific research. This course introduces students to reproducible and collaborative coding practices in the R programming language. It is specifically designed with EEB graduate students in mind, with the goal of building good coding practices for your independent data analysis needs. So, no prior coding experience is required!

Week 1 Problem Set

<https://github.com/stepfanie-aguillon/eeb201-R-course>

Tentative Schedule

Before Class - Install Software

Pre-course assignment: [install necessary programs!](#)

Week 1 - Introduction to RStudio and GitHub

Pre-class reading: [Why learn R?](#); [Why Git? Why GitHub? 1.1-1.5](#); if you're totally new to R, practice with [exercise 2.1](#)

Slides:

Problem set: [Week 1 PS](#)

Week 2 - Making figures with ggplot2

Pre-class reading:

Slides:

Problem set:

Further reading

- *The American Naturalists' Guidelines for Archiving Code with Data*
- Ivimey-Cook et al. 2023. Implementing code review in the scientific workflow: insights from ecology and evolutionary biology. *Journal of Evolutionary Biology*.
- Pereira Braga et al. 2023. Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolutionary biology. *Methods in Ecology and Evolution*.
- Powers & Hampton 2019. Open science, reproducibility, and transparency in ecology. *Ecological Applications*.