Welcome to the Treehouse Reproducible Research Practices Workshop!

April 8 & 10, 2025
Childhood Cancer Data Lab
https://alexslemonade.github.io/2025-treehouse-training/

Alex's Lemonade Stand Foundation for Childhood Cancer™

Childhood Cancer Data Lab
ALSF

Tell us about you!

- What’s your name?
- What do you work on?
- What piece of media have you recently been enjoying?

Meet your instructors

JOSH

Joshua Shapiro

he/him/his

Data Scientist @ the Data Lab

PhD Ecology & Evolution, *UChicago*
Postdoc Integrative Genomics, *Princeton*

Research interests:

- **Single cell workflows**
- **Statistical methods**

jashapiro

Meet your instructors

JACLYN

Jaclyn Taroni

she/her/hers

Director @ the Data Lab

PhD Genetics *Dartmouth*
Postdoc Computational Biology *UPenn*

Research interests:

- **Transcriptomics in rare, complex diseases**
- **Unsupervised pattern extraction**

jaclyn-taroni

Tell us about you!

- What’s your name?
- What do you work on?
- What piece of media have you recently been enjoying?

Code of Conduct

Be kind, have fun

We value the involvement of everyone in the community. We are committed to creating a friendly and respectful place for learning, teaching, and contributing.

- Use welcoming and inclusive language.
- Be respectful of different viewpoints and experiences.
- Gracefully accept constructive criticism.
- Focus on what is best for the community.
- Show courtesy and respect towards other community members.

Read the full Code of Conduct:
https://alexslemonade.github.io/2025-treehouse-training/code-of-conduct.html

If you at any time feel harassed or treated inappropriately, please contact jen.omalley@ccdatalab.org.

What you will learn (and what you won’t)

What you will learn

We will cover some common practices for reproducible computational research, including:

- Organizing your projects, including data, code, and documentation
- Tracking and automating your work with scripts
- Making your code more readable, robust, and reusable
- Maintaining and tracking changes in your projects and code over time with Git and GitHub
- Managing and tracking software and package versions for improved reproducibility
- Working with branches in Git

We'll also tell you our ***opinions*** about analytical code review and some tips for how to do it!

What you will learn (continued)

Our overarching goals:

To introduce principles and techniques to achieve reproducible results in computational cancer research and to show you commonly-used approaches that you can apply to increase the impact of your research!

To introduce the practice of analytical code review and other practices that support it so that you can think about whether it's right for Treehouse!

What you won’t learn

We won’t have time to cover:

- How to program in specific languages such as R or Python
- All the features and foibles of Git and GitHub, including all the security considerations thereof
- Workflow management systems such as CWL, Snakemake, or Nextflow

How do we pick what we teach?

We've opted to teach using `git` with GitKraken.

However, we include CLI commands in the slide for your reference.

We chose this route because:

- When introducing concepts, GitKraken visuals can be very useful.
- CLI offers the most extensibility, for example you can't use a GUI on your external server.

Schedule

Day 1 Morning

9:00 am	Welcome Why does reproducibility matter?
9:45 am	Project organization
10:30 am	Break
10:45 am	Shell scripting
11:15 am	Git, Part 1
12:15 pm	Lunch

Day 1 Afternoon

1:00 pm	Organizing code in scripts and notebooks
1:45 pm	Managing packages and environments
2:30 pm	Break
2:45 pm	Git, Part 2
3:30 pm	Working with multiple branches
5:00 pm	Adjourn

Full schedule: https://alexslemonade.github.io/2025-treehouse-training/workshop-schedule.html

Day 2 Morning

9:00 am	Pull request basics
9:30 am	Resolving merge conflicts
10:00 am	Git workflows
10:30 am	Break
10:45 am	Planning and tracking work with Github
11:30 am	Analytical code review
12:15 pm	Lunch

Day 2 Afternoon

1:00 pm	Leaving and responding to review
1:45 pm	**Optional:** Pull request responsibilities
2:30 pm	Break
2:45 pm	**Optional:** Pull request scope
3:30 pm	**Optional:** Open office hours
5:00 pm	Adjourn

Full schedule: https://alexslemonade.github.io/2025-treehouse-training/workshop-schedule.html