Welcome to the Data Lab RNA-Seq Training Workshop!

August 19 – 22, 2024
University of Minnesota
https://alexslemonade.github.io/2024-august-training/

Childhood Cancer Data Lab x Alex's Lemonade Stand

Tell us about you!

- What’s your name?
- Where are you from?
- What are you studying?
- What was your favorite moment from the Olympics?

Meet your instructors

JOSH

Joshua Shapiro

Senior Data Scientist @ the Data Lab

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

Research interests:

- **Evolutionary genomics**
- **Single cell workflows**

jashapiro

Meet your instructors

STEPHANIE
Stephanie Spielman

Data Scientist @ the Data Lab

PhD Integrative Biology *UT Austin*
Postdoc Computational Molecular Evolution *Temple*

Research interests:

- **Protein & virus evolution**
- **Reproducible genomics analysis**
- **Data science and bioinformatics education**

sjspielman

Meet your instructors

JACLYN
Jaclyn Taroni

Director @ the Data Lab

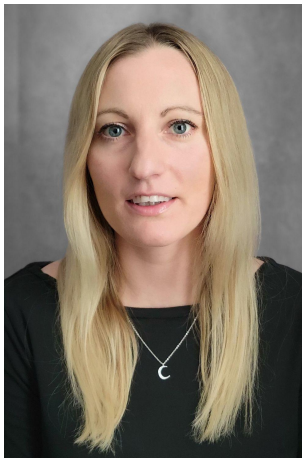
PhD Genetics *Dartmouth*
Postdoc Computational Biology *UPenn*

Research interests:

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

jaclyn-taroni

Last but certainly not least!



JEN

Jen O’Malley

Scientific Community Manager

- **Helps administer Data Lab offerings such as workshops**
- **Manages communications**
- **Saves the rest of us from ourselves**

Tell us about you!

- What’s your name?
- Where are you from?
- What are you studying?
- What was your favorite moment from the Olympics?

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 here:
https://alexslemonade.github.io/2024-august-training/code-of-conduct.html

If you at any time feel harassed or treated inappropriately, please contact ccdl@alexslemonade.org.

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

Our overarching goals

Prepare you to perform “frontline” analyses of your own data

Help to get you more comfortable reading documentation/learning new methods on your own

Give you the language and tools to collaborate more effectively with analysts when needed

What you will learn

We will introduce you to the R programming language, R Notebooks, and reproducible research practices such as:

- organizing your projects, code and documentation
- version control with Git
- managing packages and environments

We we will cover pipelines for the quality control, processing, and downstream analysis of bulk RNA-seq data

We will introduce common approaches for pathway analysis

We generally elect to go *broad* and not *deep*.

What you won’t learn

We don’t address experimental design (e.g., how many replicates you need).

We won’t compare tools (e.g., edgeR vs. DESeq2 for differential gene expression).

We won’t cover every feature (or assumption) of the tools we do present.

You may not be able to perform every analysis you need for your own work, particularly for complex experimental designs.

We present analysis as a series of *linear steps*. In practice, it’s ***not***. It’s important to consult analysis experts when you need to and to keep track of and report what you’ve done.

We won't cover all the features and foibles of Git and GitHub

How do we pick what we teach?

We want methods to be or to have:

- Useful for a wide range of experimental designs, sample sizes
- Easy to use, well-documented, and consistently updated
- Solid tutorials, a sizeable user base, and responsive authors/maintainers

We have a preference for methods that integrate easily into a single workflow that can be run on a laptop (and our own personal biases as scientists).

Schedule

Monday

Workshop Intro

Intro to R
RStudio Server
Base R
ggplot2 & tidyverse

Intro to Reproducible Research Practices
Project organization

Tuesday

Reproducible Research Practices
Unix & the command line
Git
Organizing code
Managing environments

Intro to Bulk RNA-seq

Workshop dinner
Surly Brewing Co.

Wednesday

Bulk RNA-seq
Quantification & QC
Exploratory analysis

Reproducible Research Redux
Git on the server

Bulk RNA-seq
Differential Expression
Heatmaps & clustering

Thursday

Pathway analysis
ORA
GSEA
GSVA

Participant Presentations

Full schedule: https://alexslemonade.github.io/2024-august-training/workshop/SCHEDULE.html

Daily Schedule Components

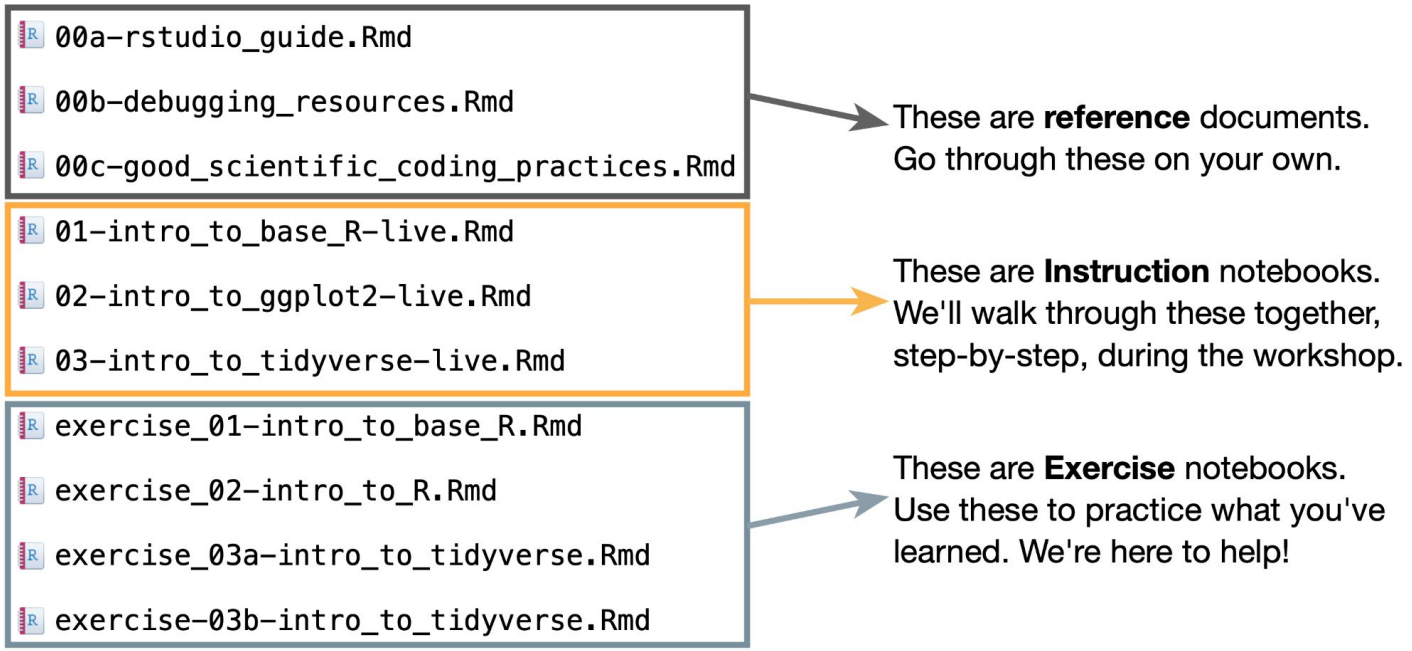
Instruction
Full group
Lectures

- Introduce concepts and background
- Demonstrate usage
- Answer general questions

Consultation Periods
Exercise notebooks
Your own data

- Ask questions of instructors and other participants
- Practice what you have learned
- Work on exercises individually or in groups
- Work with your own data

Module Layout



Module cheatsheets cover key functions

https://github.com/AlexsLemonade/training-modules/tree/2024-august/module-cheatsheets

`dplyr`

Read the `dplyr` package documentation here.
A vignette on the usage of the `dplyr` package can be found here.

Library/Package	Piece of code	What it's called	What it does
`dplyr`	`%>%`	Pipe operator	Funnels a data.frame through tidyverse operations
`dplyr`	`filter()`	Filter	Returns a subset of rows matching the conditions of the specified logical argument
`dplyr`	`arrange()`	Arrange	Reorders rows in ascending order. arrange(desc()) would reorder rows in descending order.
`dplyr`	`select()`	Select	Selects columns that match the specified argument
`dplyr`	`mutate()`	Mutate	Adds a new column that is a function of existing columns
`dplyr`	`summarise()`	Summarise	Summarises multiple values in an object into a single value. This function can be used with other functions to retrieve a single output value for the grouped values. `summarize` and `summarise` are synonyms in this package.
`dplyr`	`rename()`	Rename	Renames designated columns while keeping all variables of the data.frame
`dplyr`	`group_by()`	Group By	Groups data into rows that contain the same specified value(s)
`dplyr`	`inner_join()`	Inner Join	Joins data from two data frames, retaining only the rows that are in both datasets.

Thursday

Your own projects
Exercise notebooks

Spend time on Thursday (and earlier!) working with your own data, getting assistance as needed from Data Lab staff and each other.

Participant Presentations

Present what you worked on during the consultation times to the group!

Training Procedures

We’re going to use sticky notes and note cards…

- As an alternative to raising your hands for help
- To give feedback about the session

During the session

Use the yellow sticky note to let us know that you need help or having some trouble following along.

One of us will come over and help you.

Use the blue sticky note to let us know when a long-running step has completed.

At the end of each session

At the end of each module, write down your muddiest point on a white note card:

I do not understand _____.

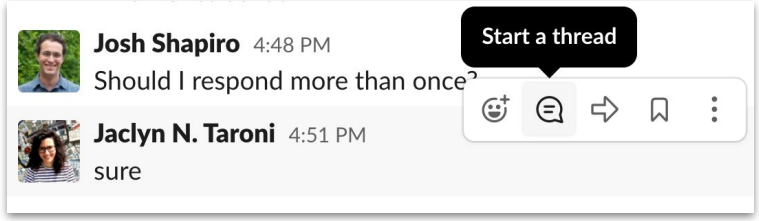
At the end of the session, use one blue note card and one white note card to write down:

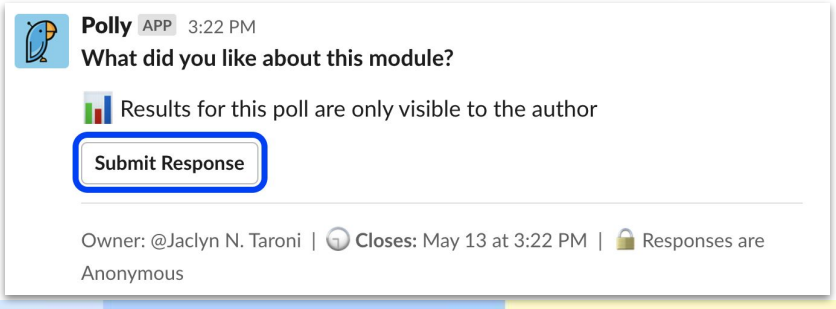
What did you like about this session?

How might we improve the session?

We encourage you to use Slack

- You have been added to the **#2024-august-training** Slack channel
- Post public questions, get help with errors and debugging, make comments, and help others!
    - Use threads to keep related content together
- Stay in touch after the workshop!





Housekeeping Notes

- Waivers - If you have not yet done so, we have printed copies you can sign!
- Where are the restrooms?
- Where is water/coffee available?
- Snacks!
- Dinner tomorrow (Tuesday) at Surly Brewing Co, 520 Malcolm Ave SE