Welcome to the June 2021 Virtual CCDL RNA-Seq Training Workshop!

June 28 – July 2, 2021
Childhood Cancer Data Lab
https://alexslemonade.github.io/2021-june-training/

Alex's Lemonade Stand
Foundation for Childhood Cancer™

CHILDHOOD
CANCER
DATA LAB
Powered by: Alex's Lemonade Stand Foundation

Meet your instructors

JOSH

Joshua Shapiro

Data Scientist @ the CCDL

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

Research interests:

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

jashapiro

Meet your instructors

CHANTE

Chante Bethell

Biological Data Analyst @ the CCDL

Bachelor’s in Bioinformatics from *Rowan University*

Research interests:

- **Functional motifs in the proteome**

cbethell

Meet your instructors

Ally

Ally Hawkins

Data Scientist @ the CCDL

PhD Cancer Biology *University of Michigan*
Postdoc Computational Biology *Cornell*

Research interests:

- **Single cell data analysis**
- **Origins of pediatric solid tumors**

allyhawkins

Meet your instructors

JACLYN

Jaclyn Taroni

Principal Data Scientist @ the CCDL

PhD Genetics *Dartmouth*
Postdoc Computational Biology *UPenn*

Research interests:

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

jaclyn-taroni

Meet your instructors

STEPHANIE

Stephanie Spielman

Assistant Professor @ Rowan University

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

Research interests:

- **Protein evolution and phylogenetic modeling**
- **Data science and bioinformatics education**

sjspielman

Other staff you may see

DEEPA

Deepa Prasad

User Experience Designer
@ CCDL

- Talks to researchers about their needs and frustrations
- Designs usable software

KURT

Kurt Wheeler

Data Engineer
@ CCDL

- Builds scalable systems
- Manages servers

Tell us about you!

- What’s your name?
- What are you studying?
- What’s a recent piece of media (book, movie, TV show, etc.) you enjoyed?

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/2021-june-training/code-of-conduct.html

If you at any time feel harassed or treated inappropriately, please contact [ccdl@alexslemonade.org.](mailto:ccdl@alexslemonade.org)

Monday

Workshop Intro

Intro to R
RStudio Server
Base R
ggplot2 & tidyverse

Consultations
Exercise notebooks

Tuesday

Single-cell RNA-seq
Quantification & QC
Filtering & Normalization

Consultations
Exercise notebooks
Your own data

Wednesday

Single-cell RNA-seq
Dimensionality reduction
Clustering
Marker identification

Consultations
Exercise notebooks
Your own data

Thursday

Pathway Analysis
Overrepresentation
GSEA

Consultations
Exercise notebooks
Your own data

Friday

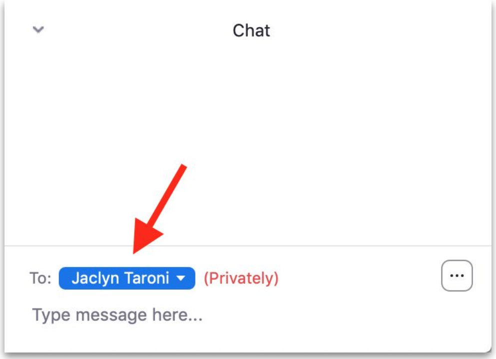
Consultations
Your own data
Exercise notebooks

Presentations

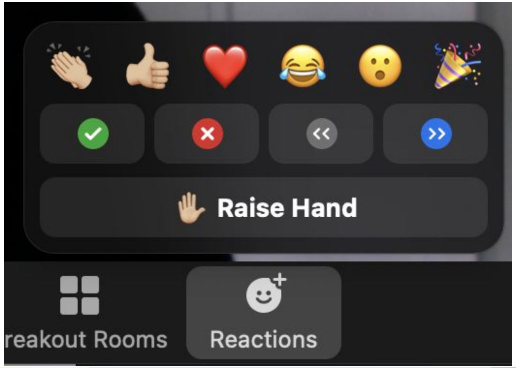
Full schedule: https://alexslemonade.github.io/2021-june-training/workshop/SCHEDULE.html

Virtual Training Procedures

General Zoom Etiquette

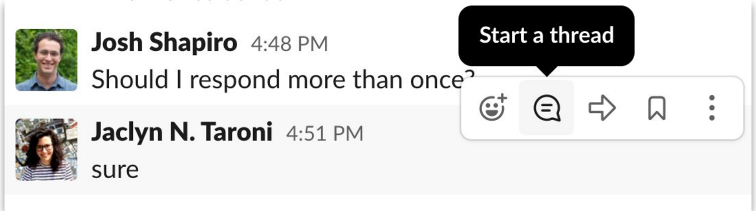


- Keep your microphone muted
- Type questions in the Chat window (directed to the host)
  - Click on the “Chat” button at the bottom of your window to open the chat.
- Use the Zoom reactions to tell us how you are doing!



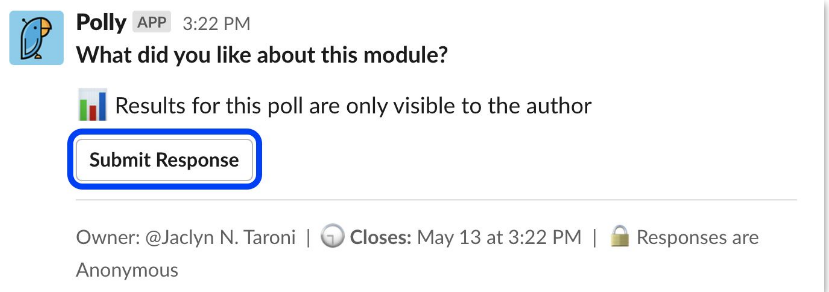
Using Slack

- Use the **#2021-june-training** Slack channel
- Post public questions, get help with errors and debugging, make comments, and help your fellow participants!
  - Use threads to keep related content together
- We will post questions at the end of each session with Polly



Josh Shapiro 4:48 PM
Should I respond more than once?

Jaclyn N. Taroni 4:51 PM
sure



Polly APP 3:22 PM
What did you like about this module?

Results for this poll are only visible to the author

Owner: @Jaclyn N. Taroni | **Closes:** May 13 at 3:22 PM | Responses are Anonymous

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

What you will learn

We will introduce you to the R programming language, R Notebooks, and some reproducible research practices.

We cover pipelines for the quality control, pre-processing, and initial analysis of single-cell RNA-seq data almost entirely through hands-on exercises.

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

Our overarching goals: To prepare you to perform “frontline” analyses of your own data, to get you more comfortable reading documentation/learning new methods on your own, and to give you tools to collaborate more effectively with analysts when needed

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., Bioconductor vs. Seurat for single cell analysis).

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. For example, in the context of single-cell analysis, we do not cover integrating data from multiple samples.

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.

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

Daily Schedule Outline

Instruction
Full group
Lectures

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

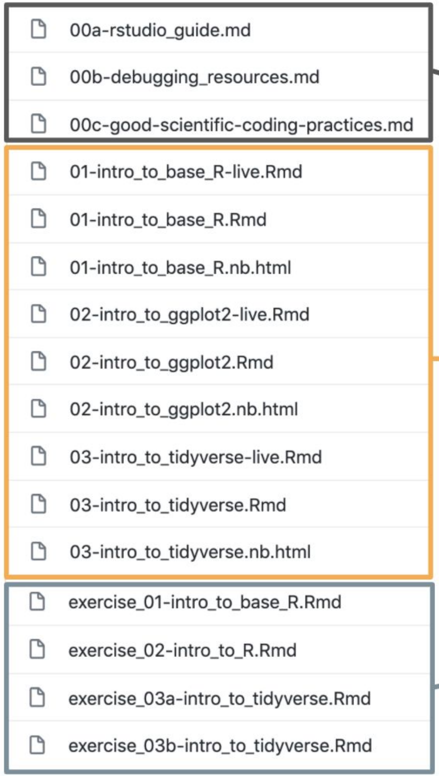
Breakout
Small groups
Start exercise notebooks

- Split up into Zoom breakout rooms
- Ask questions of instructors and other participants

Consultation Period
Exercise notebooks
Your own data

- Practice what you have learned
- Work on exercises individually or in groups
- Work with your own data

Module Layout



00a-rstudio_guide.md
00b-debugging_resources.md
00c-good-scientific-coding-practices.md

These are reference documents.
Go through these on your own.

01-intro_to_base_R-live.Rmd
01-intro_to_base_R.Rmd
01-intro_to_base_R.nb.html
02-intro_to_ggplot2-live.Rmd
02-intro_to_ggplot2.Rmd
02-intro_to_ggplot2.nb.html
03-intro_to_tidyverse-live.Rmd
03-intro_to_tidyverse.Rmd
03-intro_to_tidyverse.nb.html

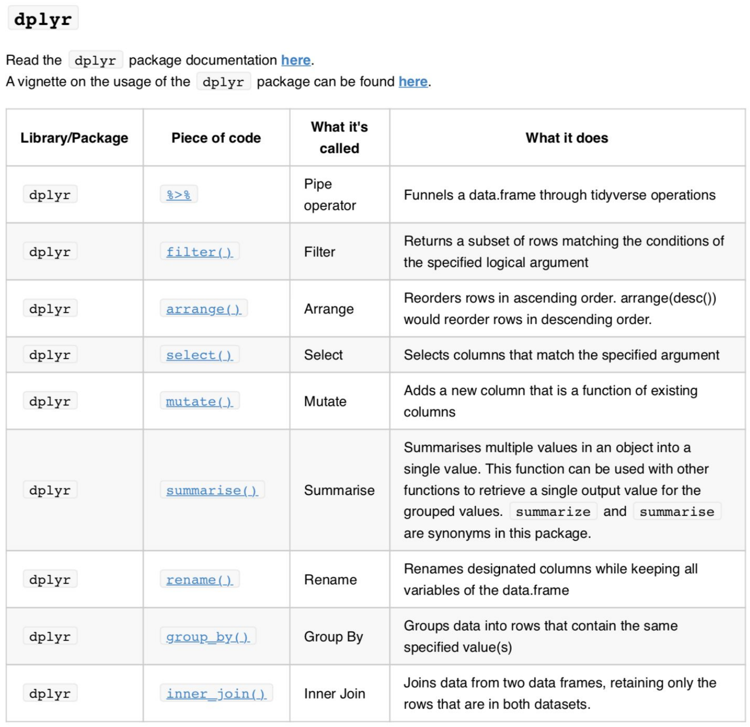
These are **Instruction** notebooks.
We'll walk through these together,
step-by-step, during the workshop.

exercise_01-intro_to_base_R.Rmd
exercise_02-intro_to_R.Rmd
exercise_03a-intro_to_tidyverse.Rmd
exercise_03b-intro_to_tidyverse.Rmd

These are **Exercise** notebooks.
Use these to practice what you
have learned. We're here to help!

Module cheatsheets cover key functions

https://github.com/AlexsLemonade/training-modules/tree/2021-march/module-cheatsheets



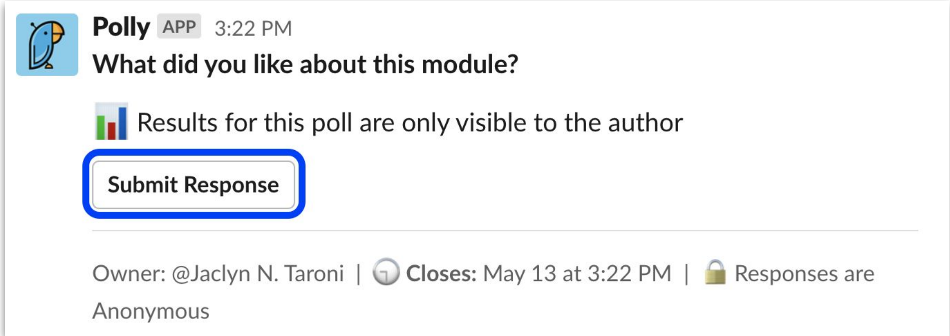
`dplyr`

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.

We want your feedback!

At the end of each module, we will post a few questions in the Slack channel.



Polly APP 3:22 PM

What did you like about this module?

📊 Results for this poll are only visible to the author

Owner: @Jaclyn N. Taroni | **Closes:** May 13 at 3:22 PM | 🔒 Responses are Anonymous

- The most difficult or confusing point of the module ("muddiest point")
  We will post additional material answering your questions the next day
  Responses to this question will appear in the channel anonymously

- What did you like about the module?
- How we can improve the module?
  These responses will be collected anonymously (and not posted)

Friday

Your own projects
Exercise notebooks

Spend Friday working with your own data, getting assistance as needed from CCDL staff and each other.

Presentations

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

Communication during instruction

zoom

- **I have an urgent question** that needs an answer before moving on:
  - *Raise Hand* or *Chat* with the room host

- **I'm stuck with an error** and can't proceed with the hands-on exercise
  - *Chat* with meeting host: Request 1:1 and you will be placed in a breakout room with a CCDL staff member

slack

- **I have an general question** that does not need an answer right away.
  - *Post* in *#2021-march-training*

- **I'm having trouble logging in** to RStudio Server
  - *Direct Message* a CCDL staff member (not the current host or instructor)

Trouble logging into Zoom and Slack? ***Email*** training@ccdatalab.org

Communication at other times (consultation periods)

slack

- I have questions about **previous instruction** or **exercise notebooks**
  - ***Post*** in *#2021-march-training*
  - If you need to share your screen, we will set up a 1:1 or group Zoom call
- **I would like to be paired up with other participants**
  - ***Post*** in *#2021-march-training*; we can set you up in a Zoom breakout room
- I have a question that is **highly specific to my data**
  - ***Direct Message*** a CCDL staff member
- **I'm having trouble logging in** to RStudio Server
  - ***Direct Message*** a CCDL staff member

Trouble logging into Zoom and Slack? ***Email*** training@ccdatalab.org