



Welcome to the August 2020 Virtual CCDL RNA-Seq Training Workshop!

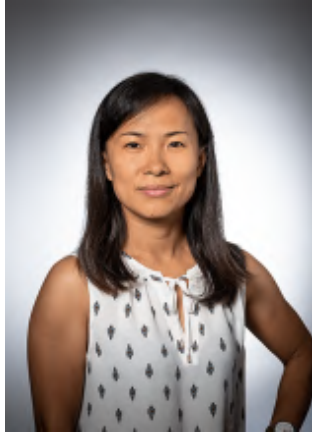
August 17-21, 2020

Carnegie Mellon University Libraries
with Childhood Cancer Data Lab

https://cmu-lib.github.io/CCDL_pgh_202008/



Meet your instructors



HUAJIN

Huajin Wang

Librarian; Program Director for Open Science & Data Collaborations @CMU Libraries

PhD: Cell Biology, *Univ. Alberta*

Postdoc: Cell Biology of Lipids, *Yale /Harvard*

Research interests:

- Lipid and membrane biology, cell imaging
- Open science, data reuse and reproducibility



huajinw



HuajinBioLib

Meet your instructors



CHRISTINA

Christina Akirtava

PhD Candidate @ Carnegie Mellon University

Bachelor's Bioinformatics & Molecular Biology
Rensselaer Polytechnic Institute

Research interests:

- **Understanding translation regulation**



bio2data



bio2data

Meet your helpers from the CCDL



JACLYN

Jaclyn Taroni, PhD
Principal Data Scientist
@the CCDL



JOSH

Joshua Shapiro, PhD
Data Scientist
@the CCDL



CANDACE

Candace Savonen
Biological Data Analyst
@the CCDL



CHANTE

Chante Bethell
Biological Data Analyst
@the CCDL

Meet workshop moderators /hosts



HANNAH

Hannah Gundermann, PhD
Research Data Management
Consultant @CMU Libraries



SARAH

Sarah Young
Social sciences librarian
@CMU Libraries



EMMA

Emma Slayton, PhD
Data Curation, Visualization, and GIS
Specialist @CMU Libraries

Why CMU Libraries?



A University Libraries program supporting collaborative, transparent, openly accessible, and reproducible research across all disciplines at Carnegie Mellon University. We recognize that having well documented and automated research workflows, code, and datasets is essential to making research more interdisciplinary, efficient, and reusable as well as allowing researchers to leverage data science techniques. This program provides services and infrastructure for open research at CMU through digital tools, training opportunities for research tools and practices, collaboration opportunities on data science projects, special events and advocacy, and a team of experts available as research consultants and collaborators.



Tools



Trainings



Events



Experts

- Tools that facilitate research data sharing and collaboration
- Training on basic data science and data management
- Events that engage with the research community to make data more reproducible and reusable
- [Data Collaborations Lab](https://www.library.cmu.edu/datapub/open-science) to bridge collaborations on data science projects



Casey Greene
Associate Professor @ UPenn
Director @ CCDL

Invited Speakers



Sean Davis

Senior Associate Scientist
National Cancer Institute, NIH



Matt Fredrikson

Assistant Professor
School of Computer Science
Carnegie Mellon University



Casey Greene

Assistant Professor of Systems
Pharmacology and
Translational Therapeutics
Perelman School of Medicine
University of Pennsylvania



Irene Kaplow

Postdoctoral Fellow
School of Computer Science
Carnegie Mellon University



<https://events.library.cmu.edu/aidr2019/>

Tell us about you!

- What's your name?
- What are you studying?
- What's one thing you are proud of?

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://github.com/cmu-lib/CCDL_pgh_202008/blob/master/code-of-conduct.md

If you at any time feel harassed or treated inappropriately, please contact
dataCoLAB@andrew.cmu.edu

Monday

Workshop Intro

Intro to R

RStudio Server
Base R

Consultations

Exercise notebooks

Wednesday

Bulk RNA-seq

Pre-processing
Exploratory data analysis

Consultations

Exercise notebooks

Friday

Consultations

Exercise notebooks
(Your own data**)

Presentations

Tuesday

Intro to R

ggplot2
Tidyverse

Consultations

Exercise notebooks

Thursday

Bulk RNA-seq

Differential expression

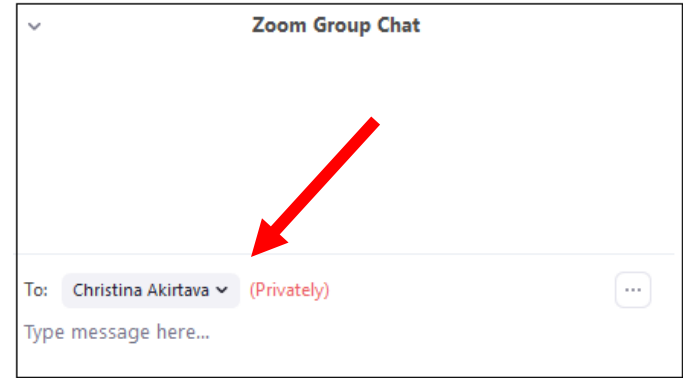
Consultations

Exercise notebooks

Virtual Training Procedures

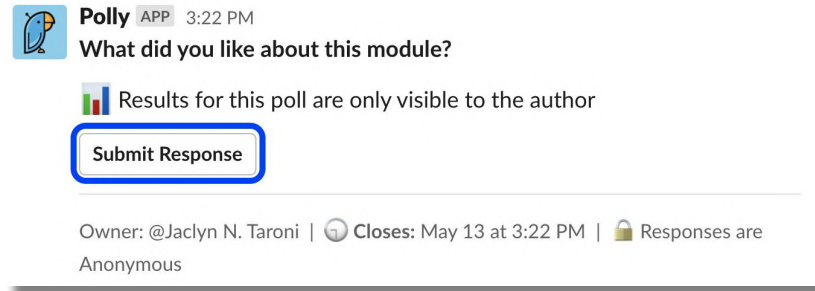
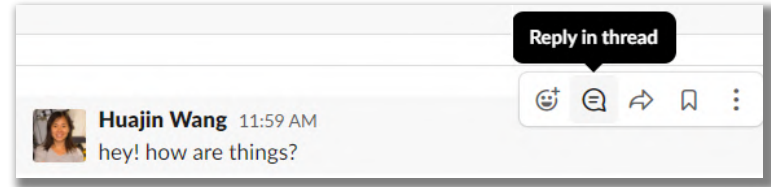
General Zoom Etiquette

- 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 status buttons to tell us how you are doing!
 - Click on the “Participants” button at the bottom of your zoom window to see these buttons



Using Slack

- Use the **#2020-pgh-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



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, processing, and downstream analysis of bulk and 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., 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 to perform 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.

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

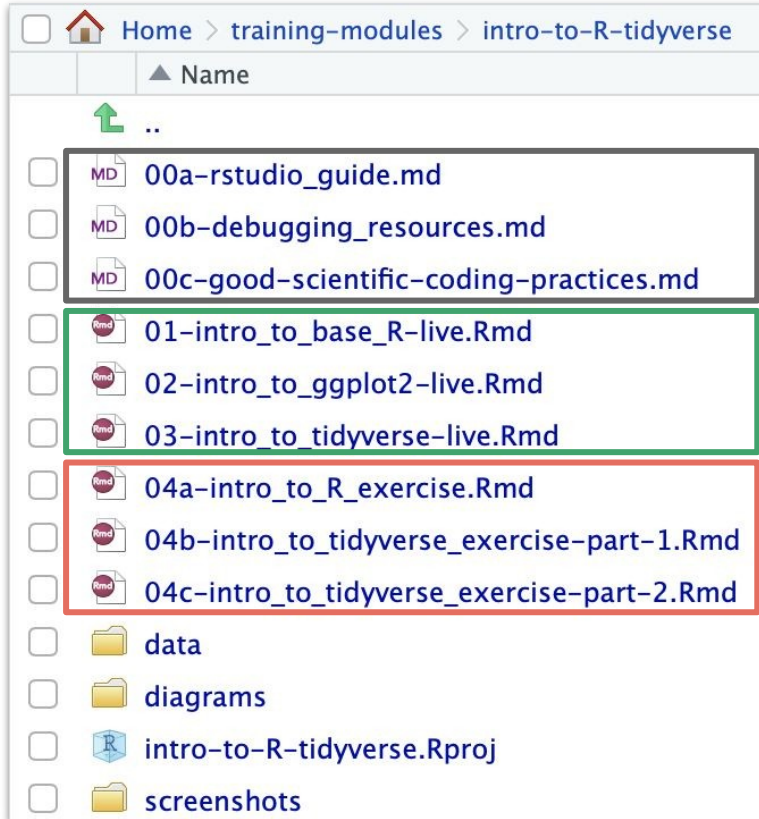
- Split up into Zoom breakout rooms
- Ask questions of instructors and other participants

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



This is a reference document.
We will not go through this.

We'll walk through these notebooks
together, step-by-step

You will practice what you have
learned. We're here to help!

Module cheatsheets cover key functions

<https://github.com/AlexsLemonade/training-modules/tree/2020-june/module-cheatsheets>

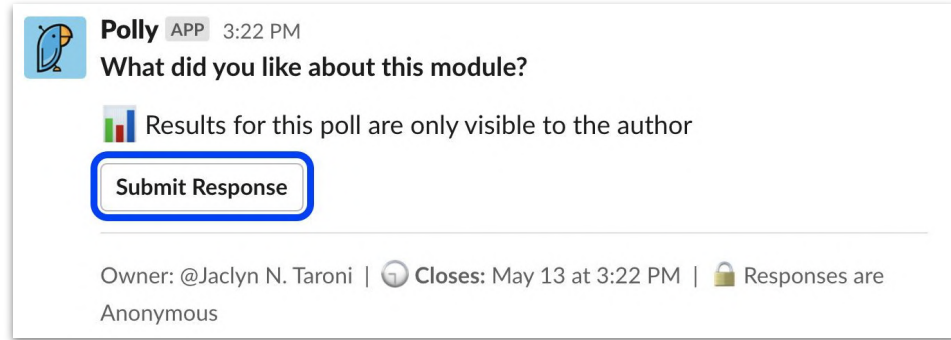
Base R

Read the Base R documentation [here](#).

Library/Package	Piece of code	What it's called	What it does
Base R	<code>library()</code>	Library	Loads and attaches additional packages to the R environment.
Base R	<code><-</code>	Assignment operator	Assigns a name to something in the R environment.
Base R	<code>c()</code>	Combine	Combines values into a vector or list.
Base R	<code>%in%</code>	"in" logical operator	Checks if the given value(s) on the left side of the operator are in the vector or other R object defined on the right side of the operator. It returns a logical TRUE or FALSE statement.
Base R	<code>rm(x)</code>	Remove	Removes object(s) <code>x</code> from your environment.
Base R	<code>==, <=, >=, !=</code>	Relational Operators	These are binary operators which allow for the comparison of values in an object.
Base R	<code>str(x)</code>	Object Structure	Gets a summary of the object <code>x</code> structure.
Base R	<code>class(x)</code>	Object Class	Returns the type of the values in object <code>x</code> .
Base R	<code>nrow(x); ncol(x)</code>	Number of Rows; Number of Columns	Get the number of rows and the number of columns in an object <code>x</code> , respectively.
Base R	<code>length(x)</code>	Number of Rows	Returns how long the object <code>x</code> is.
Base R	<code>min(x)</code>	Minimum	Returns the minimum value of all values in an object <code>x</code> .

We want your feedback!

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



The screenshot shows a Slack message from the Polly app. The message content is: "What did you like about this module?". Below the question, there is a small bar chart icon and the text "Results for this poll are only visible to the author". A blue button labeled "Submit Response" is highlighted with a blue border. At the bottom of the message, it says "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 workshop or a specific module?
- How we can improve the workshop or a specific module?
These responses will be collected anonymously (and not posted)

Friday

Exercise notebooks Your own projects

Spend Friday working on exercise notebooks and getting assistance from the staff, forming a research question and outline of data analysis plan from your own data, or begin initial analysis of your own data.

Presentations

Present what you learned, worked on, or future research plan to the group.

Communication during instruction



raise hand

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



- **I have an general question** that does not need an answer right away.
 - *Post* in #2020-pgh-training
- **I'm having trouble logging in** to RStudio Server
 - *Direct Message* a staff member (not the current host or instructor)

Trouble logging into Zoom and Slack? *Email* dataCoLAB@andrew.cmu.edu

Communication at other times (consultation periods)



- I have questions about **previous instruction** or **exercise notebooks**
 - *Post* in #2020-august-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 #2020-august-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 staff member
- **I'm having trouble logging in** to RStudio Server
 - *Direct Message* a staff member

Trouble logging into Zoom and Slack? *Email* dataCoLAB@andrew.cmu.edu