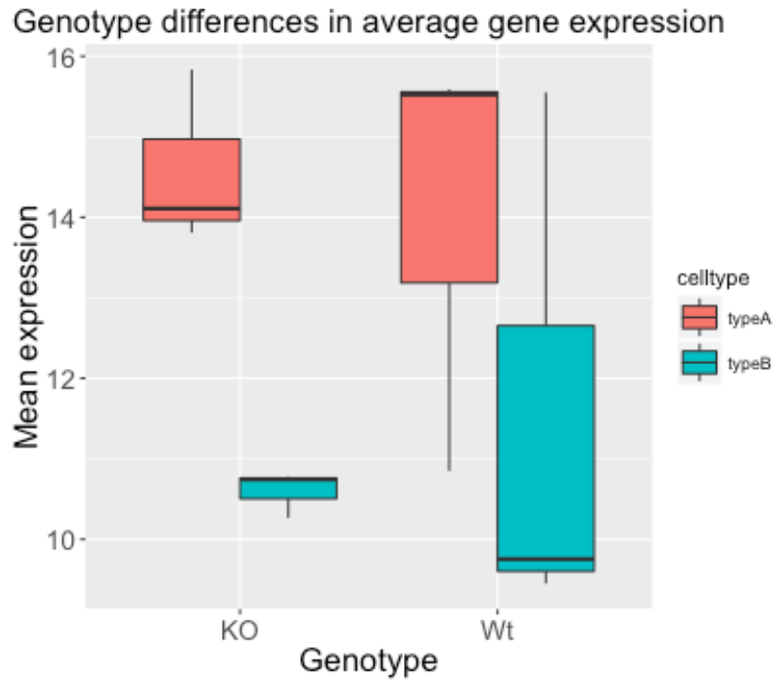
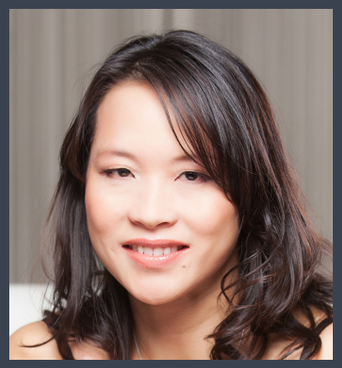


Introduction to R

Harvard Chan Bioinformatics Core

https://tinyurl.com/hbc-r-online

Sponsored by DF/HCC, CFAR, and HMS Foundry



Shannan Ho Sui
Director

Victor Barrera

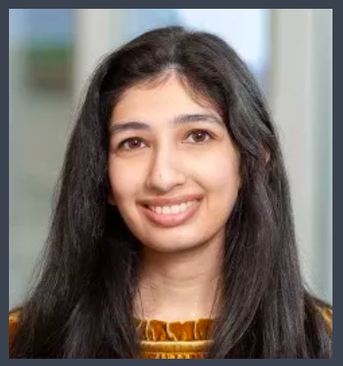
James Billingsley

Zhu Zhuo

Meeta Mistry
Interim Director of Education

Heather Wick

Will Gammerdinger



Noor Sohail

Emma Berdan

Sergey Naumenko

Maria Simoneau

Consulting

- Transcriptomics: bulk, single cell, small RNA
- Epigenomics: ChIP-seq, CUT&RUN, ATAC-seq, DNA methylation
- Variant discovery: WGS, resequencing, exome-seq and CNV
- Multiomics integration
- Spatial biology
- Experimental design and grant support

http://bioinformatics.sph.harvard.edu/

CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

Training

A key component of the HBC's mission is its training initiative. Our dedicated training team holds workshop to help researchers at Harvard better understand analytical methods for NGS data.

HBC's training team is made up of four PhD-level scientists who devote substantial time to material development, training and community building/outreach. All members of the training team also participate in consultations on research projects to ensure they remain up-to-date on current best practices in NGS analysis.

Our hands-on workshops focus on **basic data skills** and **analysis of high-throughput sequencing data**, with an emphasis on **experimental design**, current **best practices** and **reproducibility**. Our workshops are designed for **wet-lab biologists** aiming to independently design sequencing-based experiments and analysing the resulting data.

We offer three types of workshops:

1. Short, 3-hour monthly workshops (*Current topics in bioinformatics*)
2. Basic Data Skills**
3. Advanced Topics: Analysis of high-throughput sequencing (NGS) data**

***The basic data skills workshops serve as the foundation for the advanced workshops.*

http://bioinformatics.sph.harvard.edu/training/

https://hbctraining.github.io/main/

Training

A key component of the H
researchers at Harvard b

HBC's training team is m
training and community b
research projects to ensu

Our hands-on workshops
an emphasis on **experim
for **wet-lab biologists** ai
data.

We offer three types of w

1. Short, 3-hour monthly
2. Basic Data Skills**
3. Advanced Topics: Ana

***The basic data skills*

Our dedicated training team holds workshop to help
or NGS data.

o devote substantial time to material development,
raining team also participate in consultations on
best practices in NGS analysis.

ysis of high-throughput sequencing data, with
and **reproducibility**. Our workshops are designed
cing-based experiments and analysing the resulting

matics)

NGS) data**

r the advanced workshops.

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

http://bioinformatics.sph.harvard.edu/training/

https://hbctraining.github.io/main/

Training

A key component of the H
researchers at Harvard b

HBC's training team is m
training and community b
research projects to ensu

Our hands-on workshops
an emphasis on **experim
for **wet-lab biologists** ai
data.

We offer three types of w

1. Short, 3-hour monthly
2. Basic Data Skills**
3. Advanced Topics: Ana

***The basic data skills*

Our dedicated training team holds workshop to help
or NGS data.

o devote substantial time to material development,
raining team also participate in consultations on
best practices in NGS analysis.

ysis of high-throughput sequencing data, with
and **reproducibility**. Our workshops are designed
cing-based experiments and analysing the resulting

matics)

NGS) data**

r the advanced workshops.

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

DF/HCC DANA-FARBER / HARVARD CANCER CENTER

HARVARD CATALYST
THE HARVARD CLINICAL AND TRANSLATIONAL SCIENCE CENTER

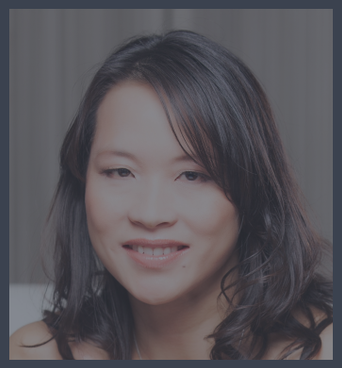
HARVARD MEDICAL SCHOOL

CFAR HARVARD UNIVERSITY CENTER FOR AIDS RESEARCH

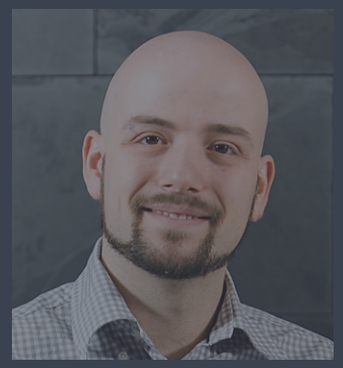
http://bioinformatics.sph.harvard.edu/training/

https://hbctraining.github.io/main/

Introductions!



Shannan Ho Sui
Director



Victor Barrera

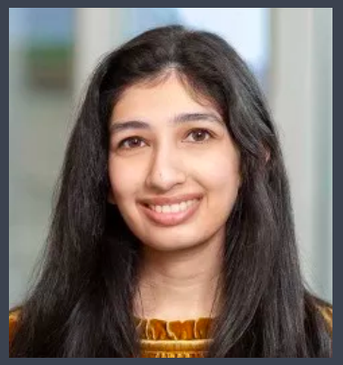
James Billingsley

Zhu Zhuo

Meeta Mistry
Interim Director of Education

Heather Wick

Will Gammerdinger



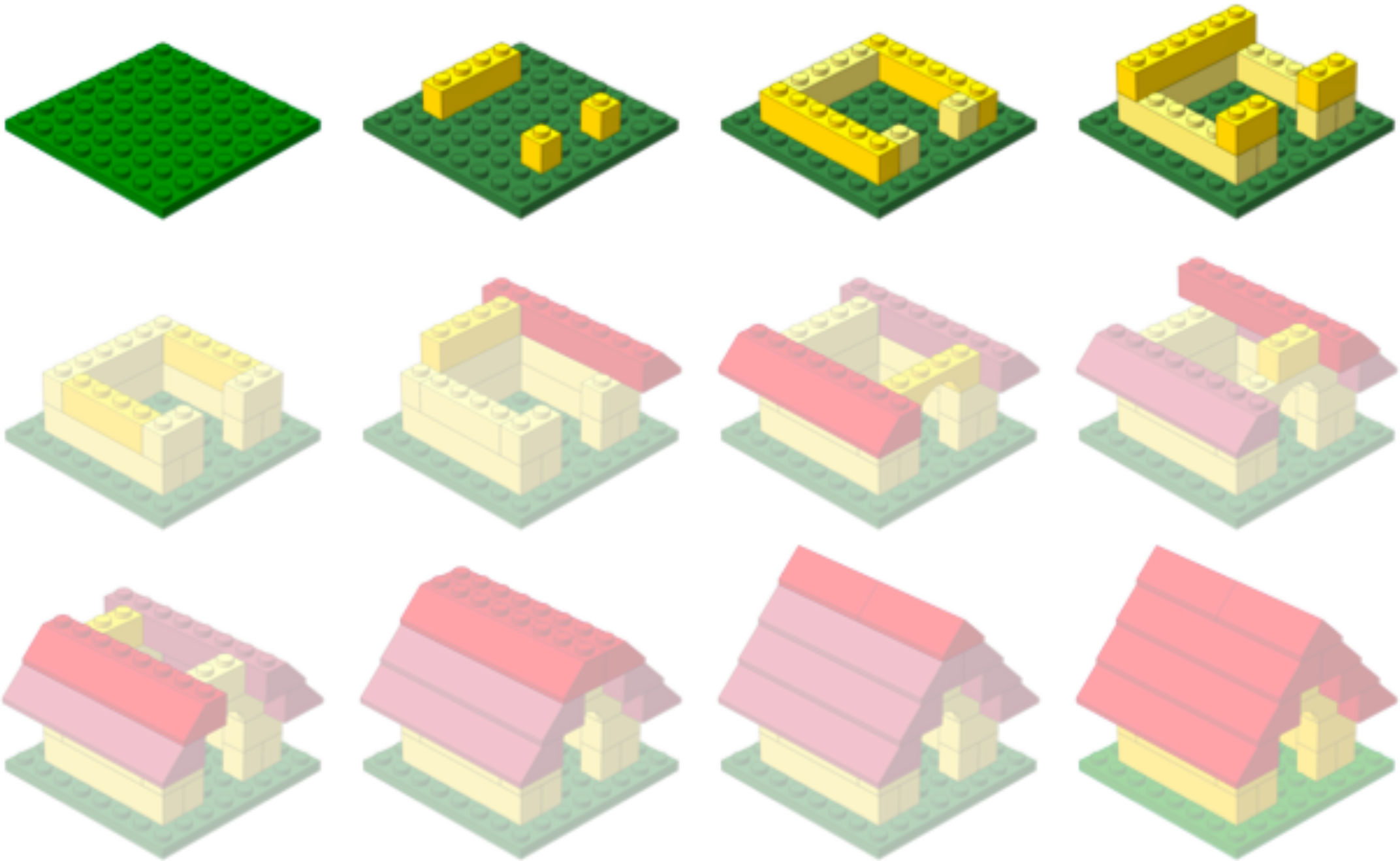
Noor Sohail

Emma Berdan

Sergey Naumenko

Maria Simoneau

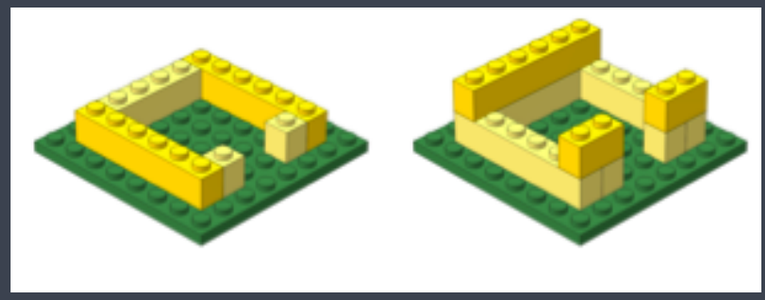
Workshop Scope…



Learning R

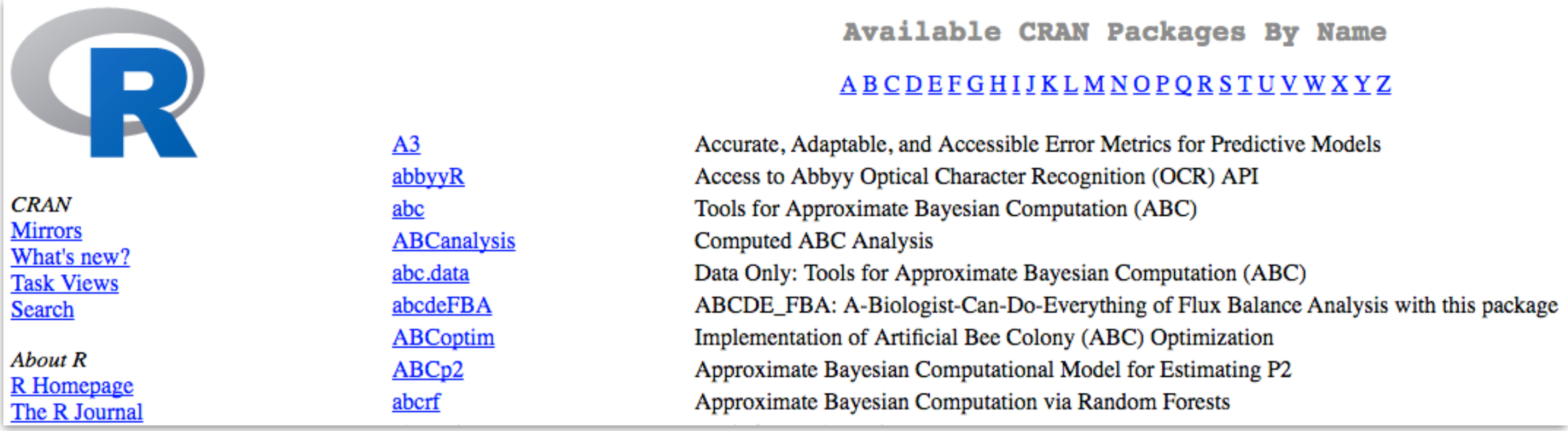
http://anoved.net/tag/lego/page/3/

Workshop Scope



- ✓ Comfortably use RStudio (a graphical interface for R)
- ✓ Fluently interact with R using RStudio
- ✓ Become familiar with R syntax
- ✓ Understand data structures in R
- ✓ Inspect and manipulate data structures
- ✓ Install packages and use functions in R

CRAN
(Comprehensive R Archive Network)



CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Available CRAN Packages By Name

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A3	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
abbyyR	Access to Abbyy Optical Character Recognition (OCR) API
abc	Tools for Approximate Bayesian Computation (ABC)
ABCanalysis	Computed ABC Analysis
abc.data	Data Only: Tools for Approximate Bayesian Computation (ABC)
abcdeFBA	ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
ABCoptim	Implementation of Artificial Bee Colony (ABC) Optimization
ABCp2	Approximate Bayesian Computational Model for Estimating P2
abcrf	Approximate Bayesian Computation via Random Forests

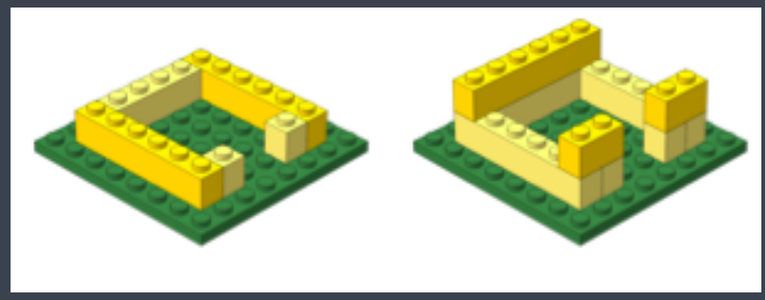
- The main repository for R packages
- Easy to install

https://cran.r-project.org/

- An alternative package repository; "..provides tools for the analysis and comprehension of *high-throughput genomic data*."
- Includes (but is not limited to) tools for:
    - performing statistical analysis
    - accessing public datasets
- Open source and open development
- Free

www.bioconductor.org

Workshop Scope



- Comfortably use RStudio (a graphical interface for R)
- Fluently interact with R using RStudio
- Become familiar with R syntax
- Understand data structures in R
- Inspect and manipulate data structures
- Install packages and use functions in R
- ✓ Visualize data using *ggplot2*
- ✓ Utilize pipes, tibbles and functions from the Tidyverse package suite

Logistics

Course webpage

https://tinyurl.com/hbc-r-online

Course webpage

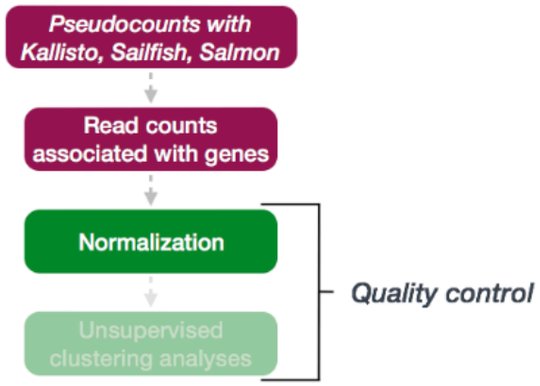
Approximate time: 60 minutes

Learning Objectives

- Explore different types of normalization methods
- Become familiar with the `DESeqDataSet` object
- Understand how to normalize counts using DESeq2

Normalization

The first step in the DE analysis workflow is count normalization, which is necessary to make accurate comparisons of gene expression between samples.



Course schedule online

Workshop Schedule

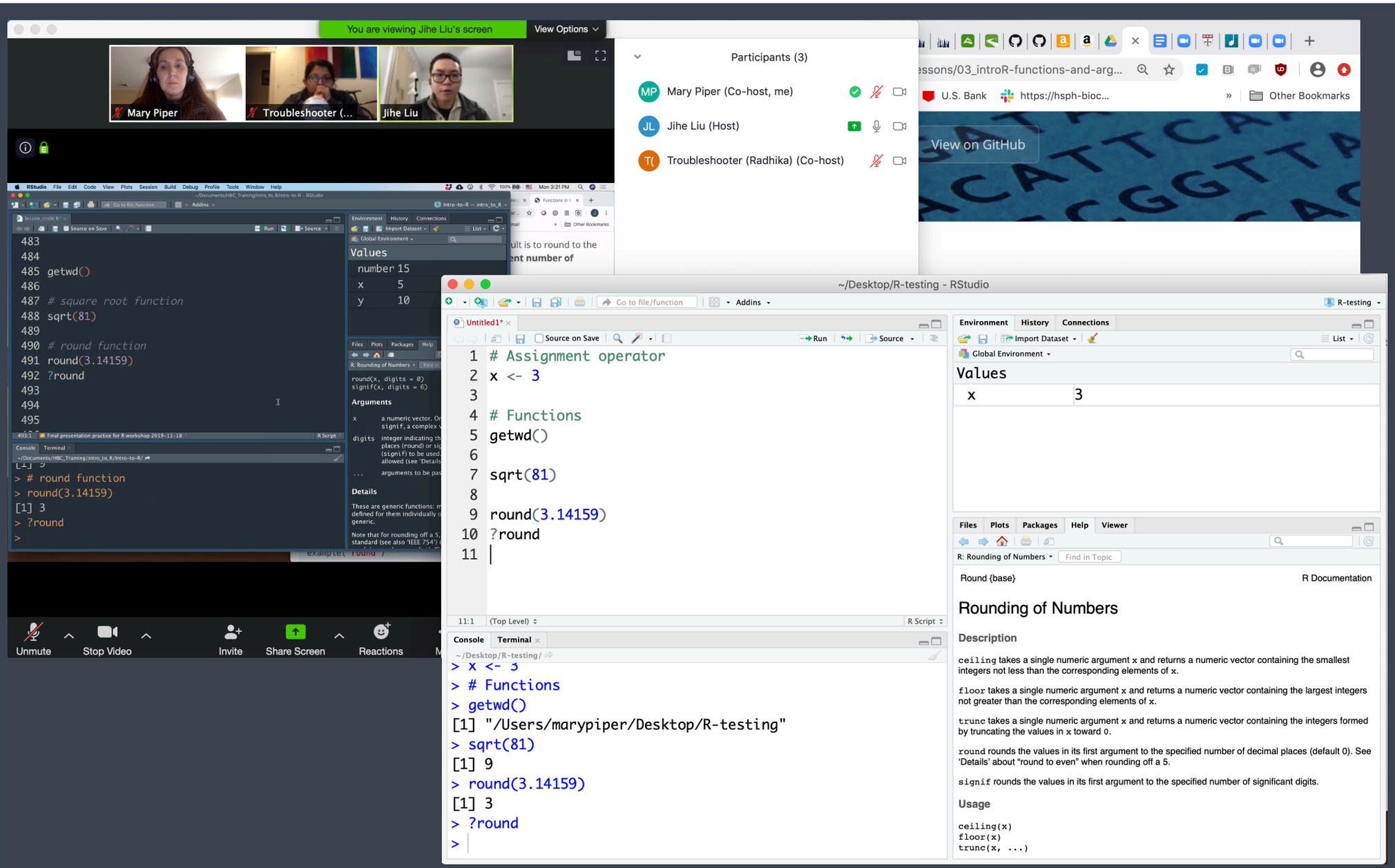
Day 1

Time	Topic	Instructor
10:00 - 10:30	Workshop Introduction	Jihe
10:30 - 11:45	Introduction to R and RStudio	Radhika
11:45 - 12:00	Overview of self-learning materials and homework submission	Mary

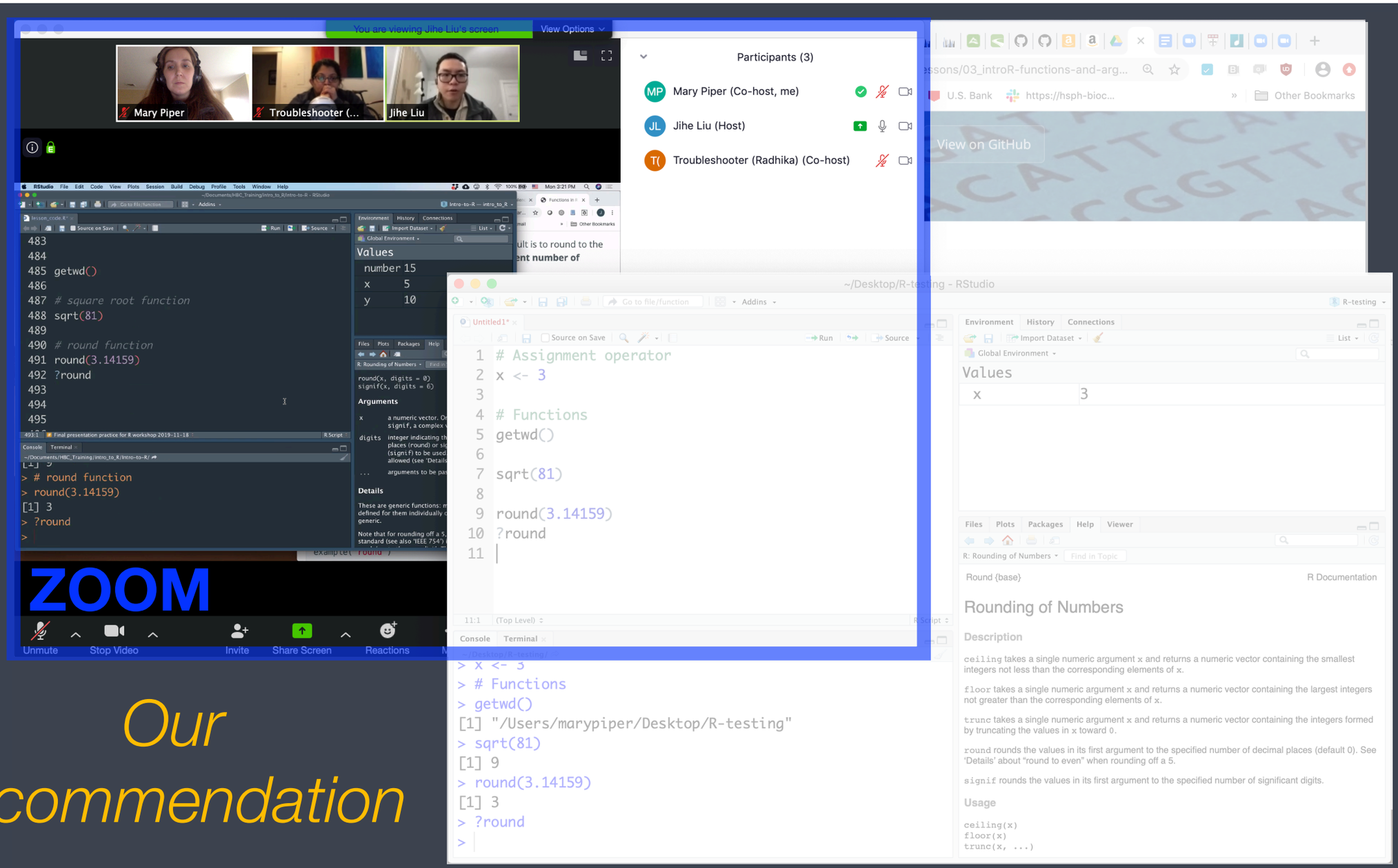
Before the next class:

1. Please **study the contents** and **work through all the code** within the following lessons:
    - R Syntax and Data Structure
    - Functions and Arguments
    - Reading in and inspecting data
2. **Complete the exercises**:
    - Each lesson above contain exercises; please go through each of them.
    - **Copy over** your code from the exercises into a text file.
    - **Upload the saved text file** to Dropbox the **day before the next class**.

Single screen & 3 windows?

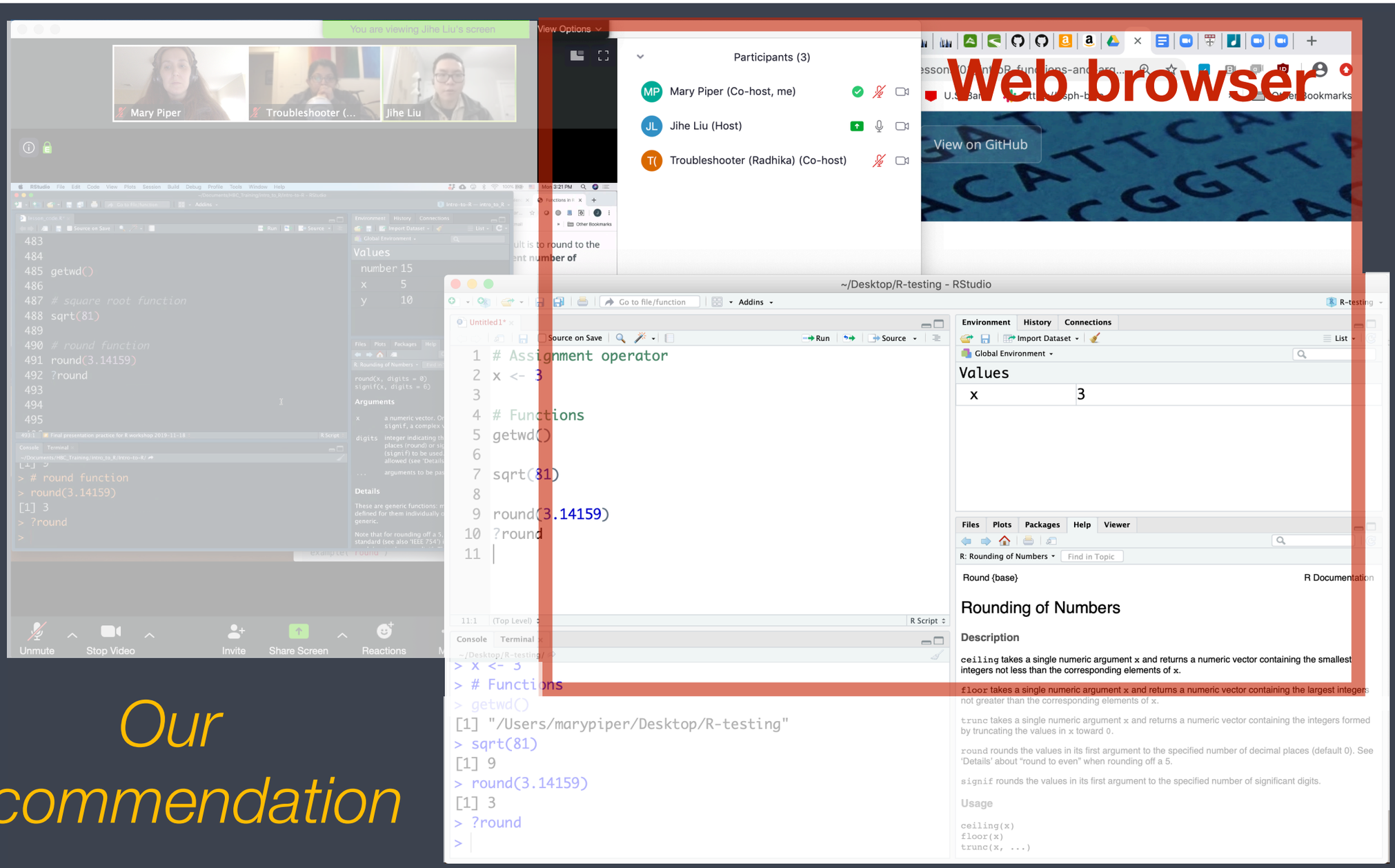


Single screen & 3 windows?



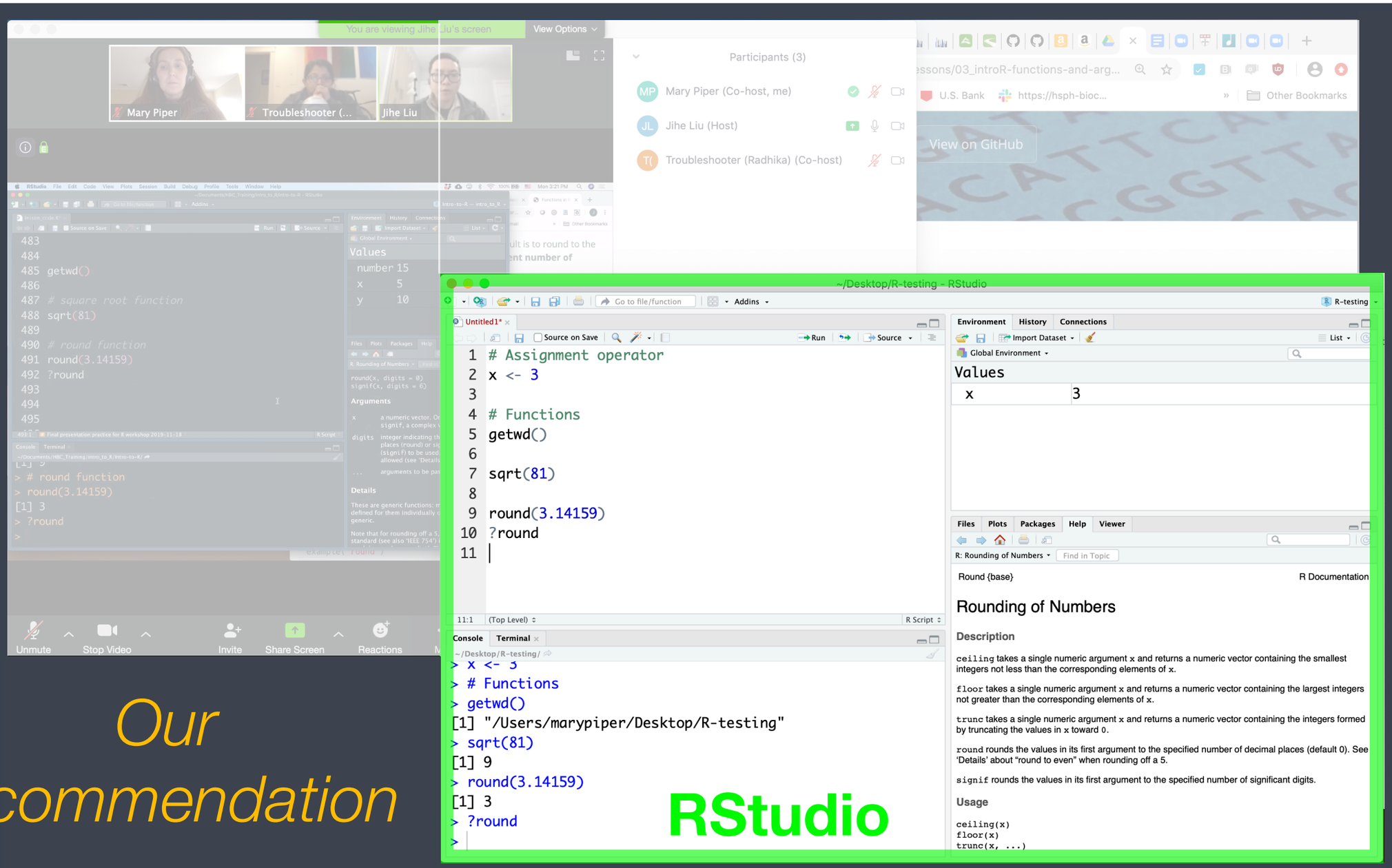
Our recommendation

Single screen & 3 windows?



Our recommendation

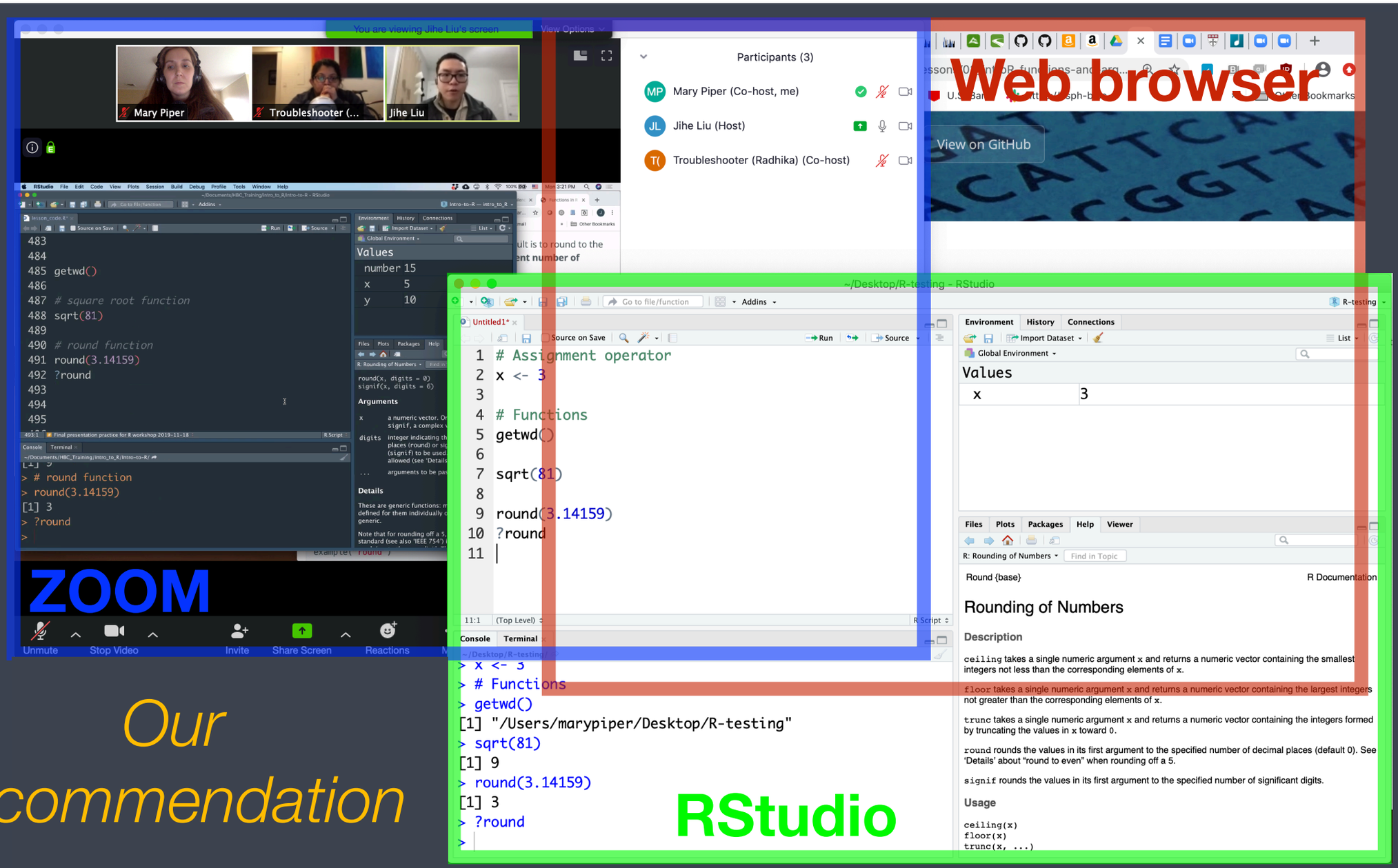
Single screen & 3 windows?



Our recommendation

RStudio

Single screen & 3 windows?



Our recommendation

Course participation

- Please keep your videos on, we would love to see your faces!
- Mandatory review of self-learning lessons and assignments
- Attendance required for all classes
- Your questions and active participation drive learning
- We look forward to all of your questions!

Homework and Expectations

- At-home lessons and exercises after each session
- Cover material not previously discussed
- Provides us feedback to help pace the course appropriately
- 3-5 hours to complete
- Homework load is heavier in the beginning of this workshop series and tapers off

Odds and Ends (1/2)

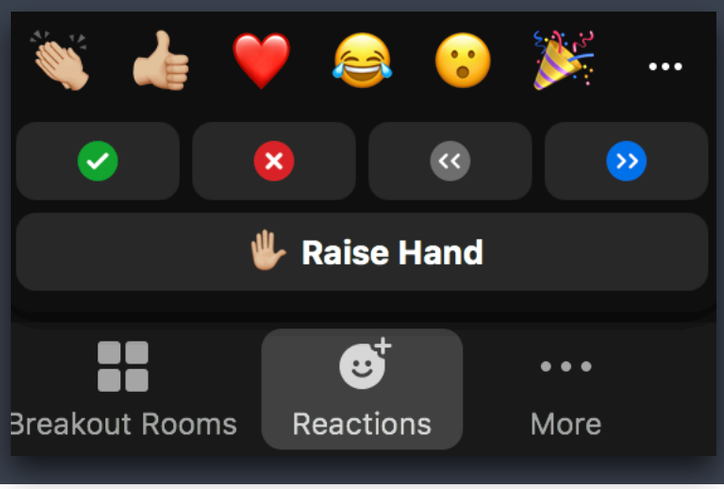
- Quit/minimize all applications that are not required for class

Odds and Ends (1/2)

- Quit/minimize all applications that are not required for class
- Captioning is available upon request

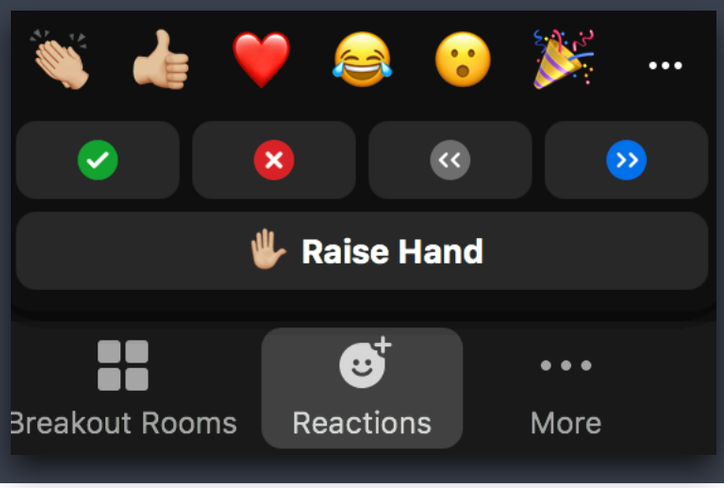
Odds and Ends (1/2)

- Quit/minimize all applications that are not required for class
- Captioning is available upon request
- Are you all set?
  -  = "agree", "I'm all set" (equivalent to a green post-it)
  -  = "disagree", "I need help" (equivalent to a red post-it)



Odds and Ends (2/2)

- Questions for the presenter?
  - Post the question in the Chat window OR
  - ✋ **Raise Hand** when the presenter asks for questions
  - Let the **Moderator** know
- Technical difficulties with software?
  - Start a private chat with the **Troubleshooter** with a description of the problem.



Contact us!

HBC training team: hbctraining@hsph.harvard.edu

HBC consulting: bioinformatics@hsph.harvard.edu

Twitter

@bioinfocore