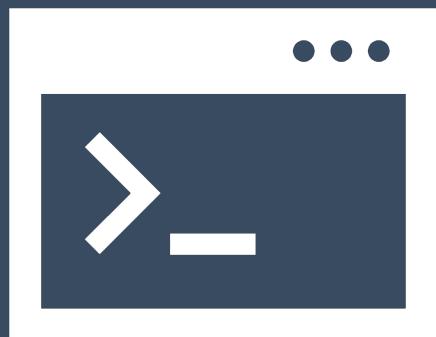




HBC
Harvard Chan Bioinformatics Core

Introduction to R

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



Harvard Chan Bioinformatics Core



Introductions!





Shannan Ho Sui
Director



Meeta Mistry
Associate Director



Lorena Pantano
Director of Bioinformatics Platform



John Quackenbush
Faculty Advisor



Upen Bhattacharai



Heather Wick



Will Gammie



Noor Sohail



Elizabeth Partan



Alex Bartlett



Emma Berdan



James Billingsley



Zhu Zhuo



Maria Simoneau



Shannan Ho Sui
Director



Meeta Mistry
Associate Director



Lorena Pantano
Director of Bioinformatics Platform



John Quackenbush
Faculty Advisor



Upen Bhattacharai



Heather Wick



Will Gammie



Noor Sohail



Elizabeth Partan



Alex Bartlett



Emma Berdan



James Billingsley



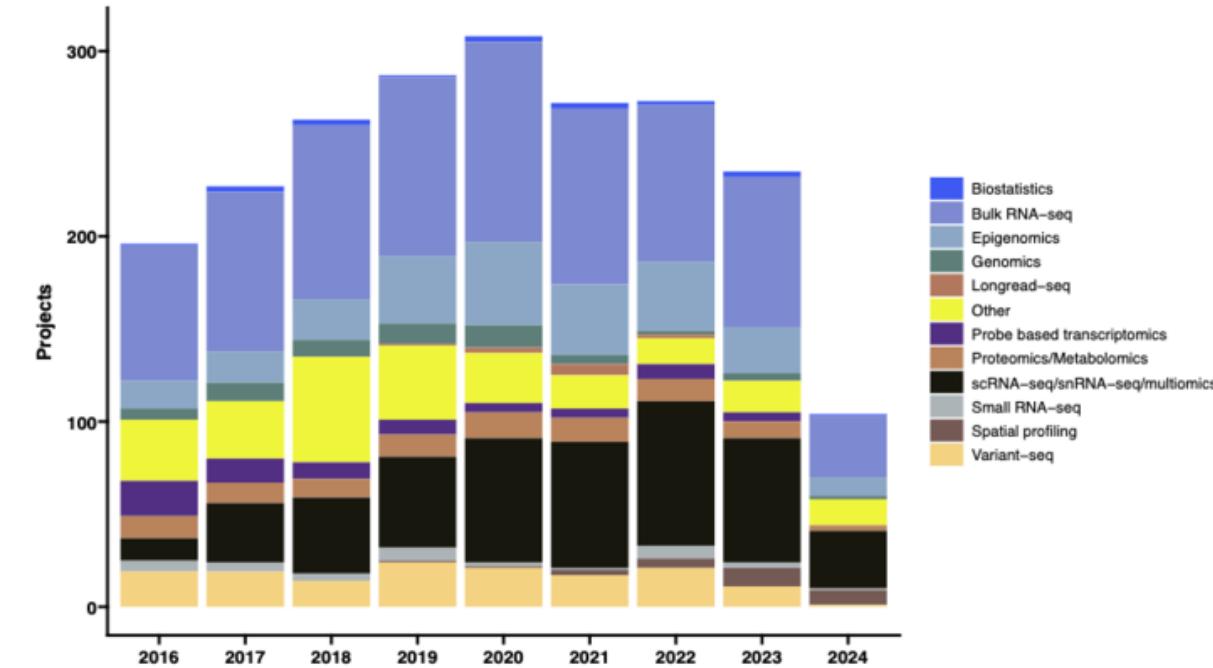
Zhu Zhuo



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



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



HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS



Training

- ❖ Hands-on workshops design to reflect best practices, reproducibility and an emphasis on experimental design
 - ❖ Basic Data Skills
 - ❖ Shell
 - ❖ R
 - ❖ Advanced Topics: Analysis of high-throughput sequencing data
 - ❖ Chromatin Biology
 - ❖ Bulk RNA-seq
 - ❖ Differential Gene Expression
 - ❖ scRNA-seq
 - ❖ Variant Calling
 - ❖ Current Topics in Bioinformatics

Training

- ❖ Hands-on workshops design to reflect best practices, reproducibility and an emphasis on experimental design
 - ❖ Basic Data Skills
 - ❖ Shell
 - ❖ R
 - ❖ Advanced Topics: Analysis of high-throughput sequencing data
 - ❖ Chromatin Biology
 - ❖ Bulk RNA-seq
 - ❖ Differential Gene Expression
 - ❖ scRNA-seq
 - ❖ Variant Calling
 - ❖ Current Topics in Bioinformatics

<https://bioinformatics.sph.harvard.edu/training>



**HARVARD
T.H. CHAN**
SCHOOL OF PUBLIC HEALTH

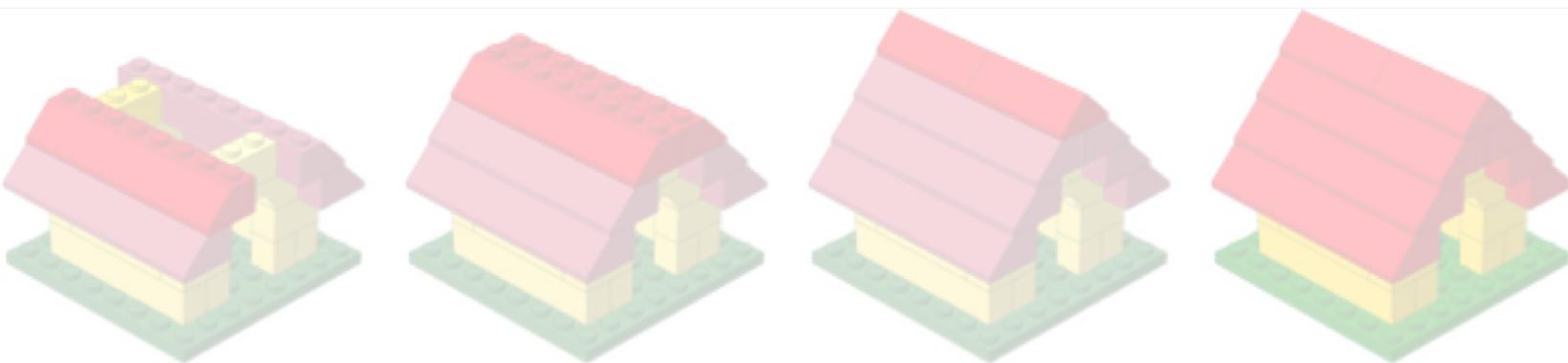
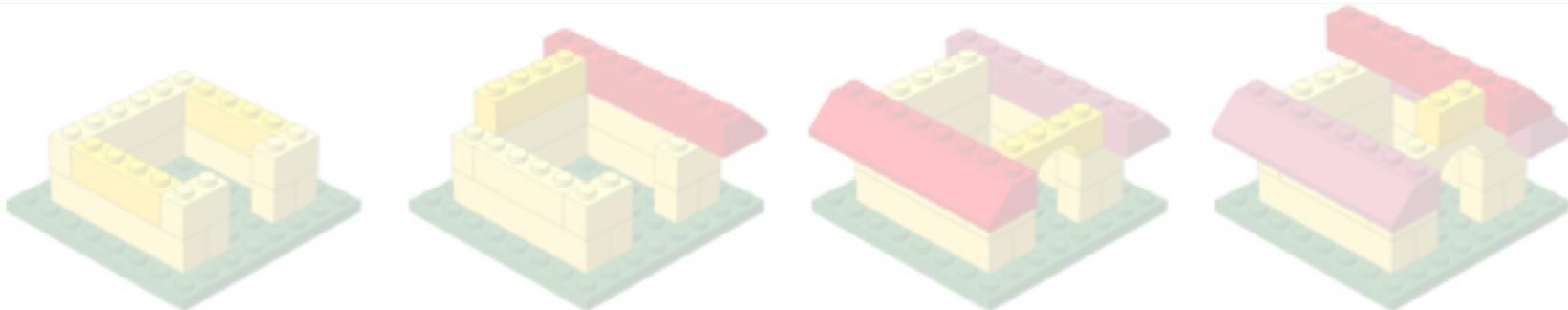
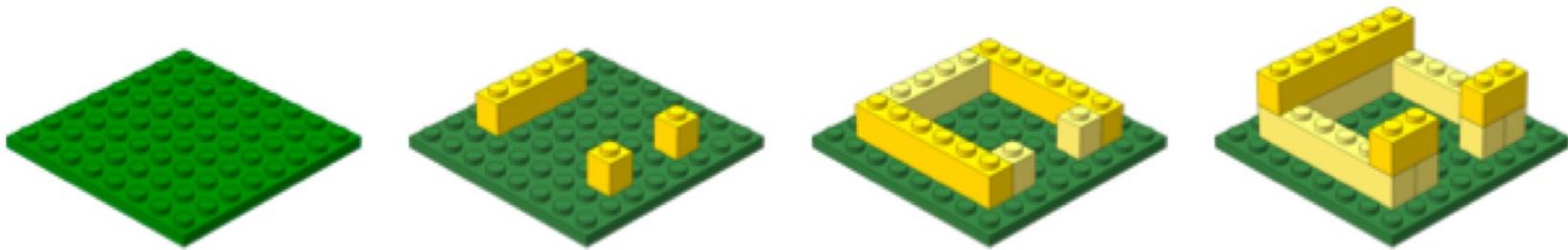
DF/HCC
DANA-FARBER / HARVARD CANCER CENTER



THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

 **HARVARD**
MEDICAL SCHOOL

Workshop scope



Learning R

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



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

About R
[R Homepage](#)
[The R Journal](#)

[A3](#)
[abbyyR](#)
[abc](#)
[ABCAnalysis](#)
[abc.data](#)
[abcdeFBA](#)
[ABCOptim](#)
[ABCp2](#)
[abcrf](#)

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](#)

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

- ❖ Comprehensive R Archive Network
- ❖ The main repository for R packages
- ❖ Easy to install



- ❖ 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

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 schedule

Workshop Schedule

Day 1

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

Before the next class:

I. Please **study the contents** and **work through all the code** within the following lessons:

1. [R Syntax and Data Structure](#)

Click here for a preview of this lesson

2. [Functions and Arguments](#)

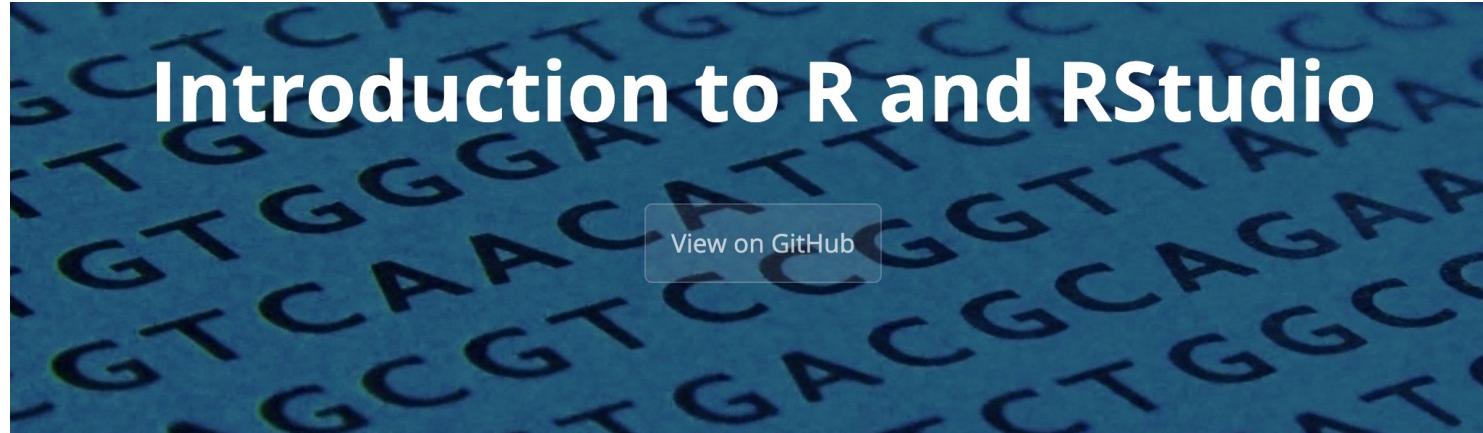
Click here for a preview of this lesson

3. [Reading in and inspecting data](#)

Click here for a preview of this lesson

Course materials

- ❖ We continuously update our materials to reflect changes in the field/software

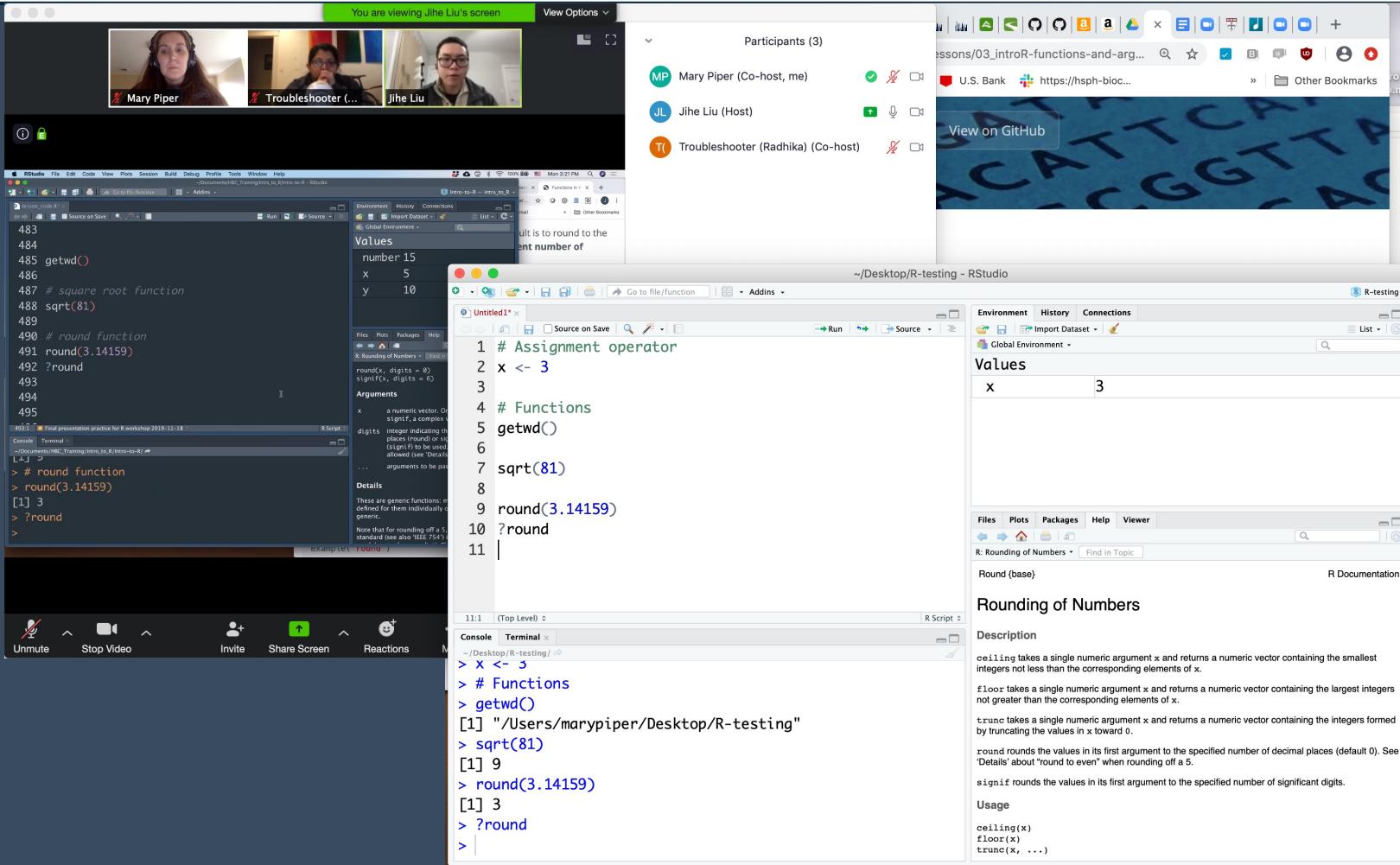


Approximate time: 45 minutes

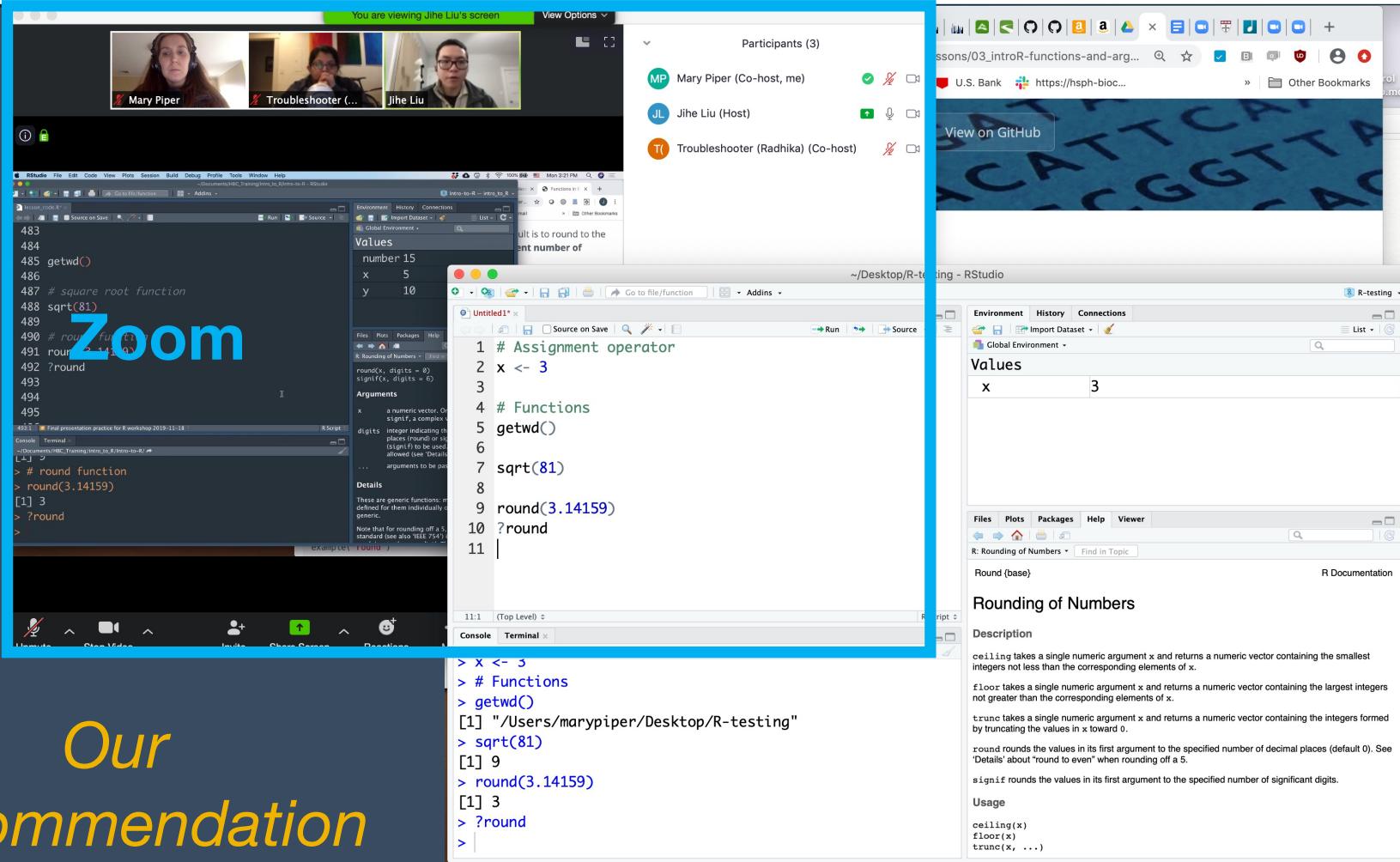
Learning Objectives

- Describe what R and RStudio are.
- Interact with R using RStudio.
- Familiarize various components of RStudio.
- Employ variables in R.

Single Screen & 3 Windows

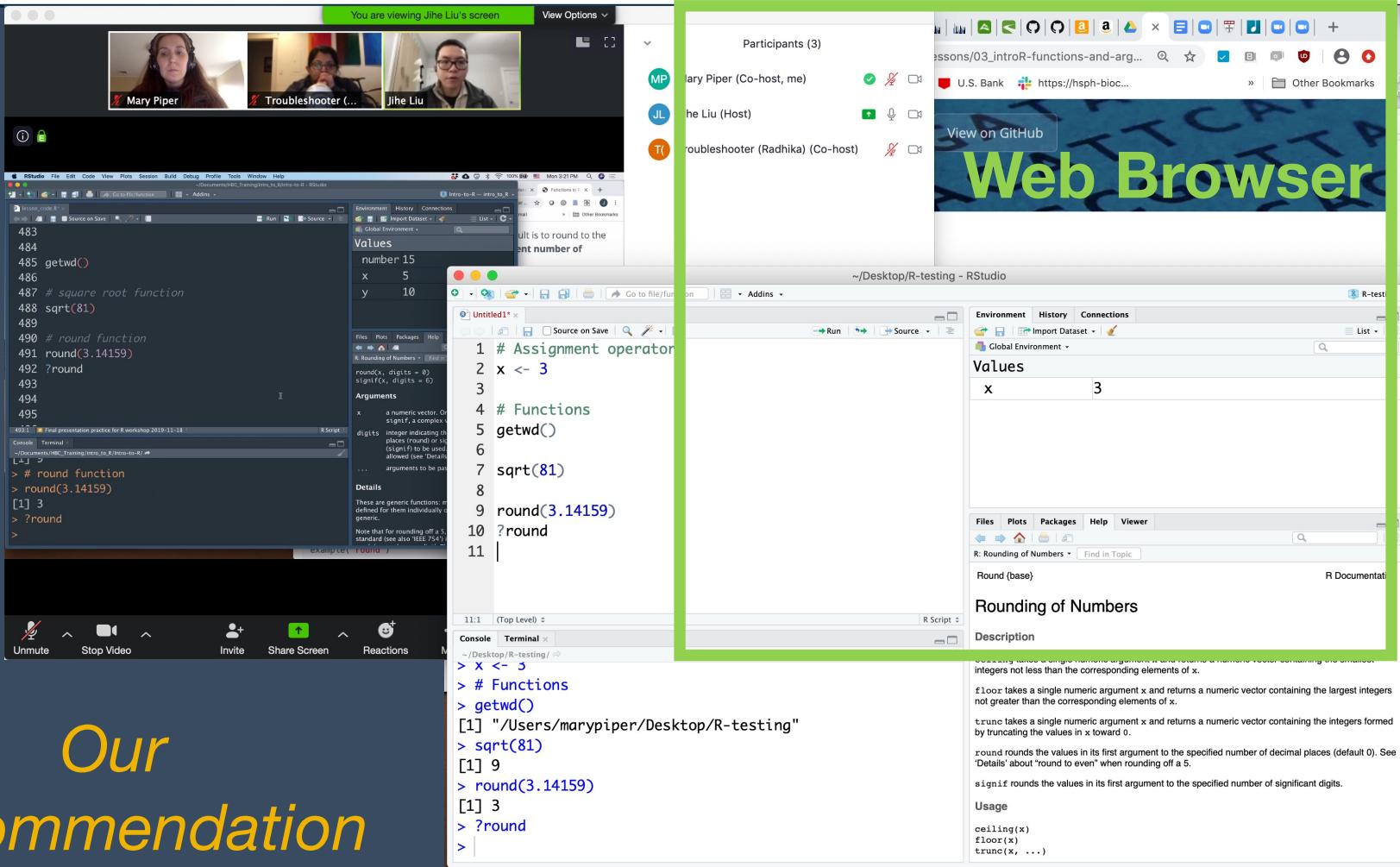


Single Screen & 3 Windows



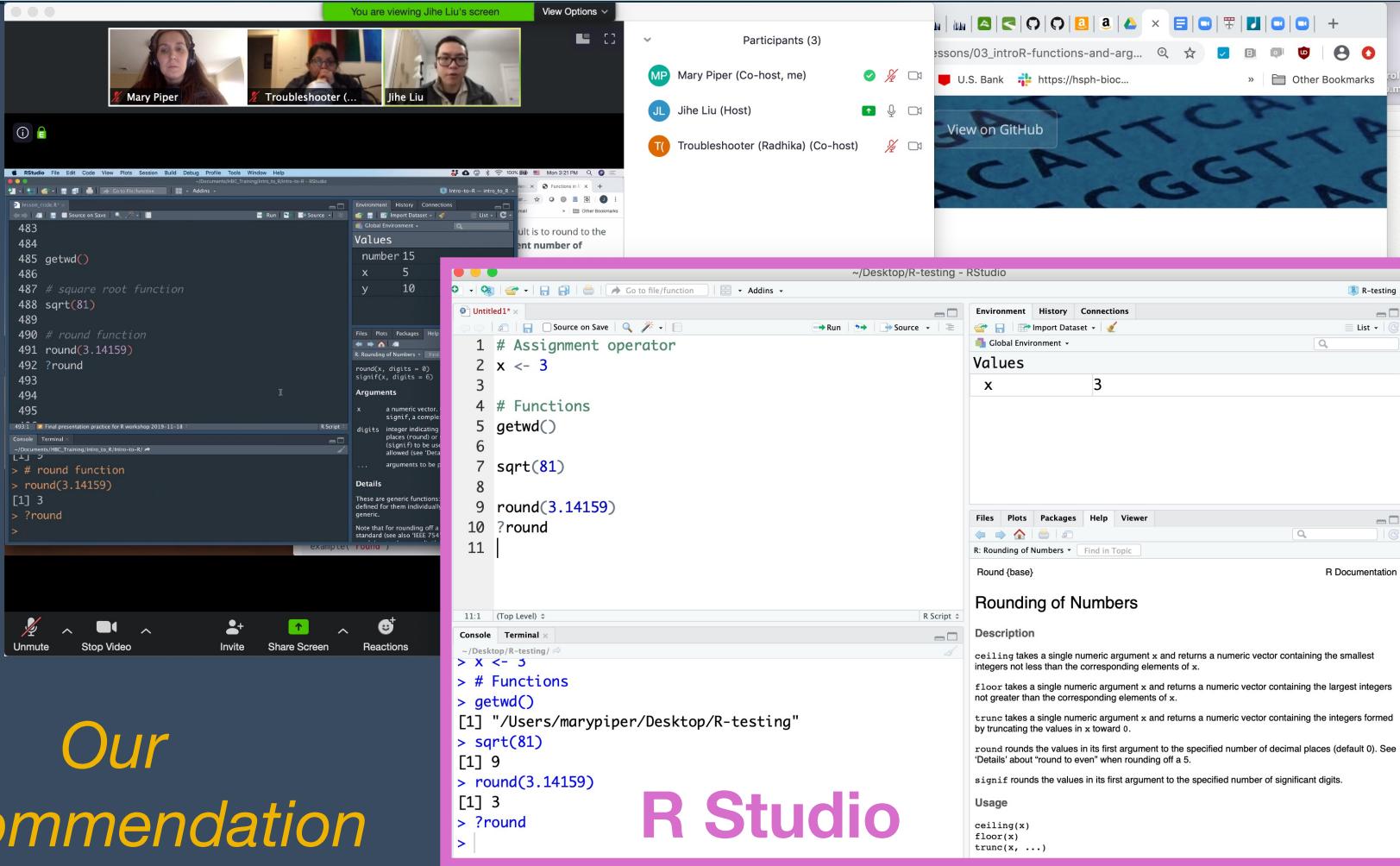
*Our
Recommendation*

Single Screen & 3 Windows



*Our
Recommendation*

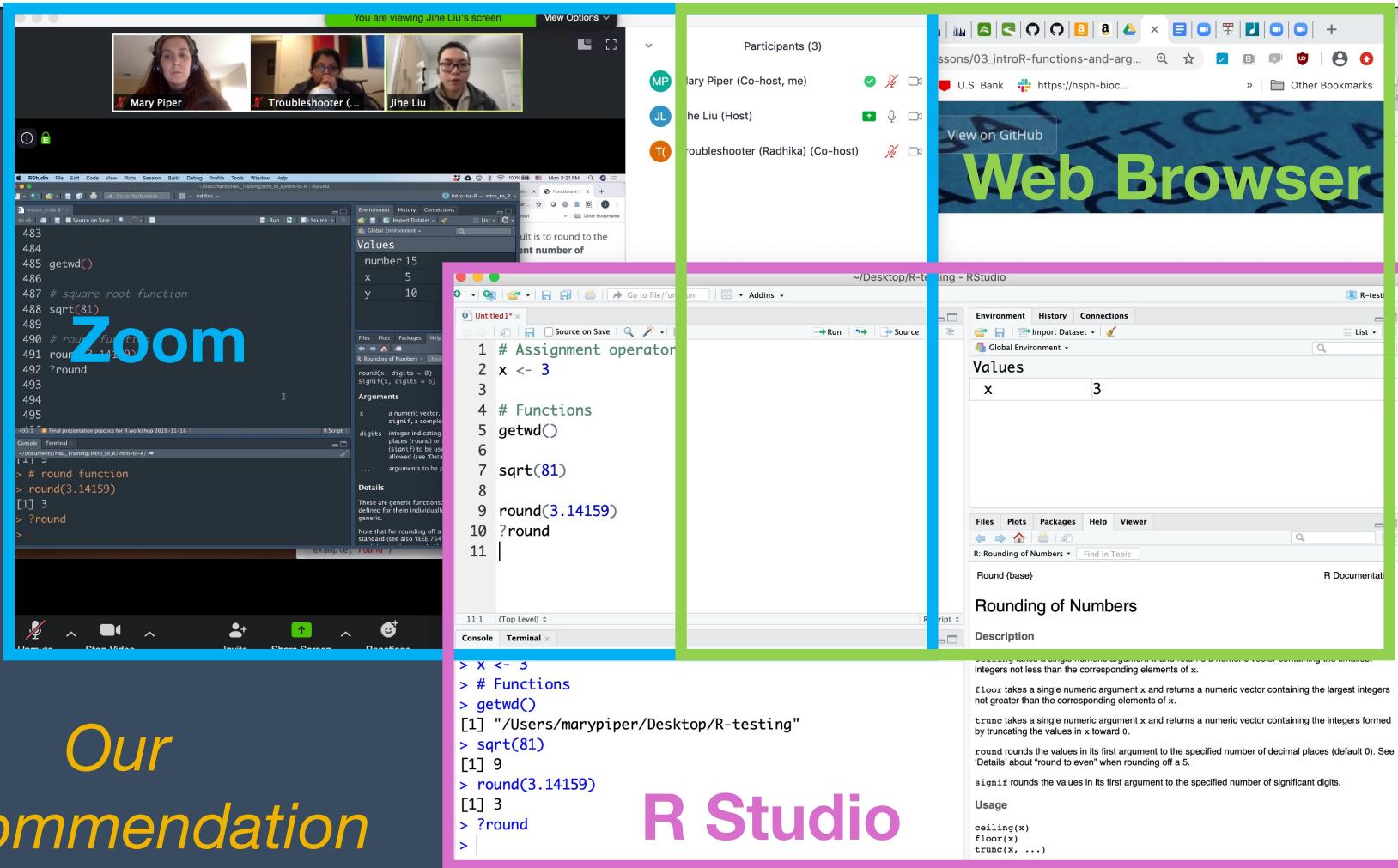
Single Screen & 3 Windows



Our
Recommendation

R Studio

Single Screen & 3 Windows



Course participation

- ❖ 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!**



Course participation

- ❖ 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

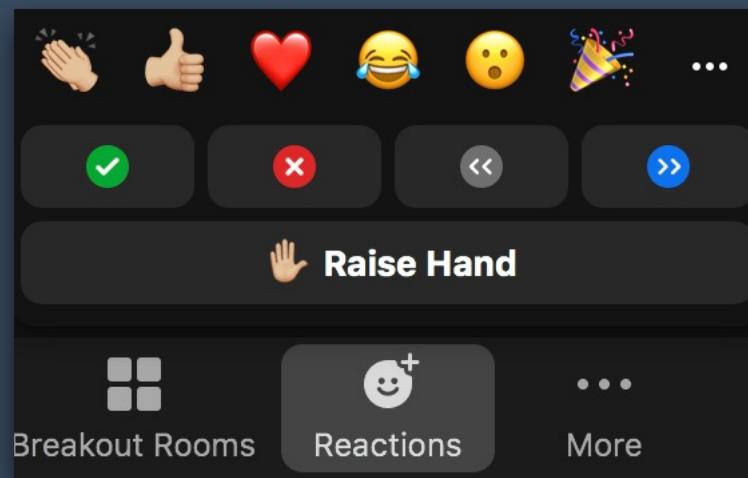
Using AI for Assignments

- ❖ Do
 - ❖ Try to resolve error messages with it
 - ❖ Test code written by AI on a dataset where you have expected results
 - ❖ Take the time to review the generated code line-by-line
- ❖ Don't
 - ❖ Implement it in replacement to learning
 - ❖ Write code that you don't understand
 - ❖ Assume the output from an AI process is correct

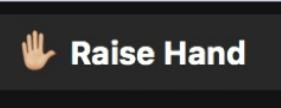
Odds & Ends

- ❖ Quit/minimize all applications that are not required for class
- ❖ Are you all set?

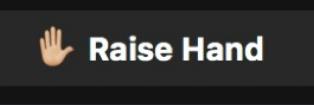
- ❖  = "agree", "I'm all set"
- ❖  = "disagree", "I need help"



Odds & Ends

- ❖ Questions for the presenter?
 - ❖ Post the question in the Chat window OR
 - ❖  when the presenter asks for questions
 - ❖ Let the Moderator know

Odds & Ends

- ❖ Questions for the presenter?
 - ❖ Post the question in the Chat window OR
 - ❖  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